Project Deliverable

Project Number:	Project Acronym:	Project Title:
287901	BUTLER	uBiquitous, secUre inTernet-of-things with Location and contExt-awaReness

Instrument:	Thematic Priority
Integrated Project	Internet of Things

Title:

D2.5 - IoT Enabling Technologies and Future Developments

Contractual Delivery Date:	Actual Delivery Date:
September 2014	October 2014

Start Date of Project:	Duration:
October, 1st 2011	36 months

Organization name of lead contractor of this deliverable:	Document Version:
Istituto Superiore Mario Boella (ISMB)	V1.1

Dissemination level (Project co-funded by the European Commission within the Seventh Framework Programme

PU	Public	**X**
PP	Restricted to other program participants (including the Commission)	
RE	Restricted to a group defined by the consortium (including the Commission)	
CO	Confidential, only for members of the consortium (including the Commission)	

Authors (organisations):

Francesco Sottile, Mirko Franceschinis, Zhoubing Xiong and Prabhakaran Kasinathan (ISMB)

Philippe Smadja and Patrick Enjolras (GTO)

Giuseppe Abreu, Stefano Severi, Omotayo Oshiga, Satya Vuppala and Simona Poilinca (JUB)

Arun Ramakrishnan and Davy Preuveneers (KUL)

Hennebert Christine, Iulia Tunaru, Benoît Denis and Luiz Henrique Suraty Filho (CEA)

Jani Saloranta, Davide Macagnano and Giuseppe Destino (OULU)

María-Fernanda Salazar and Miguel-Angel Monjas (ERC)

Foued Melakessou (UL)

Aliaksei Andrushevich (iHL)

Abstract:

In the context of the Internet of Things (IoT), the envisioned next evolution of the Internet which the BUTLER project is completely focused on, this document poses the attention on the enabling technologies, identifying three of them which assume a relevant importance for different reasons: localization, behavioral models, and security and privacy. Localization capability, the knowledge that a node has of its own absolute position along time as well as the ability of discovering in almost real-time the location of other nodes, is a fundamental key to provide data with augmented value: localization awareness is indeed one side of context awareness. On a parallel plane, behavioral models are semantic powerful tools that can equally improve the system smartness by recognizing the user contexts without depending extensively on inputs from the users. Finally, security and privacy already represent a central issue in the current Internet due to the intensive use of wireless communication technologies, particularly vulnerable to external attacks exploiting the shared nature of the access medium, but the vast set of problems behind the privacy protection and the different data security aspects are destined to exponentially arise in a much more complex and populated system as the next Internet of Things is expected to become.

Keywords:

IoT technology, privacy and security, localization awareness, behavior modeling.

Disclaimer

Revision History

The following table describes the main changes done in the document since it was created.

Revision	Date	Description	Author (Organization)
V0.0	28/01/2014	Creation	ISMB
V0.1	07/02/2014	Discussion about the ToC	All involved partners
V0.2	25/06/2014	First draft of Chapter 3	ERC, ISMB, KUL, OULU
V0.3	08/07/2014	Second draft of Chapter 3	CEA, ISMB, JUB, UL
V0.4	14/07/2014	Third draft of Chapter 3	CEA, ERC, iHL, ISMB
V0.5	16/07/2014	Fourth draft of Chapter 3	CEA, GTO, ISMB, JUB
V0.6	21/07/2014	Fifth draft of Chapter 3	CEA, GTO, ISMB, UL
V0.7	02/09/2014	First draft of Chapter 4	All involved partners
V0.8	08/09/2014	Second draft of Chapter 4	CEA, ISMB, JUB, KUL
V0.9	09/09/2014	First draft of Chapter 3 and 4	CEA, GTO, ISMB, JUB
V0.91	11/09/2014	Second draft of Chapter 3 and 4	GTO, ISMB, OULU
V0.92	03/10/2014	Third draft of Chapter 3 and 4	GTO, ISMB, KUL, UL
V0.93	06/10/2014	Fourth draft of Chapter 3 and 4	CEA, GTO, ISMB
V0.94	06/10/2014	Added Summary of chapter 3	ISMB
V0.95	10/10/2014	Fifth draft of Chapter 3 and 4	CEA, GTO, JUB, ISMB, KUL
V0.96	13/10/2014	Sixth draft of Chapter 3 and 4	JUB, ISMB
V0.97	20/10/2014	Added summary and conclusion	ISMB
V0.98	21/10/2014	Error correction	All involved partners
V1.0	30/10/2014	First completed version	ISMB
V1.1	31/10/2014	Send to BUTLER board	ISMB

Executive Summary

The BUTLER project is generally focused on the so-called Internet of Things (IoT), the emerging paradigm also known as the internet of objects and likewise identified by other more or less official names which, by means of a few terms, try to summarize the concept of an extremely large community of objects able to interact with each other. The impetuous revolution of the IoT certainly represents the ongoing evolution of the modern data communication networks, but it cannot be simply limited to that. It is indeed much more than communication. It invests the generation, the interpretation, the contextualization and the sharing of a surprisingly enormous amount of data: the raw data generated by the billions of sources represented by sensors and generic objects is indeed transformed into information by advanced distributed processing, thus extremely gaining value. It even gives life to simple inanimate objects in virtue of processing and contextualization capabilities which they are endowed with through complex and advanced algorithms in the field of artificial intelligence. Finally, it embraces sociological implications.

In the context of BUTLER project the IoT challenge is studied and analyzed from multiple perspectives, with the aim of proposing and converging towards commonly shared approaches and system architectures. In these research efforts many issues are touched and dealt with, and three of them, namely security and privacy, localization, and behavioral modeling, are the main topics faced in this report due to their primary importance, as better highlighted in the prosecution of this Executive Summary.

In a global system where incredible amounts of data are generated and distributed, security and privacy are two delicate and critical issues. Privacy means that, depending on the content, a set of data could have restrictions to its access, foreseeing different levels of accessibility. Security, on the other side, must be guaranteed on the potential alteration and falsification attempts from malicious sources. In a complex system architecture as the IoT is, protecting and safeguarding these requirements might be hard and challenging even because continuously put in discussion by adverse enemy actions. A description of the enabling technologies related to security and privacy in the context of the BUTLER Security Framework opens the way to mostly detailed issues and results, respectively faced and achieved, regarding the different security technologies, prototypes and experiments developed and realized in BUTLER. Implementation issues are likewise considered. In particular, those ones concerning the implementation of the low-level security, application level security and the problem of the security bootstrapping are described and presented in a detailed manner. Throughout the document it is shown how the BUTLER security framework enables end-to-end security between a data provider and a data consumer while the designed security protocols insure confidentiality, integrity of the messages and authentication of the peers. The privacy approach requires to dissociate the security roles. For instance, it is fundamental that data can be received only by allowed entities while, on the contrary, all of the technical components like gateways and proxies usually in charge of conveying data shall not have access to the data. This principle insures that the data cannot be retrieved and used by entities without user controls.

Another fundamental capability of IoT systems, which cannot be eluded and whose strategic importance comes from the necessity of correlating and contextualizing data (one of the pillars behind the smartness of devices), is the localization, i.e., the capability of knowing the position of an object as time passes. In general, wireless localization is a fairly mature area of research, characterized by a vast and solid literature. Nonetheless, despite the formidable effort already put into the problem, wireless positioning is still shy of its potential as a truly ubiquitous and real-time locating technology. Difficulties come from the needs requested by strict requirements such as ubiquity and real-time: ubiquity requires the technology to be available in every environment; real-time location is indissolubly associated with automatic identification and tracking of location within an environment. It is well-known that, unfortunately, wireless localization systems are still inaccurate and unreliable in places such as urban cities and indoors, which are characterized by high multipath propagation, different disturbs and noise corrupting the data estimation reliability, and scarcity of Line of Sight (LoS)

conditions. It is for this reason that the study and development of algorithms able to fight against multipath and Non Line of Sight (NLoS) conditions for accurate indoor wireless localization are an emerging issue considered in BUTLER and in this report in particular. Entering in more technical details, an accurate ranging algorithm using super-resolution techniques over phase measurements for distance estimation between devices has been developed within BUTLER and is presented in this document. This algorithm has even already been implemented in ZIGPOS Localization Architecture. The possibility of performing multipoint ranging via this algorithm using orthogonal set of Golomb rulers has been envisioned and is reported as one of the localization-related results achieved. With distance estimation discussed, accurate, robust and efficient positioning algorithms for target localization using algebraic confidence via circular interval scaling and a hybrid cooperative algorithm are presented. To improve the performance of the positioning algorithms, therein target coordinates, a cooperative technique for detecting NLoS measurements is also presented.

A third support for high-level data processing and contextualization, equally essential, is the exploration of behavioral models. The goal is to recognize the user contexts, both effectively and efficiently, without depending extensively on input from the users. In line with the scope of the horizontal architecture of BUTLER [1], algorithms and tools have been developed or enhanced to this aim. One of the most relevant contributions is in advancing the integration between deterministic and probabilistic modelling of human behaviors. Two major blocks are at the basis of the proposed system. The first one concerns the behavioral SmartServer (SAMURAI), which seamlessly integrates various technologies such as semantic reasoning, Complex Event Processing (CEP), stream mining, as building blocks in a scalable way. The second one regards various algorithms that exploit in the best manner correlations between user contexts and causal information thus enabling versatile and robust recognition of user contexts. Multi-modal user behavior modelling and recognition is enhanced by indirect inference from semantic locations and improved electrical appliance usage algorithms. Since behavior recognition systems are complex and characterized by lot of parameters, the methodology definitively selected learns deployment trade-offs in order to facilitate the efficient deployment of software components in different BUTLER platforms. Other works in the domain of transfer learning aim at reducing user efforts in training as well as aids demographic analysis of user behaviors. Also, the work on contextual networking pays more attention to mobility support for masses by providing context-aware adaptation of various networking mechanisms. Moreover, from the user perspective, context synthesis and management approach proposed on the basis of CEP would enable them to define their own contexts of their interest even without advanced technical skills.

Considerable effort has been devoted within BUTLER project to the development and improvement of algorithms for data security and integrity and for privacy protection, the design and the implementation of robust localization protocols finalized to support context awareness services, finally the exploration of behavioral models able to enrich the semantic data processing and to let data further gain value. Nonetheless, the research on these topics is still at the preliminary stages and immense spaces are left for future achievements. An overview of possible research lines and in-depth analysis to face the challenges put in place by the identified issue of data security and privacy safeguard, localization capability for enabling context awareness, behavioral modelling derivation.

The chapters and the sections in which this document is structured give a detailed overview of the technical achievements with extensive citations to relevant scientific publications and level of integration of those scientific contributions in BUTLER platform, as well as the main challenges .

Contents

1 Acronyms

3GPP The 3rd Generation Partnership Project
6LoWPAN IPv6 LoW Power wireless Area Networks
AES Advanced Encryption Standard
AES-CCM AES Counter with CBC-MAC
AH Authentication Header
AoA Angle of Arrival
AODV Ad-hoc On-demand Distance Vector
AP Access Point
API Application Programming Interface
ARC Available Routing Construct
ARFF Attribute-Relation File Format
ARM Architectural Reference Model
ARX Autoregressive Exogenous Model
AWGN Additive White Gaussian Noise
BFS Breadth First Search
BN Bayesian Network
BP Belief Propagation
BSF Bootstrapping Server Function
BT BrTree
BUTLER uBiquitous, secUre inTernet-of-things with Location and contExt-awaReness
CBC Cipher Block Chaining
CBC-MAC Cipher Block Chaining Message Authentication Code
CDF Cumulative Density Function
CDMA Code Division Multiple Access
CEP Complex Event Processing
CEPFC Complex Event Processing Functional Component
CFD Coordinator-Function Device
CIS Circular Interval Scaling
CLT Central Limit Theorem
C-MDS Classical-Multidimensional Scaling
CoAP Constrained Application Protocol
CPU Central Processing Unit
CRLB Cramér-Rao lower bound
CWRR Continuous Wave Radar Ranging
DDN Dynamic Decision Networks
DFS Depth First Search
DI Directed Information
DM Device Manufacturer
DoA Direction of Arrival

DoS Denial-of-Service
DRL Drools Rule Language
DSRC Dedicated Short-Range Communication
DTLS Datagram Transport Layer Security
DzS Dantzig-Selector
ECA Event-Condition-Action
ECAES Elliptic Curve Augmented Encryption Scheme
ECC Elliptic Curve Cryptography
ECDH Elliptic Curve Diffie Hellman
ECDLP Elliptic Curve Discrete Logarithmic Problem
ECDSA Elliptic Curve Digital Signature Algorithm
ECIES Elliptic Curve Integrated Encryption Scheme
EE Edgeworth Expansion
EP Expectation Propagation
ERQ Equivalent Ranging Quality
ESP Encapsulating Security Protocol
ETSI European Telecommunications Standards Institute
EU European Union
FFD Full-Function Device
FFT Fast Fourier Transform
FHT Fixed Handover Threshold
FIM Fisher Information Matrix
FPT Fixed Prediction Threshold
FRA Fair Resource Allocation
FQDA Fuzzy Quantitative Decision Algorithm
GDOP Geometric Dilution of Precision
GK Gaussian Kernel
GNSS Global Navigation Satellite Systems
GP Global Platform
GPS Global Positioning System
GSM Global System for Mobile communications
HC-PF Hybrid-Cooperative Particle Filter
HCPP Hard-Core Point Process
HHT Hysteresis Handover Threshold
HP Hard Proactive
HPT Hysteresis Prediction Threshold
H-SPAWN Hybrid Sum-Product Algorithm over a Wireless Network
HTTP HyperText Transfer Protocol
HTTPS HyperText Transfer Protocol Secure
i.i.d. independent and identically distributed
IEEE Institute of Electrical and Electronic Engineers

iHL iHomeLab
IKE Internet key Exchange
IMSI International Mobile Subscriber Identity
INS INertial Sensors
IoT Internet of Things
IPSec Internet Protocol Security
IR-UWB Impulse Radio Ultra Wide Band
ISMB Istituto Superiore Mario Boella
ISP Internet Security Protocols
ITS Intelligent Transportation Systems
JSON JavaScript Object Notation
JUB Jacobs University Bremen
KL Kullback-Leibler
KMP Key Management Protocol
KUL KU Leuven
LAN Local Area Network
LARSO Least Absolute Residual and Selection Operator
LASSO Least Absolute Shrinkage and Selection Operator
LoS Line of Sight
LS Least Square
LTE Long Term Evolution
M2M Machine to Machine
MAC Medium Access Control
MANET Mobile Ad-hoc Network
MCS Mobile Crowd Sensing
MEMS Micro-Mlectro-Mechanical Sensors
MI Mutual Information
MNO Mobile Network Operator
MODA Multicriteria Optimization and Decision Analysis
MP Multilayer Perceptron
NARVAL Network Analysis and Routing eVALuation
NFC Near Field Communication
NGSI Next Generation Service Interfaces
NIALM Nonintrusive Appliance Load Monitoring
NLoS Non Line of Sight
NLoS2 Non Line of Sight Square
NSCL Network Service Capability Layer
NSM Network Subscription Manager
OMA Open Mobile Alliance
OSI Open Systems Interconnection
P2P Peer-to-Peer

PaaS Platform as a Service
PAWN Personal Area Wireless Network
PDF Probability Distribution Function
PDoA Phase-Difference of Arrival
PEB Position Error Bound
PF Particle Filter
PGFL Probability Generating Functional
PKC Public Key Cryptography
PKI Public Key Infrastructure
PPP Point-to-Point
PUF Physically Unclonable Function
PVT Position-Velocity-Time
QDV Quantitative Decision Algorithm
QEV Quantitative EValuation
QoC Quality-of-Context
QoS Quality of Service
RAM Random Access Memory
RF Radio Frequency
RFD Reduced-Function Device
ROM Read-Only Memory
RPL Routing Protocol for Low power and Lossy Networks
RSA Rivest Shamir Adleman
RSSI Received Signal Strength Indicator
SA Security Association
SaaS Software as a Service
SAMOA Scalable Advanced Massive Online Analysis
SAMURAI Streaming Architecture for Mobile and Ubiquitous RESTful Analysis and Intelligence
SDP Security Deployment and Maintenance
SE Secure Element
SIM Subscriber Identity Module
SKG secret key generation
SMACOF Scaling by Majorizing a COmplicated Function
SMDS Super MultiDimensional Scaling
SMO Sequential Minimal Optimization
SMS Short Message Service
SNR Signal-to-Noise Ratio
SP Soft Proactive
SPA Sum-Product Algorithm
SPARQL SPARQL Protocol and RDF Query Language
SRC Short-Range Communication
SSL Secure Sockets Layer

TA Trusted Applications
TEE Trust Execution Environment
TLS Transport Layer Security
TM Trust Manager
ToF Time of Flight
UDP User Datagram Protocol
UL University of Luxembourg
UMTS Universal Mobile Telecommunications System
URL Uniform Resource Locator
USB Universal Serial Bus
UTC Coordinated Universal Time
UWB Ultra Wide Band
VANET Vehicular Ad-hoc Network
WAN Wide Area Networks
WBAN Wireless Body Area Networks
WG Work Group
WSN Wireless Sensor Network
ZR ZeroR

2 Introduction

The core of this document is contained in Chapter 3 and Chapter 4, both parallel organized into three main sections (3.1, 3.2 and 3.3 for the former, 4.1, 4.2 and 4.3 for the latter) devoted to security and privacy, localization, and behavioral models, respectively. While Chapter 3 is dedicated to presenting the work developed around such issues within the BUTLER project, describing algorithms and protocols, introducing models, providing implementations where available, Chapter 4 is instead interested in highlighting the main challenges today posed by these research issues and generally addressing the future research on the se topics. The document finally includes Chapter 5 where conclusions are drawn.

3 Advances Achieved in the Development of Integrated IoT Enabling Technologies

3.1 Privacy and Security

This section gives a description of the enabling technologies related to security and privacy. Starting by a summary of the BUTLER Security Framework, it presents the results and issues of the different security technologies, prototypes and experimentations. In the last subsection, it gives a summary of the main issues related to the implementation of the low level security, application level security and the problem of the security bootstrapping. The Butler security framework enables end-to-end security between a data provider and a data consumer. The security protocols insure confidentiality, integrity of the messages and authentication of the peers. For privacy perspective, it is important to dissociate the security roles. For instance, it is important that data can only be received by allowed entities. All technical components (gateways, proxies...) used to transport data shall not have access to the data. This principle insures that the data cannot be retrieved and used by entities without user controls.

3.1.1 Summary of the BUTLER Security Framework

BUTLER has designed a Security Framework which is able to deal with major of IoT solutions and provide simple, lost cost and deployable security framework. The framework is described in details in Deliverable D5.1 - BUTLER Platforms and Pervasive Functionalities. The security framework is designed to provide end-to-end security from entity providing data (on actuating on data) and entity using data (or sending data to actuators). The first entities are called Resource and the second are Resource Consumers (or application).

The Security Framework assumes that the entities are able to communicate. The communication can be direct or indirect. For direct communication, the application accesses directly the resource thru the network. This kind of communication can be provided over IPV6 network or IPV4 network. The communication can use routers to distribute packets over the network; for direct communication routers do not store data locally. Indirect communications are the most used communication types. Sending entities push data to Service Platform, and Consuming entities retrieve data from this Service Platform.

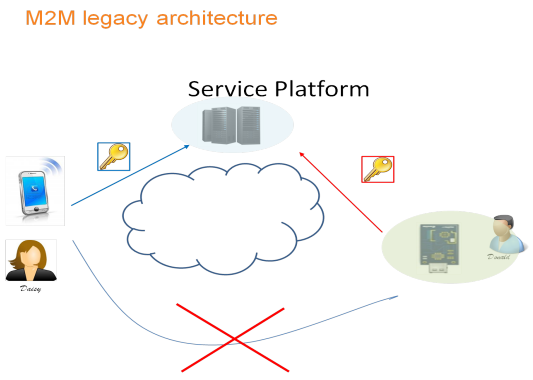


Figure 1: M2M Legacy Architecture.

The legacy M2M architecture is presented in Figure 1. In legacy M2M architecture, devices regularly sends data to Service Platform and Applications retrieve such data from the Service Platform. The two links can be secure point-to-point but there is no end-to-end security between the consuming application and the device. This architecture may pose a problem of privacy because the

data is kept in clear at Service Platform and may be used in fraudulent way or without control of user. In addition, the Service Platform must follow data storage and protection regulation rules which may higher the operational cost of the Service Platform. OneM2M standard organization is taking into account this architectural issue by specifying end-to-end security based on authorization mechanism. The work is in progress and Gemalto is strongly involved in OneM2M organization for pushing the authorization paradigm.

In BUTLER we have prototyped an IoT Security Framework supporting authorization paradigm and end-to-end security between devices and applications. The logical architecture is presented in Figure 2. The application must securely retrieve authorization token and security keys from an Authorization Server and applies the Security Protocol to retrieve data from the Service Platform. The Device also applies the Security Protocol to send data to the Service Platform for the application. Only the application can decrypt the encrypted data. The data is never kept in clear at the Service Platform and therefore the framework preserves data privacy. The point-to-point security between the device and the Service Platform can be implemented using different mechanisms described in subsection “Prototypes and Experimentations”.

The security protocol has been patented under the reference “System and method for securing machine-to-machine communications” EP13306900.5.

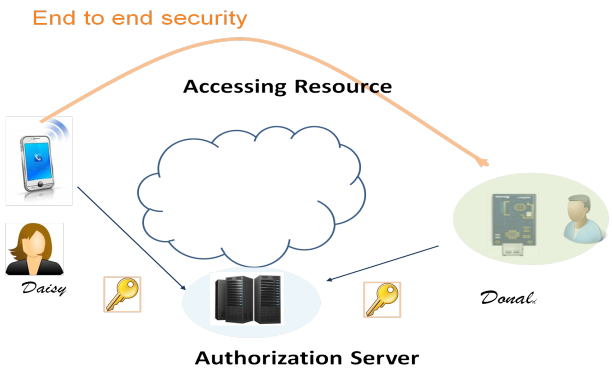


Figure 2: BUTLER Security – Nominal use case.

The main use case can be summarized as follow.

Resource - Registration.

1. The user is provided with a BUTLER compatible device. This device runs the BUTLER security framework and communication mechanisms allow its reachability on the network.
2. The user already has an account at the Trust Manager/Authorization Server for using the security and authorization mechanisms.
3. The user connects to the Authorization Server and registers a new resource. The resource consists of a Resource (unique) Identifier (generally an URL) and resource security credentials. At the resource, the user shall be able to configure (or retrieve) resource security credentials. The method for setting (or retrieving) the resource security credentials is one of the major issue to boostrap security. The section “Prototypes and Experimentations” studies different mechanisms to initialize the credentials.

   Once the registration is done, the user digitally owns the device.

Application - Registration.

Applications must be registered inside the Trust Manager. At application registration, the Trust manager generates application credentials which consist of application identifier and password. Using such data, application can authenticate to the Authorization Server to retrieve token to access a resource on behalf of an authenticated user.

Accessing the resource - Nominal use case.

Application authenticates to the Trust Manager and requests an access token and security materials for accessing a specific resource.The protocol may require user authentication. Application accesses the related resource by providing the access-token and using the security protocol. It must be noted the Security Protocol does not assume secure point-to-point communication between the application and the resource but provides its own security - for now - based on symmetric cryptography.

The application prepares the command; the command includes the following items:

1. the retrieved access-token
2. authentication-data. This data is computed by the application using the security material. Once receiving the authentication-data, the resource is able to verify that the application got the security material associated to the access-token.
3. encrypted and signed payload of the command. It is computed using the security material.

The application sends the access-token and the encrypted command to the resource [1]. The resource receives the sent items.

1. Using the embedded resource security material, the resource checks the signature of the access token and decrypts the access token
2. From decrypted access token, the resource gets the application authentication security material.
3. With the application authentication security material, the resource checks the signature of the authentication data and decrypts the authentication data.
4. The resource checks that the decrypted authentication data is consistent with the access token. If yes, the application is authenticated.
5. The resource authenticates to the Trust Manager for receiving the request/response encryption and signature keys of the command payload. For security reason, the key identifiers are provided by the application inside the authentication-data and not inside the acces-token.
6. The resource checks the signature of the command and decrypts the command.
7. The resource does its works and builds the response.
8. The resource encrypts and signs the response.
9. The application receives the signed and encrypted response, checks the signature and decrypts the response.

The protcol implements end-to-end security between the application and the resource. As said, the protocol is secure by itself and does not require secure transport protocol such as Transport Layer Security (TLS) or Datagram Transport Layer Security (DTLS). It implements authentication of the application to the resource, confidentiality and integrity of the command payload and implements anti-replay mechanism. If the protocol is transported over unsecure link, the access-token can be retrieved by unexpected entity that may analyze the request/response characteristics (source

[1] The protocol is self secure, therefore a secure transport protocol is not necessary at this level

address, destination address, data size, http header, ...) and deduce some characteristics of the peers. In consequence, it is more secure to transport the BUTLER protocol over secure link like TLS or DTLS. Anyway, it is not always possible to have a secure link from source (application) to destination (resource) because they are not always directly reachable on the network or because one entity cannot implement secure link due to device or network constraints.

In conclusion the BUTLER security protocol implements a good security mechanism for constrained devices taking into account network IoT characteristics.

ETSI – M2M architecture overview

To illustrate the underlying architecture of the ETSI – M2M standard, let's imagine that one device (device1) needs to exchange data with another device (device2) via a network Service Capability Layer (NSCL).

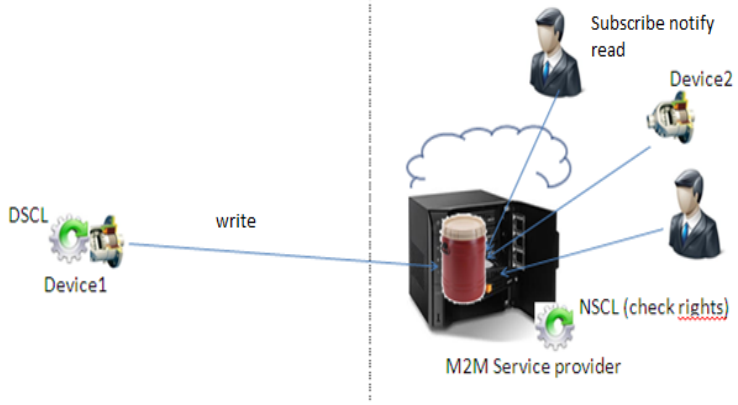


Figure 3: Illustration of M2M data exchange with ETSI M2M architecture

Typically a bucket resource will be created in the Service Capability Layer. Bucket resources have specific properties and can be managed using the resource management functionalities described by ETSI M2M. Access rights to this bucket will need to be acquired in read/write mode by the emitting device (device1), and possibly in read-only mode by the receiving device (device2). Both devices will probably first bootstrap their security with the M2M service provider and then connect to the M2M service as described above.

The receiving device (device2) may request to be notified whenever some data are transmitted by device1. In order to transmit data, device1 will issue a write request in the bucket resource. This will trigger the NSCL to send a notification to device2 which in turn will issue a read request in the bucket resources. This flow of operation is summarized in the below illustration. It can be seen that all functions involved can be provided in a restful way, while the process of setting up a data connection and using it for data exchange between devices is by nature a stateful process.

Leverage ETSI – M2M standard with BUTLER Security Framework.

The nominal use case assumes direct access to the resource. Anyway, in many IoT solutions compatible with ETSI M2M architecture, the resource is not directly reachable but can provide data to Service Platform for applications.

The end-to-end security is implemented as follows.

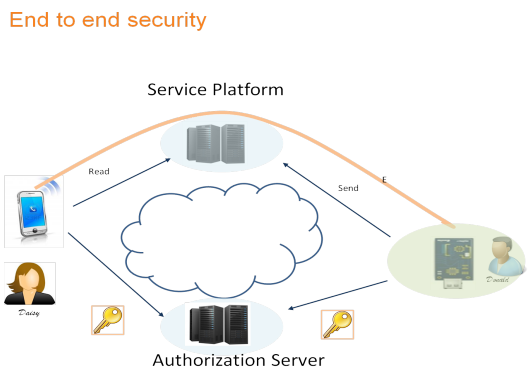


Figure 4: BUTLER end-to-end security

In this example, the Daisy's application exposes a resource which is able to use Donald's data. At the Trust Manager Daisy shall grant Donald the permission to access such Daisy's resource. Donald runs an application which shall retrieve an access token for accessing the Daisy's resource and shall push the access-token and the encrypted data to the Service Platform. Daisy runs an application which connect to the Service Platform, reads the access-token and the encrypted data. The Daisy's application checks the access-token, authenticates to the Trust Manager to get the security material associated to the access-token, decrypts the encrypted data and uses it. In this use case, Donald does not receive any feedback.

This way, we implemented end-to-end security between the Donald's application sending encrypted data for Daisy's resource.

At the Service Platform, the ETSI-M2M access right feature can be implemented using the Security Framework nominal use case. To implement it, the Service Platform exposes a storage resource (bucket resource in the ETSI parlance), allows Donald's application to write to the storage resource and the Daisy's application to read from the storage resource. Anyway, the point-to-point security for accessing the Service Platform can be implemented using other security paradigms up to the IoT solution - for instance using CoAP / DTLS or HTTPS. It is to be noted that the data are encrypted for the destination resource and cannot be used by the Service Platform, therefore the use case support end-to-end security and privacy.

BUTLER Security Framework - summary

The BUTLER Security Framework provides a simple security and privacy paradigm that can be implemented in low cost device. Using the Trust Manager, the user specifies its resources and manages the access permissions to its resources. The Trust Manager is not involved during the transfer of data between the resources and the applications and in consequence, business data management is the responsibility of the applications. The applications may use Service Platform to perform access to resources - for instance for resources that are not directly available on the Internet. The Security framework gives a way for applications to access resource (or act on resource) using a Service Platform and such Service Platform may be seen as (encrypted) Data Access Enabler and can provide other services like device management and let the applications managing the business data. In this sense, the Service Platform does not have access to clear data and cannot valuate clear business data. To do that, the Service Platform shall act as application and be allowed to use data on behalf of the user.

3.1.2 Prototypes and Experimentations

3.1.2.1 6LoWPAN Security

3.1.2.1.1 SIMULATION

The NARVAL toolbox has been updated in order to be operational in the new Scilab release 5.5.0 (new structures and functions). Network Analysis and Routing eVALuation, referenced as NARVAL has been designed on top of the Scilab environment (http://atoms.scilab.org/toolboxes/NARVAL). This Scilab External Module is focusing on the analysis of network protocols and algorithms. Each network of communicating devices such as computers, phones or sensors, needs to follows specific rules in order to organize and control the data exchange between source and destination nodes. Communication protocols enable to discover the network topology, and to propagate the data traffic between network entities. The main goal of our toolbox is to provide a complete software environment enabling the understanding of available communication algorithms, but also the design of new schemes in order to evaluate and improve the traffic behavior and distribution on network topologies defined by the user. NARVAL permits to generate random topologies according to various algorithms such as Locality, Waxman, Barabasi-Albert and hierarchical models. The user can also design his own topology by providing nodes' coordinates, visualization parameters, and also links' information that are necessary for path calculation. The combination of these functions enables to build a large range of topologies with distinct routing properties. The NARVAL module permits to study the impact of routing algorithms on the effectiveness of transmission protocols used by data communications on a defined network topology. We provide a set of basic functions in order to create network graphs, compute routing algorithms (AODV, BFS, DFS, Bellman-Ford, Dijkstra, Flood, Floyd-Warshall, Multiple Paths, RPL, ARC, etc.) on them and finally make statistical analysis on the efficiency of data communications. The mobility of nodes (Mobile/Vehicular Ad hoc NETwork MANET/ VANET) is also supported according to models such as Random Direction, Random Walk, Random Way Point, etc. The target audience of this external module includes academics, students, engineers and scientists. The description of each function has been carefully done in order to facilitate the end users' comprehension. It is often accompanied with explicit diagrams.

New functions related to Security had been designed during the Butler project:

- NL_S_AESAddRoundKey - Addition (element to element in F256) between a state matrix and a round key.
- NL_S_AESDecryption - Perform the AES decryption.
- NL_S_AESEncryption - Perform the AES encryption.
- NL_S_AESInitialization - Perform the initialization structure of the AES algorithm from a cryptographic key.
- NL_S_AESInitializationM - Perform the vectors Sbox, InvSbox, ExpoToPoly and PolyToExpo used during the initialization of the AES algorithm.
- NL_S_AESKeyExpansion - Perform the key Expansion.
- NL_S_AESMixColumns - Mix columns of a matrix.
- NL_S_AESMixColumnsR - Mix columns of a matrix (reverse).
- NL_S_AESPolynomialMult - Perform the polynomial multiplication in a finite field.
- NL_S_AESShiftRows - Shift each row of a 4x4 matrix to the left.
- NL_S_AESShiftRowsR - Shift each row of a 4x4 matrix to the right.
- NL_S_AESSubBytes - Substitution of a matrix state in respect with a table Sbox.
- NL_S_AESSubBytesR - Substitution of a matrix state in respect with a table InvSbox.
- NL_S_RSADecryption - Perform the decryption of a message in respect with the RSA scheme.

- NL_S_RSAEncryption - Perform the encryption of a message in respect with the RSA scheme.
- NL_S_RSAKeys - Perform the public and private keys of the RSA algorithm.
- NL_S_RSAKeysE - Perform the public and private keys of the RSA algorithm (additional input E).

The paper entitled “Network Analysis and Routing eVALuation, the Scilab Module Dedicated to the Analysis of Network Protocols and Algorithms” has been accepted at the 6th International Scilab Users Conference (ScilabTec’14). RPL (6LoWPAN) has been implemented and simulated on top of the NARVAL module. A new routing paradigm entitled ARC has been also implemented in the simulation environment. The ARC paradigm is based on the innovative concept of routing construct made of a sequence of nodes and links with 2 outgoing edges. As a consequence, each node can still reach one of the outgoing edges upon a single breakage. This permits to improve the network utilization in respect with the fault-tolerance intrinsic properties and also the fast re-routing, load balancing and multiple-paths features. A new paper entitled “Towards a New Way of Reliable Routing: Multiple Paths over ARCs” has been submitted to the 3rd Conference on Advance Computing, Communication and Informatics ICACCI’14. Traditional routing in data networks is based on a forwarding scheme where a path, composed by an ordered list of network nodes is performed for each connection between a source and a destination. Thus if a simple failure appears on this path, the route becomes useless. In fact the path needs to be computed again. This process adds delays in order to locally update the network routing tables impacted by the node or link breakage. Unfortunately a single failure on any new path needs also to trigger a re-routing calculation process. We present in this work a new routing algorithm named Available Routing Construct (ARC). The main advantage of this two-edged routing construct relies on the fast recovery feature in case of network failures. For instance alternative routes can be retrieved in the case of local traffic congestion. Smart routing enabling multipath and load balancing is also natively supported. ARC is based on the innovative concept of routing construct made of a sequence of nodes and links with 2 outgoing edges. As a consequence, each node can still reach one of the outgoing edges upon a single breakage. Thus each ARC provides its own independent domain of fault isolation and recovery as an ARC topology is resilient to one breakage per ARC. ARC can significantly improve the network utilization in respect with its fault-tolerance intrinsic properties and also its fast re-routing, load balancing and multiple- paths features. This work is based on a theoretical modeling supported by simulations developed on top of the Scilab environment.

3.1.2.1.2 TEST of Security Solutions (6LoWPAN)

Several security schemes had been tested for IoT devices running on TinyOS (e.g. TinySec, AES Encryption of CC2420, MiniSec, Relic and TinyECC) or Contiki (e.g. ContikiSec, Contiki-TLS-DTLS, Contiki IPsec, CoAPs: COAP over DTLS/TLS). UL improved its Coap testbed within its IoT-lab (TelosB/sky working now on Tinyos). CoAPs (CoAP over DTLS/TLS) had been released for the Contiki OS and ported for the Arago WiSMote sensors. Due to resource constraints, CoAPs cannot be applied to TelosB sensors. In fact the source code does not compile for TelosB/sky motes, due to memory requirements. However low level security can be ensured at the link layer in respect with the Advanced Encryption Standard (AES) scheme. Pre-shared keys are used. A simple Graphical User Interface (web) based on Json and Google Chart APIs has been designed. This small wireless Sensor Network (proof-of-concept) is dedicated to the real-time monitoring of temperature and humidity in an office environment. One computer is reachable with a public IPv6 address. A border router is used as “802.15.4/IPv6” gateway. 3 sensor nodes are already deployed and attached to the border router. The IPv6 connectivity had been set up. The nodes are accessible from the Internet through IPv6 connections. In this scenario, the sensed information is collected on a database in order to enable on/off-line data processing. Two new types of sensors had been added in the IoT-Lab (Libelium Smart Environment: Waspmote Plug & sense with temperature,

humidity, CO2, NO2, O2 and CO connected through WiFi, and equipped with GPS and solar panel, and Watteco SmartPlugs.

Hardware Platform and Operating Systems used for Tests

802.15.4 TelosB mote modules manufactured by the Spanish manufacturer Advanticsys have been used during the real Wireless Sensor Network Deployment. Three CM5000 motes are available (motelist):

- Mote 1: Reference MFV69HQG Description FTDI MTM-CM5000MSP
- Mote 2: Reference MFV6KBXA Description FTDI MTM-CM5000MSP
- Mote 3: Reference MFV69BLV Description FTDI MTM-CM5000MSP

The CM5000 TelosB sensor is an IEEE 802.15.4 compliant wireless sensor node based on the original open-source TelosB/Tmote Sky platform design developed and published by the University of California, Berkeley. The included sensors measure temperature, relative humidity and light. Each node is equipped with a TI MSP430F1611 Microcontroller, a CC2420 RF Chip, User & Reset Buttons, 3 Leds, an USB Interface and 2xAA Batteries. CM5000 motes support TinyOS and Contiki. Advanticsys provides a Vmware Image in order to start working with TinyOS. Instant Contiki is an entire Contiki development environment, based on an Ubuntu Linux virtual machine that runs in VMWare player or VirtualBox, and has Contiki and all the development tools, compilers, and simulators used in Contiki development already installed. There exist many alternative solutions that provide security during the deployment and operation of embedded devices within a wireless sensor network. We describe in the next section some of them, based on TinyOS and Contiki.

AES in TinyOS

Security can be ensured at the link layer in respect with the Advanced Encryption Standard (AES) scheme. In this case, a single hop security is provided. A basic example can be tested according to the CoapBlip application, already included inside the TinyOS environment. Two motes are needed: one server and one Point-to-Point Protocol (PPP) router.

User can modify the list of available resources of the server mote inside the Makefile, such as:

- DCOAP_RESOURCE_TEMP,
- DCOAP_RESOURCE_HUM,
- DCOAP_RESOURCE_VOLT,
- DCOAP_RESOURCE_LED.

He can also change the IPv6 prefix DIN6_PREFIX. The mote is assumed connected to the port /dev/ttyUSB0. The motelist command permits to know where each sensor is connected (USB0, USB1, etc.). The compilation is done in respect with the following command, where X is the address:

```
make telosb blip coap install, X bsl,/dev/ttyUSB0
```

The PPP router is assumed to be connected to the port /dev/ttyUSB1. Its set up is done according to the following command:

```
make telosb blip install bsl, /dev/ttyUSB1
```

Then the user needs to start the driver of the Ppp Connection with the following command:

```
sudo pppd debug passive noauth nodetach 115200 /dev/ttyUSB1 nocrtscts
nocdtrcts lcp-echo-interval 0 noccp noip ipv6 ::23,::24
```

Then the PPP connection is established with the command (the right IN6_PREFIX must be used):

```
sudo ifconfig ppp0 add fec0::100/64
```

Then the user can request (GET/PUT) a resource (Voltage:sv, Temperature:st, Humidity:sh, Leds:l, AES key:ck) from the server mote with the coap-client application:

```
./coap-client -m <CMD> coap://[fec0::X]:61616/<URI> -t binary
```

For instance, the following command permits to actuate on the LEDs of the server mote:

```
echo -e -n \\x02 | ./coap-client -m put coap://[fec0::3]:61616/l -T 3a -t binary -f -
```

For a CM5000 mote, here is the matching table between LED values and colours:

- x00: no LED
- x01: Red
- x02: Yellow
- x03: Yellow and Red
- x04: Blue
- x05: Blue and Red
- x06: Blue and Yellow
- x07: Blue, Red and Yellow

Wireshark permits to capture the data traffic of the ppp0 connection. When a resource request is launched, two packets are collected:

```
 \x42\x01\xA8\x65\x11\x2A\x82\x73\x74
send to [fec0::2]:61616:
  pdu (9 bytes) v:1 t:0 oc:2 c:1 id:43109 o: 1:'*' 9:'st'
Jul 24 13:36:13 ** received from [fec0::2]:61616:
  pdu (8 bytes) v:1 t:2 oc:1 c:80 id:43109 o: 1:'*'
  data:'\xF5w'
Jul 24 13:36:13 *** removed transaction 43109
** process pdu: pdu (8 bytes) v:1 t:2 oc:1 c:80 id:43109 o: 1:'*'
  data:'\xF5w'
**.Temperatur: 307.09 K
```

The user can set up a cryptographic key composed by 25 values used by the AES encryption:

```
Echo -e -n \\xFF\\x01\\x02\\x03\\x04\\x05\\x06\\x07\\x08\\x01\\x02\\x03\\x04
\\x05\\x06\\x07\\x08\\x09\\x10\\x11\\x12\\x13\\x14\\x15\\x16 >> key
./coap-client - m put coap://[fec0::2]:61616/ck -T 3a -t binary -f key
```

TinyECC2.0 in TinyOS

The development of Elliptic Curve Cryptography (ECC) has proven that Public Key Cryptography (PKC) can be used in WSNs. ECC, proposed independently by Neal Koblitz [2] and Victor Miller [3], is based on the algebraic structure of elliptic curves over finite fields. The harder it is to solve a mathematical problem, the more secure the algorithm. For instance, the Rivest Shamir Adleman (RSA) algorithm is based on the Integer Factorization problem that has a sub-exponential solution. ECC is based on the Elliptic Curve Discrete Logarithmic Problem (ECDLP) where the solution is fully exponential. As a consequence, ECC can offer the same level of security than RSA with much smaller key sizes. A 160-bit ECC key provides the same level of security as a 1024-bit RSA key, and 224-bit ECC is equivalent to the 2048-bit RSA. Smaller keys result in faster computations, less memory, power and bandwidth consumption. Using ECC results in approximately 1.5 times smaller memory consumption on the motes and in 4 to 5 times faster cryptography operations when compared to using RSA [3]. TinyECC is a portable and efficient library developed at the North Carolina State University. It has been released by A Liu et al. [4]. It provides a digital signature scheme, named Elliptic Curve Digital Signature Algorithm (ECDSA), a key exchange protocol named Elliptic Curve Diffie Hellman (ECDH) and a public key encryption scheme (ECIES). The last release of TinyECC2.0 supports different platforms such as MICA2/MICAz, TelosB/Tmote Sky, BSNV3 and Imote2 motes. The user must install TinyOS 2.1.1 or a later version. He/She can also use a virtual machine where the TinyOS is already set up, such as the one provided by the Advanticsys manufacturer. The user should extract TinyECC-2.0.zip to the following location:

```
sudo unzip TinyECC2.0.zip -d /opt/tinyos-2.x/apps/TinyECC-2.0.zip
```

Here is the list of interfaces provided by TinyECC2.0:

- NN.nc defines the interface NN implemented by NNM.nc, related to big natural number operations.
- ECC.nc defines the interface ECC implemented by ECCM.nc. It provides the basic and enhanced (sliding window method and projective coordinate system) elliptic curve operations.
- ECDSA.nc defines the interface ECDSA (ECDSAM.nc), which provides the ECDSA signature generation and verification.
- ECIES.nc defines the interface ECIES (ECIESM.nc), which performs the ECIES encryption and decryption.
- ECDH.nc defines the interface ECDH (ECDHM.nc), which computes the ECDH key establishment.
- SHA1.nc defines the interface SHA1 (SHA1M.nc), which provides the SHA-1 functions.
- CurveParam.nc defines the interface CurveParam, which permits to get the parameters of elliptic curves and to optimize multiplication with omega. secp128*.nc, secp160*.nc, secp192*.nc implement this interface in order to provide parameters for Standards for Efficient Cryptography Group (SECG) defined elliptic curves. Curve name needs to be defined in the makefile to select the elliptic curve parameters.

TinyECC is written in NesC. TinyECC is a ready-to-use software package enabling ECC Public Key Cryptography (PKC) operations. It includes optimizations for ECC operations. A. Liu et al. compared the execution time, ROM/RAM and energy consumptions for different platforms [4]. TinyECC supports ECC schemes such as ECDSA, ECDH and ECIES, defined in the Standards for Efficient Cryptography [5], and elliptic curve parameters such as secp160k1, secp160r1 and secp160r2 [6]. ECDH (respectively ECDSA) is a variant of the Diffie-Hellman key agreement protocol (respectively Digital Signature Algorithm). ECIES, also known as Elliptic Curve Augmented Encryption Scheme (ECAES), supports semantic security. Optimizations for ECC include Barrett Reduction, Hybrid Multiplication, Hybrid Squaring, Projective Coordinate System, Sliding Window Method, Shamir's Trick and Curve-Specific Optimization [4]. TinyECC supports 128-bit, 160-bit and 192-bit ECC parameters. We remind that 160-bit ECC parameters provide the same security level as 1024-bit RSA. In their evaluation, A. Liu et al. selected secp160r1 to evaluate each optimization technique. Enabling all optimizations requires long pre-computation and the largest ROM and RAM consumptions.

ECDSA: testECDSA.nc and testECDSAM.nc are related to ECDSA and permits to measure its execution time. As we are using TelosB/Tmote Sky mote, the installation is done as what follows. User must change the COMPONENT inside the Makefile of the folder TinyECC-2.0 with the correct value (in this case testECDSA). It is possible to switch between seven curve parameters, i.e. DSECP128R1, DSECP128R2, DSECP160K1, DSECP160R1, DSECP160R2, DSECP192K1, DSECP192R1. The application is then compiled and uploaded into a mote connected through a USB port.

```
make telosb install
```

The serial forwarder can be initialized with the following command (see Figure 5):

```
java net.tinyos.sf.SerialForwarder -comm serial@/dev/ttyUSB0:telosb &
```

The user can start the application by typing the following command from a shell:

```
cd /opt/tiny/opt/tinyos-2.1.1/apps/TineECC-2.0/
java show_ecdsa
```

The output is directly displayed on the console. Many rounds are generated and provide the following information:

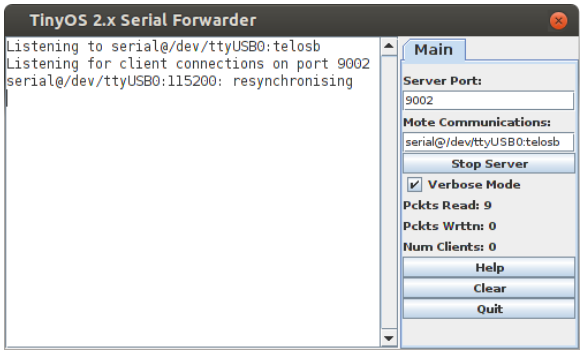


Figure 5: TinyOS Serial Forwarder

- the private key d
- the time of ECC.init() ecci
- the public key x and y
- the time of public key generation pkg
- the time of ECDSA.init() ecdsai
- the message msg
- the message signature r and s
- the time of signature generation sg
- the time of signature verification sv

The average timing result is also computed after each new round.

```
-------------------- round 1 ---------------------
Private key:
d: 58a9a30d5937a46a5a8d5780483bf86e533ad08
[ time of ECC.init() is 2.462 sec ]
Public key:
x: b23aecbc4a85845975f1c9dd3776aa07d0e0d728
y: 97544c3f505ce484beb29c851432abf85efcb138
[ time of public key generation is 2.631 sec ]
[ time of ECDSA.init() is 4.76 sec ]
content and signature
msg: bd65d5b57ce2dba944973174f7f1f4f7f8efc19d2c42963e63d9ad459c2751b4
72fbe0dfa85aa653b964d7b87aebc98d0c021b20
signature
r: fc0b8add4d9a036e18c04cd872cd3297bf4b1ed6
s: 0220bb36907191b99db37c53721b6343e8495535
[ time of signature generation is 2.832 sec ]
[ time of signature verification is 3.665 sec ] (pass)
Average timing result
ECC.init(): 2.4635
ECDSA.init(): 4.754
public key gen: 2.6605
sign: 2.8705
verify: 5.491
```

We performed 100 runs for each curve parameter. Statistics (average time in second) are provided in the following table:

The execution time increases with the length of the key.

ECIES: testECIES.nc and testECIESM.nc are related to ECIES and permits to measure its execution time. As we are using TelosB/Tmote Sky mote, the installation is done as what follows. User must change the COMPONENT inside the Makefile of the folder TinyECC-2.0 with the correct value (in this case testECIES). It is possible to switch between seven curve parameters, i.e. DSECP128R1, DSECP128R2, DSECP160K1, DSECP160R1, DSECP160R2, DSECP192K1, DSECP192R1. The application is then compiled and uploaded into a mote connected through a USB port.

Table 1: ECDSA.

ECDSA	ecci	pkg	ecdsai	sg	sv
DSECP128R1	1.7947	2.7943	3.4575	2.874	3.6845
DSECP128R2	1.7913	3.0516	3.4553	3.0996	4.0409
DSECP160R1	2.4613	2.6855	4.7523	2.8799	3.6717
DSECP160R2	2.4364	2.7759	4.7258	2.9456	3.7664
DSECP192K1	3.4083	4.1254	6.5745	4.3815	5.2008
DSECP192R1	3.5185	4.0747	6.4897	4.3352	5.4671

```
make telosb install
```

The serial forwarder can be initialized with the following command:

```
java net.tinyos.sf.SerialForwarder -comm serial@/dev/ttyUSB0:telosb &
```

The user can start the application by typing the following command from a shell:

```
cd /opt/tiny/opt/tinyos-2.1.1/apps/TineECC-2.0/
java show_ecies
```

The output is directly displayed on the console. Many rounds are generated and provide the following information:

- the plaintext message m
- the time of ECIES.init() eciesi
- the private key d
- the public key x and y
- the time of public key generation pkg
- the time of ECIES.encrypt eciese
- the ciphertext message c
- the time of ECIES.decrypt eciesd

The average timing result is also computed after each new round.

```
6bc0aa4b890d0c02163379edcc821b2944923b69cd851c2751bd65d5b57ce2dba9449
73174f7f1f4
[ time of ECIES.init() is 2.471 sec ]
Private key:
d: 3bd90d63163e9396d140fb63e775b1ed5679dea1
Public key:
x: e727d37663ef9b77fdfeba4ebee979d0c6b89ad
y: f2735e2aee45d84fd2bada7360de6a174f1bc2ca
[ time of public key generation is 2.719 sec ]
0210da4922e2993ea55344a7321c42eacd11a01e107332b4dbc397ddf9191941e46bf
a6745d46f7f3c69660cd6eb1c728f14d59b48c6751b882ab5e344565024eef4ed0dd7
b80ae025b4854c72b687b471
[ time of ECIES.encrypt() is 6.005 sec ]
6bc0aa4b890d0c02163379edcc821b2944923b69cd851c2751bd65d5b57ce2dba944973174f7f1f4
[ time of ECIES.decrypt() is 3.945 sec ]
Average timing result for 1 rounds
ECIES.init(): 2.471
public key gen: 2.719
encrypt: 11.919001
decrypt: 7.8929996
```

Table 2: ECIES.

ECIES	eciesi	pkg	eciese	eciesd
DSECP128R1	1.8021	2.7933	6.1389	3.7353
DSECP128R2	1.7938	2.9911	6.5543	3.9329
DSECP160R1	2.4554	2.6841	5.92	3.9022
DSECP160R2	2.4336	2.7727	6.0931	4.0463
DSECP192K1	3.4171	4.1437	9.0265	6.0687
DSECP192R1	3.4041	4.0614	8.8462	5.8449

We performed 100 runs for each curve parameter. Statistics (average time in second) are provided in the following table: The execution time increases with the length of the key.

ECDH: testECDH.nc and testECDHM.nc are related to ECDH and permits to measure its execution time. As we are using TelosB/Tmote Sky mote, the installation is done as what follows. User must change the COMPONENT inside the Makefile of the folder TinyECC-2.0 with the correct value (in this case testECDH). It is possible to switch between seven curve parameters, i.e. DSECP128R1, DSECP128R2, DSECP160K1, DSECP160R1, DSECP160R2, DSECP192K1, DSECP192R1. The application is then compiled and uploaded into a mote connected through a USB port.

```
make telosb install
```

The serial forwarder can be initialized with the following command:

```
java net.tinyos.sf.SerialForwarder -comm serial@/dev/ttyUSB0:telosb &
```

The user can start the application by typing the following command from a shell:

```
cd /opt/tiny/opt/tinyos-2.1.1/apps/TineECC-2.0/
java show_ecdh
```

The output is directly displayed on the console. Many rounds are generated and provide the following information:

- the time of ECDH.init ecdhi
- the private key1 d1
- the public key1 x1 and y1
- the time of public key1 generation pkg1
- the private key2 d2
- the public key2 x2 and y2
- the time of public key2 generation pkg2
- the established key1 ek1
- the time of ECDH.key_agree(1) ecdhka1
- the established key2 ek2
- the time of ECDH.key_agree(2) ecdhka2

```
[ time of EDH.init() is 2.462 sec ]
Private key1:
76f12bf705749a315d95497b8e81a4876ead61c5
Public key1:
x: f6c6d38995b21ef32826c30fe94bd2977fc85d2
y: a2bf24aaf788a969724cdf5436e5a56b1e652423
[ time of public key 1 generation is 2.74 sec ]
Private key2:
53ad39590c23169e93c4dca1f6f931c7964d8695
Public key2:
```

```
x: 4c84d0d86fdc064dfea10cbac47d833b7f149809
y: dd7537d9dda79514c40948c7dcb8b7acf0cb0a15
[ time of public key 2 generation is 2.731 sec ]
established key1: 011294f5a4dbdbda23380c6e9ec1e48e84356200
[ time of ECDH.key_agree() for 1 is 3.198 sec ]
established key2: 011294f5a4dbdbda23380c6e9ec1e48e84356200
[ time of ECDH.key_agree() for 2 is 3.21 sec ]
Average timing result for 1 rounds
ECDH.init(): 2.462
PK1: 2.74
PK2: 2.731
key_agree1: 3.198
key_agree2: 3.21
```

We performed 100 runs for each curve parameter. Statistics (average time in second) are provided in the following table: The execution time increases with the length of the key.

Table 3: ECDH.

ECDH	ecdhi	pkg1	pkg2	ecdhka1	ecdhka2
SECP128R1	1.7962	2.7843	2.7756	3.355	3.3412
SECP128R2	1.7958	3.0017	3.0108	3.5633	3.5713
SECP160R1	2.4627	2.7106	2.7003	3.1826	3.1847
SECP160R2	2.4353	2.7704	2.778	3.259	3.282
SECP192K1	3.41	4.1236	4.1328	4.8097	4.8212
SECP192R1	3.4065	4.062	4.0707	4.7305	4.7389

IPSec in Contiki

IPsec (Internet Protocol Security) operates at the network layer. It is based on two protocols, i.e. the Authentication Header (AH) and the Encapsulating Security Protocol (ESP). AH provides integrity and data origin authentication with optional anti-reply features while ESP also offers confidentiality. Two modes are available, i.e. transport and tunnel. In the transport mode, the IPsec header is inserted after the IP header. Protection is primarily provided for next layer protocols. In fact only the payload of the original IP packet is encrypted and/or authenticated. In the tunnel mode, a complete IP packet is encrypted and/or authenticated. Then it is encapsulated into a new IP packet with a new IP header. A Security Association (SA) is an association between two IPsec endpoints. The flow of information in one direction is protected. As a consequence, two SAs are needed for a bidirectional communication. SAs can be established and maintained manually. However the Internet Key Exchange protocol (IKE or IKEv2) is used to automatically do this. Mutual authentication is performed between two parties. The IKE security association is used to efficiently establish SAs for ESP or AH. All IKE communications consist of pairs of messages, called “exchange”. The IKE SA INIT exchange negotiates security parameters (nonces and Diffie-Hellman) for the IKE SA. The IKE AUTH exchange transmits identities, authenticates them and initializes the SA. IKE is responsible for key management needed in all encryption and authentication operations based on keys. IPSec for 6LoWPAN has been released by Dr Shahid Raza from the Swedish Institute of Computer Science (SICS) [7]. The source code is available here:

```
svn co https://contikiprojects.svn.sourceforge.net/svnroot/contikiprojects/sics.se/ipsec ipsec
```

The set-up is based on 2 motes connected through USB ports to a computer running Instant contiki 2.6. The first mote M1 is used as a border router.

```
M1: MAC 00:12:74:00:10:21:b4:d8
aaaa::212:7400:1021:b4d8
```

The second mote is used as Application Example.

```
M2: MAC 00:12:74:00:10:f4:10:3a
aaaa::212:7400:10f4:103a
```

IPsec is first configured inside the Ubuntu virtual machine (Instant contiki). The package ipsec-tools must be installed.

```
sudo apt-get install ipsec-tools
```

The file ipsec-tools.conf must be edited in order to specify SAs. The IPv6 address of the Mote 2 (Application Example) is aaaa::212:7400:10f4:103a. In order to start the service, the following command should be run:

```
sudo chmod 750 ipsec-tools.conf #(conf file not readable to the world)
sudo /etc/init.d/setkey start
```

```
user@instant-contiki:/etc sudo /etc/init.d/setkey start
* Loading IPsec SA/SP database:
[ OK ]
```

For the part related to the configuration of IPsec in the Contiki mote, it is mentioned that, the user should set in project.conf the selected modes of operation. The key are stored in ipsec/aes-ctr.c and ipsec/aes-xcbc-mac.c. The key is the same than the one used during the IPSec configuration in Ubuntu (0xcf5faaca70ee5ec4c8f43158a45c0363):

```
{0xcf,0x5f,0xaa,0xca,0x70,0xee,0x5e,0xc4,0xc8,0xf4,0x31,0x58,0xa4,0x5c,0x03,0x63}
```

The nonce is 0x69,0xbb,0xc0,0xc9. This IPsec solution uses pre-shared keys in order to establish SAs. In other words, the key management scheme of the IPsec suite, namely the IKEv2 protocol has not been done. S. Raza et al described in [8] their on-going research on Lightweight IKEv2. Finally here is the usage with the Tmote sky from Linux. The border router M1 is compiled with the following commands:

```
cd contiki-2.6/examples/ipv6/rpl-border-router/
sudo make TARGET=sky border-router.upload MOTE=1
```

The border router is initialized with the command:

```
make connect-router
```

The application is now installed on the second mote M2 with the following command:

```
cd contiki-2.6/examples/ipv6/ipsec/
sudo make ipsec-example.upload MOTE=2
```

The user should finally execute the commands to test the application:

```
cd ~/contiki-2.6/examples/ipv6/ipsec/scripts
./client.py aaaa::0212:7400:13b7:7f30
```

We made some experiments with AH, ESP, and AH+ESP. Data are composed by a succession of 32 'a'. The client application provides the following result:

```
./client.py aaaa::0212:7400:13b7:7f30
Duration: 165 ms
Request: aaaaaaaaaaaaaaaaaaaaaaaaaaaaaaaa
Response: bbbbbbbbbbbbbbbbbbbbbbbbbbbbbbbb
```

The compressed version of IPsec has been implemented for the Contiki OS. For the AH Authentication, the mode HMAC-SHA1-96 needs 24 Bytes (possibly reduced to 16 Bytes for compressed IPsec). In comparison, 802.15.4 in AES-CBC-MAC-96 mode needs 12 Bytes. For the ESP Encryption, AES-CBC needs 18 Bytes (possibly reduced to 12 Bytes for compressed IPsec). In comparison, 802.15.4 in AES-CTR mode needs 5 Bytes. For the ESP Encryption and Authentication, the mode AES-CBC and HMAC-SHA1-96 needs 30 Bytes (possibly reduced to 24 Bytes). In comparison, 802.15.4 in AES-CCM-128 needs 21 Bytes.

TinyDTLS in Contiki

O. Bergmann designed a specific library enabling a datagram server with DTLS support [9]. The current release is tinydtls-0.4.0 (July 2013). It also provides an application example. The source files are available here: (http://sourceforge.net/projects/tinydtls/files/). We installed tinydtls on the Instant Contiki 2.6 virtual machine. The set-up is done as what follows. The source file should be uncompressed in a defined location. The configuration is done with the command:

```
./configure
```

Thereafter we perform the installation.

```
make
sudo make install
```

User just needs to add tinydtls to the variable APPS in the Makefile.

3.1.2.1.3 Internet Key Exchange Mechanism for IPSec in 6LoWPAN

This subsection describes about an experimental prototype implementation of Internet Key Exchange (IKEv2) mechanism for IPSec in 6LoWPAN. Previously, IPSec (Internet Protocol Security) as the security option for 6LoWPAN based smart objects was described in [10], it includes a brief description of IPSec and the IKEv2 (Internet Key Exchange version 2) protocol's key exchange procedure. This following describes a brief description and requirements of the performed experiments.

IKEv2 implementation:
A partial implementation of IKEv2 protocol [11] was developed under the EU FP7 funded project 'CALIPSO' [12] is available open source. Some important software and hardware requirements are the following,

- A test platform with ample resources such as the RAM, Flash memory, etc. One of them is the WISMOTE platform and it's shown in Figure 6.
  - This IPsec patch must be compiled using the MSP430X instruction set (20-bit memory instructions) as the memory space provided by the 16-bit MSP430 is not enough, therefore it is recommended to build using IAR compiler or MSP430-GCC compiler of version 4.7.2 or greater (earlier versions have memory leaks).
- The Linux system should support IPSec implementation within its kernel. Strongswan [13], one of the most popular IPSec implementation for Ubuntu based linux system is preferred. Strongswan has IKEv2 keying daemon named Charon, a keying daemon was built from scratch to implement the IKEv2 protocol for Strongswan. The BR node acts as a middle man to exchange messages from node to the Internet through the Linux system.

More information about the features that has not been implemented in this IPSec patch for Contiki can be found in README.md [11]. The author has implemented only the ESP encryption mode of IPSec.

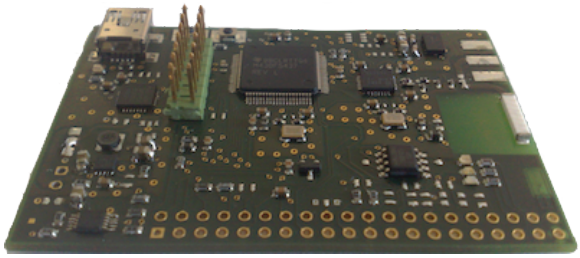


Figure 6: WISMOTE platform

Contiki, the open source operating system for 6LoWPAN based devices provide a simulation tool named Cooja. Also, it provides a native platform to test applications with Cooja. This IPSec implementation [11] has been tested with Cooja. The test set-up is shown with a screen shot from cooja simulator in Figure 7 which consists of two nodes,

- Node 1: a border router application with IPv6 address aaaa::1
- Node 2: an IPSec enabled simple application node with IPv6 address aaaa::2

Before starting the simulation in cooja, the following command has to be executed in the linux terminal:

```
make TARGET=cooja connect-router-cooja
```

The above contiki command creates a virtual TUN interface with the BR, thus provides message exchange between the linux machine and the cooja simulated environment, therefore enabling the secure IPSec end-to-end connection between the two nodes.

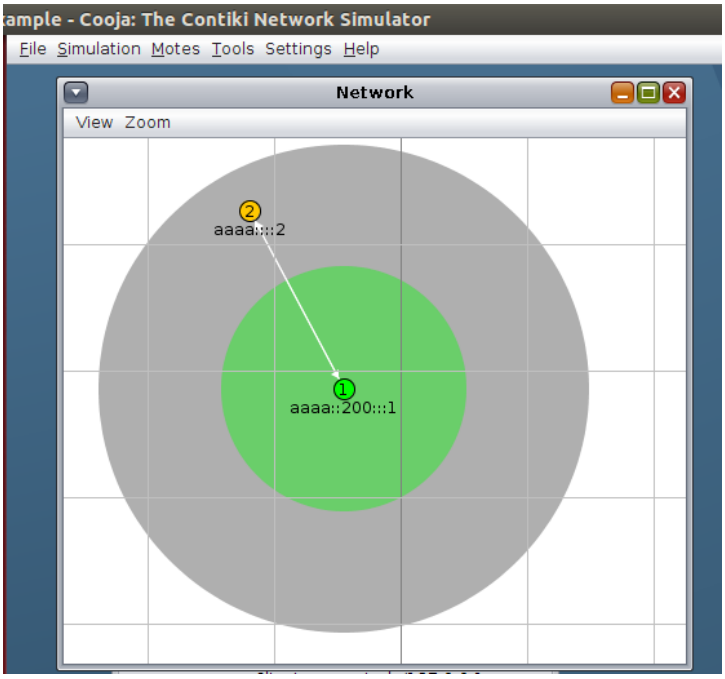


Figure 7: cooja simulator

The security parameters between two nodes are exchanged when one of the node initiates communication with another node. For example, a ping request from BR node1 (aaaa::1) to IPSec application node2 (aaaa::2) can be initiated by using the ping6 linux command or by enabling automated request from the sensor node. Following the request, an IKEv2 key exchange mechanism is initiated, Figure 8 represents the wireshark capture on the virtual TUN interface created by BR application with a successful IKEv2 key exchange with 4 messages being exchanged in an appropriate order was shown in Figure 8. Later, the application specific data are exchanged as encrypted security payload (ESP) packets.

3.1.2.2 SmartObject Security based on ZigBee

3.1.2.2.1 Security in the ZigBee stack protocol

The ZigBee stack is built upon the PHY / MAC layer of the IEEE 802.15.4 standard. It uses the IEEE 802.15.4 model where the Network Layer has direct access to the MAC. The ZigBee Network (NWK) layer provides network topology management, MAC management, routing and discovery protocol. Security management tools are provided both by the ZigBee Network layer and the Application sub-layer. The ZigBee network can overlay its security service to the additional security service

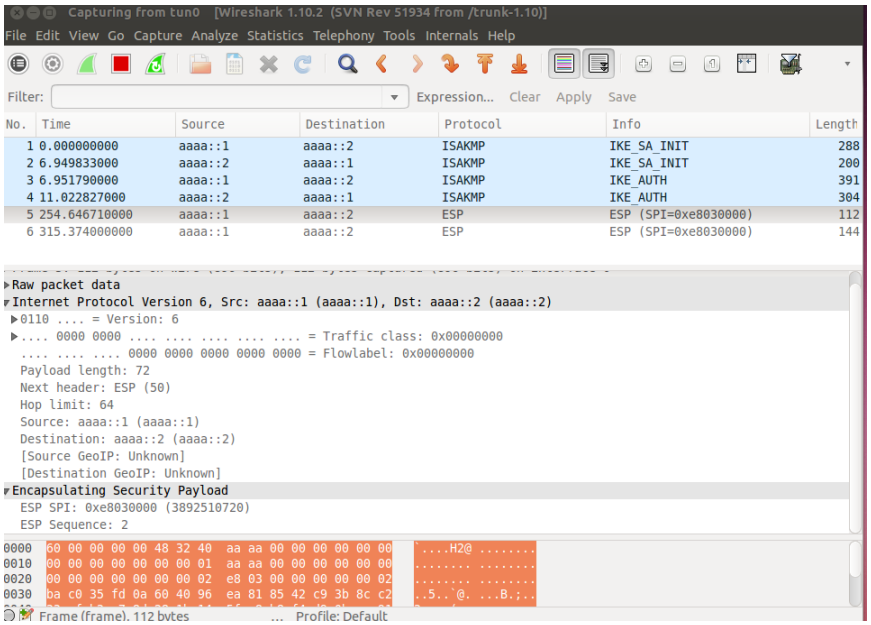


Figure 8: Successful IKEv2 parameters exchange with cooja

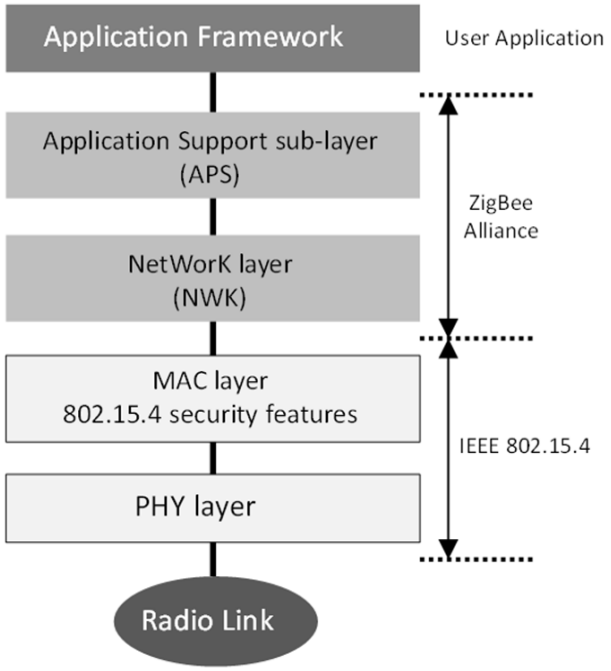


Figure 9: No name one

available in the MAC layer. The user application builds on the ZigBee stack through its Application Interface (API).

ZigBee uses three keys at different level to manage the security:

- The **master key** is the deep secret key of the nodes. It can be used as an initial secret key shared between two peers and pre-deployed via an out-of-band channel,
- The **network key** ensures the security at the network level and is shared by all the nodes of the network,
- The **link key** is optional. This key can be deduced from the master key and ensures the security of the link between two peers at the application level.

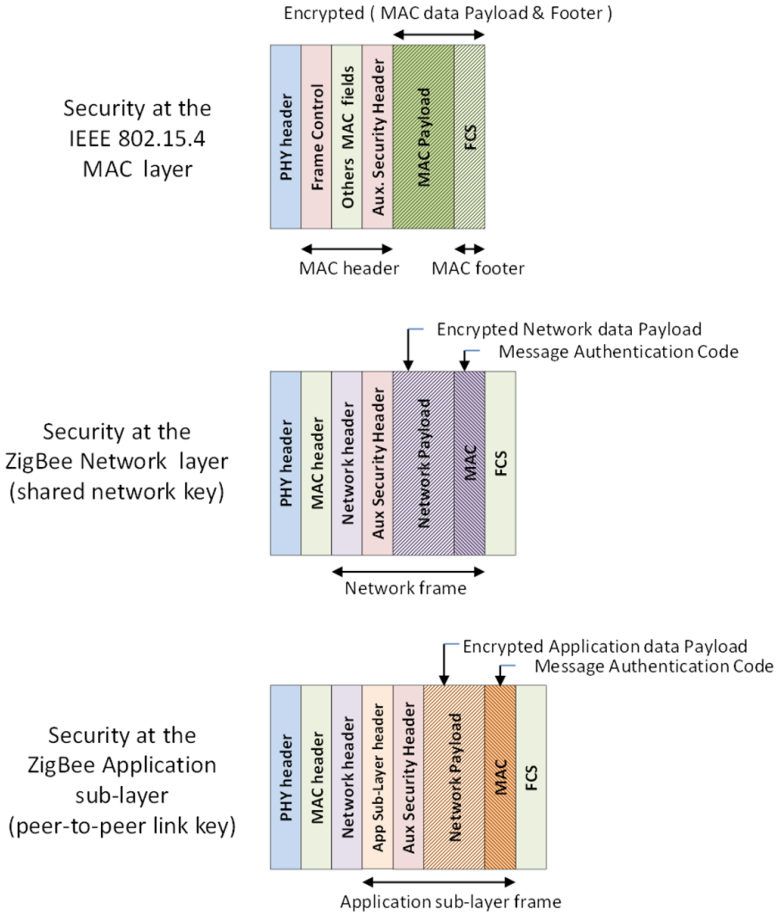


Figure 10: No name two

The security protection of the different layers can be combined.

ZigBee security provides support for key management, key transport, frame protection and key establishment. It holds a Trust Center that is usually the network coordinator. Its role is to ensure:

- The security management, to authenticate the nodes that request to join the network
- The network management to maintain and distribute network keys
- The configuration management to enable the security between several devices

3.1.2.2.2 Security framework based on ZigBee standard for Butler platform

For the Butler project, we used the ZigBee standard with the motes Waspmote from Libelium technology. We considered sensor network architecture organized in a “mesh” topology with three types of elements:

- The coordinator which is at the gateway interface between the LAN (local area network) and WAN (wide area network). This element must be “trust” because it provides key management in sensor network and stores cryptographic keys, information or resources. He plays the role of “Trust Center” for the sensor network Waspmote.
- Nodes routers only provide the routing function.
- Nodes sensors or actuators that are leaves of the network or “end-devices” and generate the resources to protect.

We have designed a homogeneous secure framework that can support the co-existence of several technologies based on IEEE 802.15.4 wireless standard. It handles motes of different capabilities and resources. A homogeneous policy and the associated protocols have been set to enable the coordinator to manage the security for three kinds of sensor network capabilities:

1. Symmetric key pre distribution via an “out-of-band” channel (for very constrained nodes)
2. Use of Raw Public Key thanks to a Lightweight DTLS handshake without certificate (for constrained nodes)
3. PKI scheme with certificate managed by the coordinator (for others nodes)

In the following, we address the two first techniques as the last one (PKI scheme) is reserved for the biggest motes with wide resources and can be achieved with open source library found in the literature. We focus on the deployment and the establishment of the “master” key of each node. This is the crucial phase of the security bootstrapping in the LAN. The others keys, at network or at application levels are derived from this key.

The ZigBee Cluster Library (ZCL) is organized into seven functional domains (in black in Table 1) specified in [14]

Table 4: Functional Domain for ZigBee Clusters

Functional domain	Cluster ID range
General	0x0000 to 0x00FF
Closures	0x0100 to 0x01FF
HVAC	0x0200 to 0x02FF
Lighting	0x0300 to 0x03FF
Measurement and Sensing	0x0400 to 0x04FF
Security and Safety	0x0500 to 0x05FF
Protocol interfaces	0x0600 to 0x06FF
SDM	0x0700 to 0x07FF

The “Security and Safety” domain is dedicated to a wireless Intruder Alarm System. Additional functions designed in this domain should be integrated into the IAS. ZigBee provides also Security Services for Key establishment but only for the Link key located at the Application layer and derived from the Master key. The master key is supposed to be secretly deployed and stored in the device. The question of the master key deployment is not tackled by the ZigBee standard. So, we will focus on this question and add features to bring solutions.

Into the ZCL, we define a new functional domain called “Security Deployment and Maintenance” (SDP) (in green in Table 4) to handle the secure bootstrapping operation. The associated Cluster ID range can be 0x0700 to 0x07FF.

For each security management scheme, we define the required cluster to achieve the master key deployment, the master key update, the certificate update (for the third scheme), the certificate (or public key) revocation.

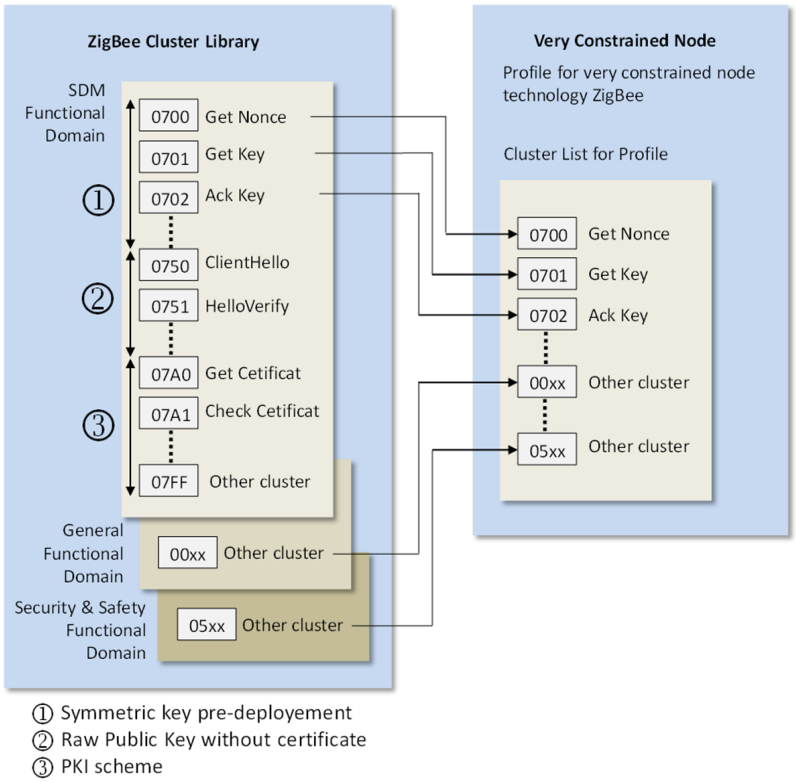


Figure 11: New Functional domain and Cluster ID for ZigBee

According to its capabilities, a node gets a Profile and is provided by the corresponding clusters of the SDP functional domain. This enables the binding between the node and the gateway/coordinator that manages the security and holds the master key features.

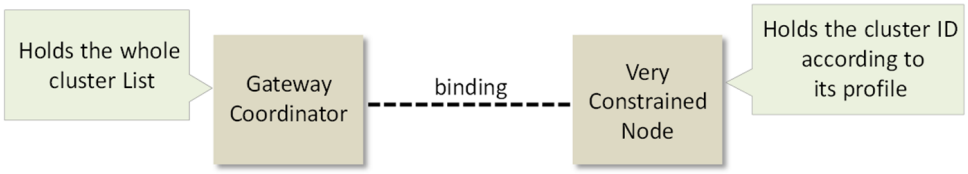


Figure 12: Binding between the coordinator and the node

3.1.2.2.3 Symmetric key distribution

We propose a security scheme based on the use of symmetric keys. It completes the security features provided by the ZigBee standard in order to reinforce the security of the master key that is the deep secret of the system, and to provide more scalability when several technologies co-exist.

ZigBee standard provides:

- A master key as the seed of the security
- A network key shared by the nodes at the network layer

- A link key between two peers at the application level.

We have added a functional domain to manage the deployment, the maintenance and the revocation of the master key that is the deep security feature of the node.

In our scheme, an initial master key is pre-deployed inside the node at manufacture via an out-of-band channel. This enables to get a pre-secured channel at start. When the node bootstraps to the network for the first time, this initial key is updated by the “first” operational master key. During the lifespan of the network, the master key will be periodically updated at the initiative of the administrator. An additional symmetric key is provided into the node at manufacture: the “global” key. This key is used at the MAC layer, is shared by all the nodes of the wireless sensor network and provides a security “at low layer” for the whole network.

So, the network key provided by the ZigBee standard can be a “Group Key” derived from the current master key and managed by the “Group Cluster”. This enables to address securely a group of nodes that share a common feature. The Link Key at application layer is also derived from the master key and enables the security management by application.

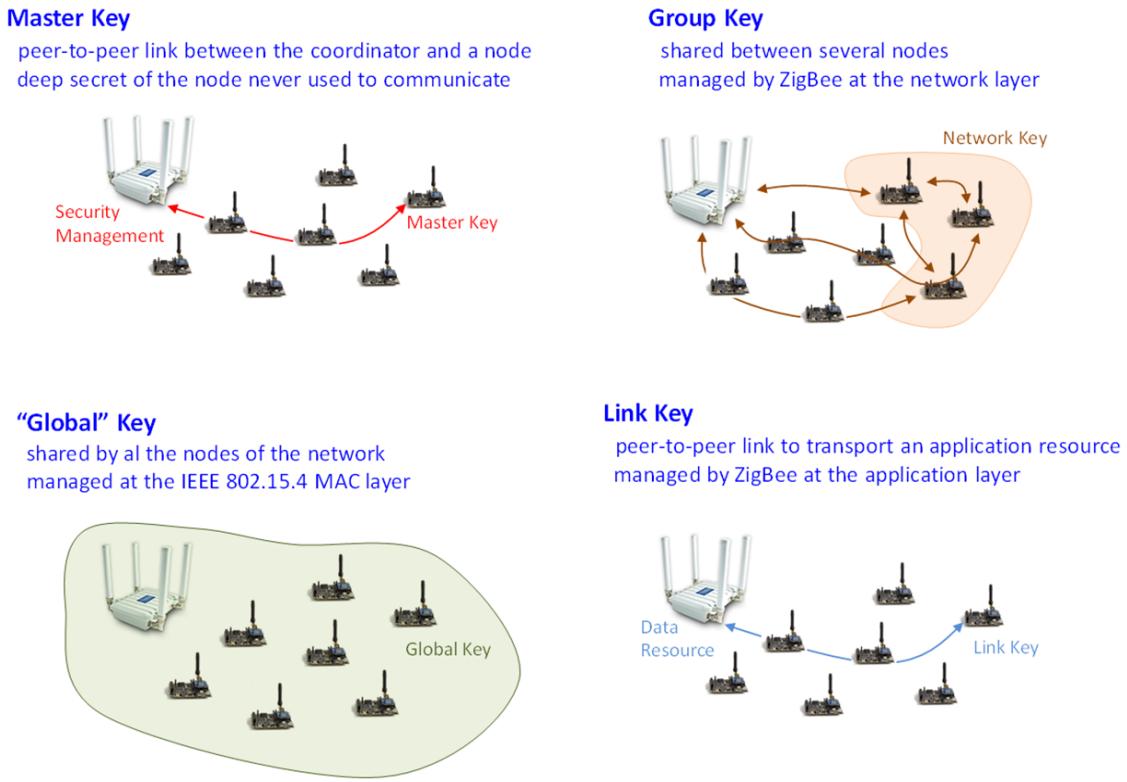


Figure 13: Four security keys for the security framework

3.1.2.2.4 The use of Raw Public Keys

In the following, we propose a lightweight handshake mechanism in the LAN to address the security deployment using Raw Public Keys of constrained data sources connected to the Internet via a trusted gateway and communicating in the LAN wirelessly with a low-power communication standard such as IEEE 802.15.4. This scheme can be used to deploy and to maintain the master key for constrained nodes.

This new lightweight authentication protocol is designed to ensure:

- The security of the wireless communication channel for low power communication standard in order to secure the communication between the constrained data sources and the gateway - i.e. the resource provider - that exposes the data over traditional Internet. The security scheme should protect against eavesdropping, replay attacks, attacks by packet injection, man-in-the-middle attacks, some deny-of-service attacks, spoofing of device identity.
- An easy deployment of the security mechanism at large scale in the LAN.
- The continuity of the security on the main trust boundary located at the gateway at the interface between the LAN and the WAN worlds.

This lightweight handshake crosses the layers of the OSI model, as it exploits some capabilities of the Physical layer able to sense the node environment thanks to the radio link, the physical sensors or internal characteristics such as internal clock jitter.

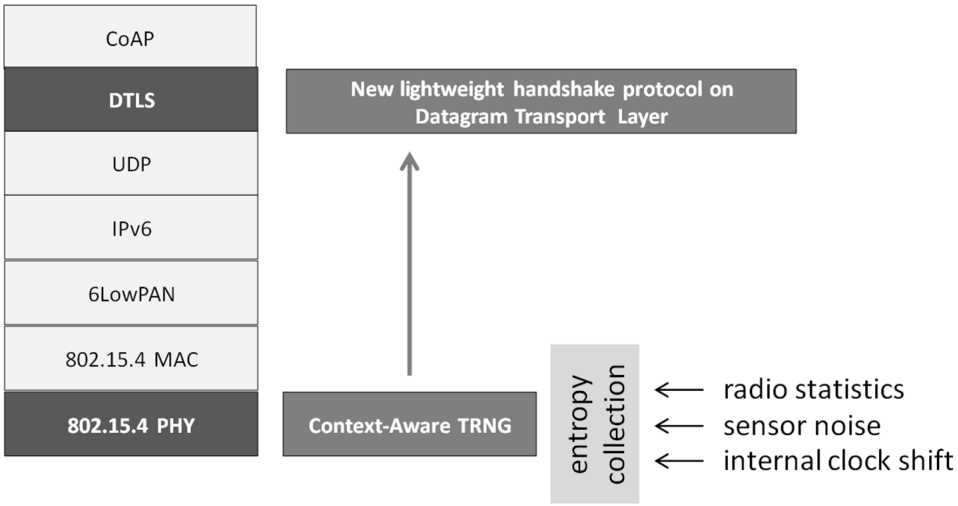


Figure 14: Cross-Layer security scheme

At the PHY layer:

The studies realized in [15], [16], [17] and [18] aim to harvest entropy form the physical environment of the node to enable constrained headless devices to generate “true” random numbers. This offers to the nodes new capabilities as generating their own cryptographic features, especially their asymmetric secret key that is never disclosed, nor stored in a remote “master” device.

Security on UDP (transport) layer:

DTLS proposes three modes for authentication and/or session key negotiation between a constrained data source - i.e. the node - and a gateway:

1. **Pre shared key:** This mode only uses symmetric cryptography. A shared key is stored inside the nodes before the deployment. Generally, the same key is shared between several nodes that form a group or a sub-network.
2. **Raw Public key:** This mode is based on an asymmetric cryptosystem and the pre-deployment of a couple of secret key/public key (SK/PK) into each node.
3. **Handshake protocol:** The complete DTLS handshake protocol using X.509 certificate standard.

The handshake protocol is the most complete mode enabling both authentication and session key negotiation based on an asymmetric cryptosystem and the use of certificates following the standard X.509. This protocol is quite big in terms of number of messages exchanged, overhead, memory

space and power consumption. It cannot be integrated in constrained devices. The raw public key protocol needs a pre-deployment of an asymmetric couple of keys into each node which is tedious at large scale.

In contrast, the Pre Shared Key protocol does not handle any authentication mechanism except the owning of the initial symmetric key shared by several entities.

It is not conceivable to embed into constrained devices complete handshake protocol. DTLS has been designed with the focus to ensure the interoperability with TLS for communication over UDP and not with the focus to be embedded into constrained devices. Yet, the use of asymmetric cryptography is seen as a benefit as it facilitates the deployment of the nodes and provides a higher security level and security scalability.

A new lightweight handshake protocol for secure bootstrapping

Based on the assumption that the nodes are provided with a new capability that enables the generation of their own secret key and public key with good entropy characteristics, we have designed a new lightweight handshake protocol that can fit into constrained devices.

Each node embeds a “true” random bit generator and an elliptic curve cryptosystem (e.g. secg-256r1) enabling its secret key generation. The gateway embeds an initial couple of asymmetric keys, the initial secret key, the public key and the lightweight associated ECC-based certificate signed by the authorization server master key ($SK_{initial}/PK_{initial}$ and $Cert_{initial}$). We suppose that each legitimate node knows the public material $PK_{initial}$ and $Cert_{initial}$ before deployment. The knowledge of the public cryptographic materials makes the node legitimate in the LAN as this public material is not really “public”, but is restricted and only disclosed to the legitimate nodes before deployment. This enables to initiate the security authentication protocol and session key agreement detailed in Figure 15.

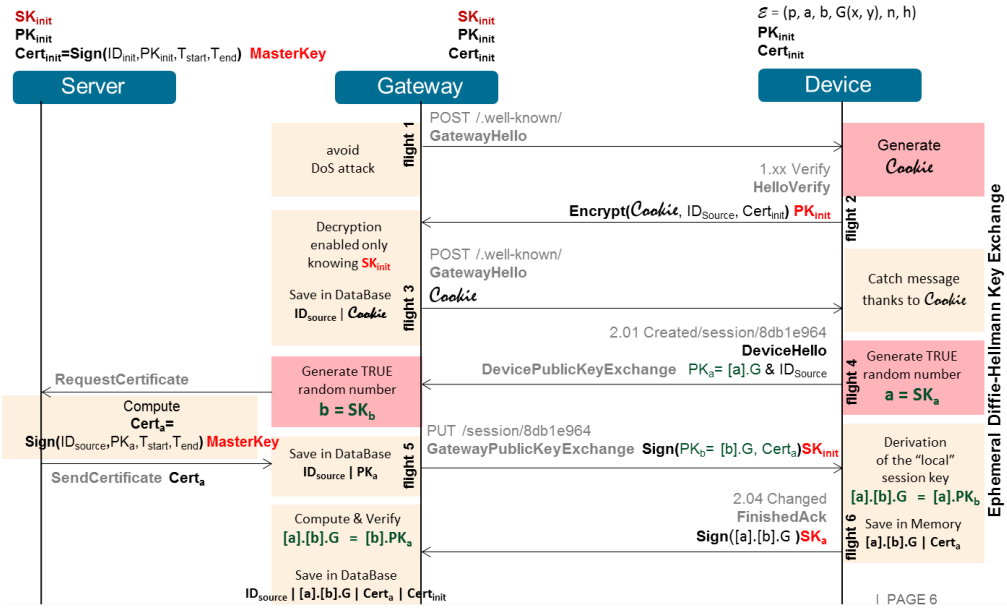


Figure 15: Overview of the lightweight handshake

The first challenge/response between the gateway and the data source is a ClientHello message following by a HelloVerify message containing a cookie. To avoid eavesdropping following by packet injection attack, the cookie, the data source identity (ID_{Data_Source}) and $Cert_{initial}$ should be encrypted with the initial gateway public key ($PK_{initial}$) stored by the data source before deployment. The role of the ciphered content is as follow:

- The cookie is a random number generated by the data source in order to avoid some deny-of-service attacks (attack by amplification and flooding the node by successive ClientHello

messages). In our scheme, the cookie plays a second role consisting in pseudo-naming the node during the handshake protocol.

- The data source identity is linked to the cookie and encrypted to preserve its confidentiality until the session key negotiation.
- The $Cert_{initial}$ encapsulated into the ciphered data content is an initial secret shared between the data source and the gateway used to prove the node legitimacy. It could be verified by the authorization server. In fact, the HelloVerify message cannot be signed as the node has not yet published its own public key.

The gateway is the unique device able to decrypt the HelloVerify response with the initial secret key $SK_{initial}$. It stores the data source ID and uses the cookie in another ClientHello challenge. Only the data source at the origin of this cookie will take the challenge in consideration. It replies with a ServerHelloAck message in two parts sent in clear text: its public key PK_a and its data source identity ID_{Data_Source}, where $a = SK_a$ is the secret key of the node generated by the embedded "true" random number generator.

Optionally, to improve the security in the case where the messages are lost by the legitimate receiver and simultaneously eavesdropped by a malicious entity, the second ClientHello including the cookie may be signed with $SK_{initial}$ and the ServerHelloAck content may be encrypted with $PK_{initial}$. In fact, even if the gateway is the unique entity to know the cookie after the flight 2, an attacker that has intercepted the cookie in clear text in the flight 3, possibly missed by the legitimate data source is able to flood the node by sending repeated messages including this cookie. For the ServerHelloAck message, the encryption of the content - i.e. $Cert_a \ \& \ ID_{Data\_Source}$ - avoids that an attacker injects fake public key associated with the correct data source identity if the legitimate ServerHelloAck message (flight 4) has been missed by the gateway.

When the gateway has sent the second ClientHello challenge, it starts a timer, waiting for the ServerHelloAck response including the data source identity associated with the cookie. When it receives the ServerHelloAck response in the given time interval, it checks that the ID_{Data_Source} is the correct one. Then, it transferes the data source public key and identity to the authorization server via a secure SSL/TLS ethernet link to request for a certificate. The authorization server creates the data source certificate signed with its master key and responds to the gateway. The gateway uses the associated public key PK_a to compute the session key by $SK_b \cdot PK_a$ where $SK_b = b$ is the gateway current secret key.

Thanks to the flight 5, the Diffie-Hellman mechanism enables to share a session key and to send to the data source its certicate signed by the server. The gateway sends a Finished message with its current public key and the certificate PK_b and $Cert_a$ in clear text. Upon reception, the data source stores its certificate and computes the session key by $SK_a \cdot PK_b$. As $SK_b \cdot PK_a = SK_a \cdot PK_b$ the gateway and the data source share a "local" session key. The data source acknowledges the Finished challenge by a Finished response handling the digital signature of the session key. During this last message exchange, the data source and the gateway could also agree on a symmetric cipher suite.

Once the handshake achieved, the data source may erase the initial public key and certificate ($PK_{initial}$ and $Cert_{initial}$) stored before deployment and remove the cookie.

This security scheme facilitates large scale deployment of additional data source over the LAN as all the nodes are provided by the same software and the same initial public key ($PK_{initial}$) at start. Thanks to the use of certificate, the cryptographic materials can be managed and maintained. If a node is considered as malicious, the authorization server can revoke it.

Nevertheless, lightweight and dedicated certificates based on ECC cryptosystem should be designed to enable exchanges of short messages on the throughput.

The session key renewal

The symmetric session key may be used for a given time to secure the communication between a node and an identified gateway. Its renewal is initiated by the authorization server when the associated certificate expires or when all the initialization vector values have been used for this session key. The authorization server will send to the gateway a request to launch a new handshake mechanism with the node. At the issue of the handshake, a new session key with a new certificate associated is established (see Figure 16).

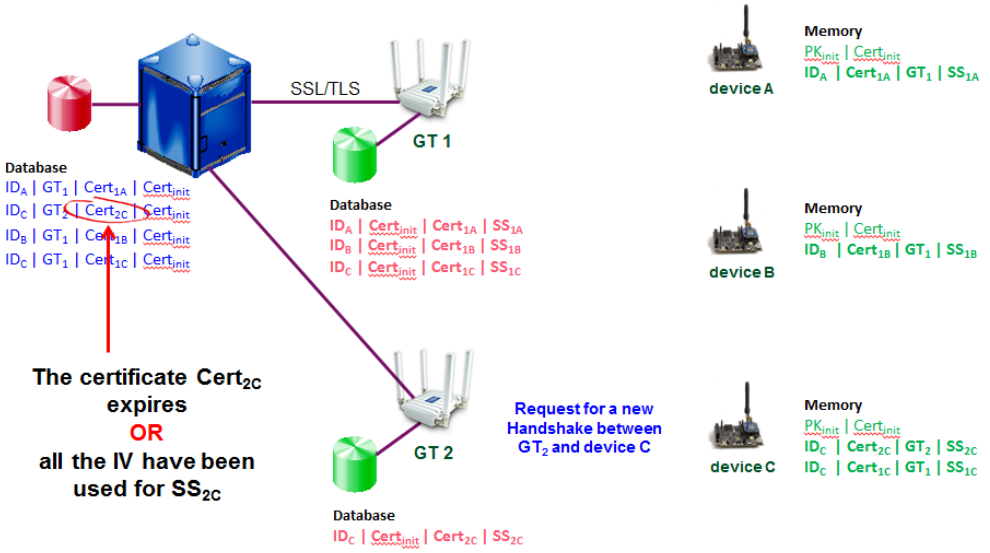


Figure 16: Session key renewal between a node and a gateway

The initial key renewal

The initial key and the initial certificate are a secret disclosed only to the legitimate nodes before deployment. It must frequently change to avoid its capture by an attacker. Their renewal is decided by the authorization server when $Cert_{initial}$ expires. The authorization server generates a new couple of asymmtric public and secret key and the associated new certificate noted $SK_{initNEW}/PK_{initNEW}$ and $Cert_{initNEW}$. Then, the server sends these new security credentials to the gateways for update via a secure SSL/TLS link. All the future nodes to deploy will embed the public material $PK_{initNEW}$ and $Cert_{initNEW}$.

All the nodes that are already deployed will receive a request from a gatway in their area to show the initial certificate they embed. This protocol is launched thanks to the channel secured with the current session key. If the initial certificate of the node is the old one, it is updated. This enables the node to move, to change of area and to establish a new handshake with another gateway belonging to the same infrastructure and knowing the new initial credential. If a node is not reachable during the time of the renewal of the intial material, it will not be able to establish a new handshake in the future (see Figure 17).

This scheme enables also to revoke nodes or cryptographic element that has been compromise. The revocation process is managed by the authorization server. Certificates are used to manage the cryptographic keys along time, in a future work, lighweight certificates based on ECC cryptography should be designed to enable an efficient management of the security of the constrained devices.

3.1.2.3 Theoretical Security under Fading Channels

A continuing trend of miniaturisation and a growing demand for information are two powerful drivers of new communication systems, which therefore increasingly rely on embedded technology, as

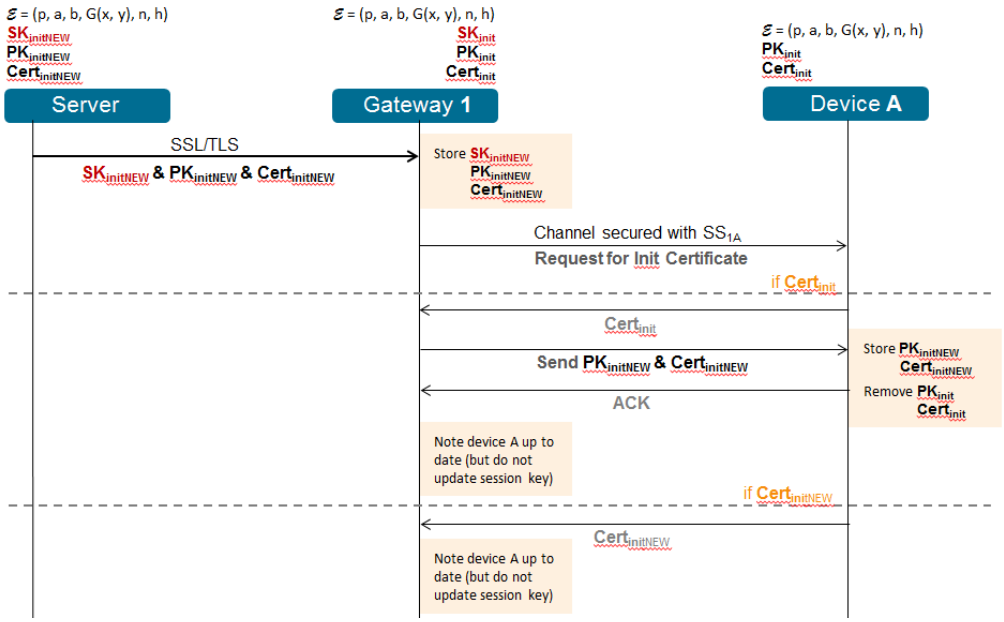


Figure 17: Initial key renewal

illustrated by IoT. The embedded nature of future wireless networks implies not only power-limitation of devices, but also a likelihood that a greater share of traffic will be of highly sensitive and personal information, which together pressure for **new wireless security mechanisms** that do not rely on the overhead-heavy and coordination-intense protocols of today.

One approach to meet such requirements is physical-layer information-theoretical security [19], which aims at eliminating the need of cryptography altogether. These seminal works on secrecy [20–22] established the notion that any wireless channel has an intrinsic **secrecy capacity**, which is fundamentally determined by how the power of the signal at a legitimate destination compares against that at an eavesdropper. After a break of three decades, these ideas have recently re-emerged and shown to apply also to wireless fading channels [23, 24]. In this context, a number of previously ignored factors gain significance, including the facts that: a) wireless devices **form networks of unknown topology**; and b) **devices in wireless networks interfer with each other**.

It is therefore fundamental to better understand the inherent secrecy of wireless systems under more realistic conditions. To this end, Stochastic Geometry [25] and the notion of **secrecy graphs** have emerged [26, 27], such that the study of achievable **secrecy rates of random networks** have initially followed steps similar to those taken to characterise the communication rates of random networks [28, 29]. Examples of such an approach are the secrecy rate scaling laws derived in [30] and [31], the secrecy transmission capacity of homogeneous networks studied in [32, 33]. The same approach has been adopted also when considering the effect of interference onto the achievable secrecy rates in wireless networks [24, 26, 33, 34]. It has been shown that the Poisson Point Processes (PPPs) can accurately model the majority of wireless systems of interest. Our works are formulated on the latter premises. Specifically, we will study the inherent secrecy in random networks, with the particular aim of **modelling topological models** and **considering the effect/role of interference**, as detailed below.

3.1.2.3.1 Results Related to the Networks of Unknown Topology

Order Statistics and Network Topology

First, consider a random network in an unbound Euclidean space of dimension d, modeled by a stationary Poisson point processes (PPP) [35] of intensity λ in $\mathbb{R}^d$, and select an arbitrary refer-

ence point defining the origin of the space, ordering the remaining points $k \in \mathbb{N}$ according to their Euclidean distances r_k to such a reference.

It is well-known that the ordered distances r_k are such that $r_k \neq r_j$ for $k \neq j$, almost surely [25]. This property is implicitly used henceforth to support the assumption that each node in the network can be **unequivocally identified** by its distance to the origin (source).

Let the aforementioned model be applied to two overlaid networks of **legitimate** nodes and **eavesdroppers**, respectively, with corresponding densities λ_{L} and λ_{E}. Consider that a source located at the origin (without lack of generality) wishes to unicast to the legitimate node k in the presence of an eavesdropper located at the unknown distance r_{E}, subjected to nakagami-m fading and path loss governed by the exponent α. Then, the **secrecy capacity** of the unicast channel can be re-written as [21, 24]

$$\mathcal{C}_{s:k} = \log_2\!\left(1+\frac{|h_k|^2 P}{r_k^{\alpha} N_0}\right) - \log_2\!\left(1+\frac{|h_{\mathrm{E}}|^2 P}{r_{\mathrm{E}}^{\alpha} N_0}\right) \quad \text{b/s/Hz}, \tag{1}$$

where we have implicitly defined the quantities $\zeta_{\mathrm{L}} \triangleq \frac{|h_{\mathrm{L}}|^2}{r_{\mathrm{L}}^{\alpha}}$ and $\zeta_{\mathrm{E}} \triangleq \frac{|h_{\mathrm{E}}|^2}{r_{\mathrm{E}}^{\alpha}}$, which denote the channel gains of the legitimate node and the eavesdroppers, respectively.

In order to obtain an expression for the secrecy outage probability – or equivalently, for the secrecy non-outage probability defined equation (1), – the distributions of the path gains ζ_{L} and ζ_{E} need to be derived.

To this end, we derived the distribution of the path gain to the k-th closest legitimate node in [36], obtaining

$$p_{\zeta_{\mathrm{L}}}(x; k, m, \eta_{\mathrm{L}}) = A(k; m, \eta_{\mathrm{L}}) \cdot \frac{x^{m-1}}{(\eta_{\mathrm{L}}\, x + 1)^{m+k}}, \tag{2}$$

where the auxiliary function $A(k; m, \eta)$ is defined as

$$A(k; m, \eta) \triangleq \frac{\Gamma(m+k)\eta^m}{\Gamma(m)\Gamma(k)}, \eta \triangleq m/(\lambda\pi). \tag{3}$$

In the secrecy capacity analysis, we now notice that for a given legitimate path gain $\zeta_{\mathrm{L}:k}$ what determines the secrecy capacity of a channel subjected to fading is not any specific eavesdropper, but rather the eavesdropper with the **maximum** (instantaneous) path gain amongst those present. Consequently, concerning the eavesdropping network and assuming that the communicating pair is exposed to an **unknown** number K of eavesdroppers, the distribution of interest is an **extreme value distribution**, namely, the statistics of the quantity

$$\bar{\zeta}_{\mathrm{E}} \triangleq \max\{\zeta_{\mathrm{E}:1}, \cdots, \zeta_{\mathrm{E}:K}\}. \tag{4}$$

We have shown in [36] that $\bar{\zeta}_{\mathrm{E}}$ is accurately modelled by a two-parameter Generalized Extreme value (GEV) approximation for $\bar{\zeta}_{\mathrm{E}}$, then

$$p_{\bar{\zeta}_{\mathrm{E}}}(x; \nu, \theta) \approx \frac{1}{\theta}\left(\tfrac{x-\nu+\theta}{\theta}\right)^{-2} e^{-\left(\frac{x-\nu+\theta}{\theta}\right)^{-1}}, \tag{5}$$

where the parameters ν and θ are the location and scale parameters, respectively.

Here, yet another comment is relevant, namely, that **the approach leading to the result described by equation** (5) **is even more flexible to generalisation then that of equation** (2). Indeed, it is well known [37] that order statistics distributions are robust to variations of the shapes of distributions, depending mostly on the tail's decay rate.

Back to our discussion, using the path gain distributions of legitimate and eavesdropper nodes, one can compute the secrecy non-outage probability between the source and k-th node (in the presence of a randomly located multiple eavesdroppers) as

$$\tilde{\mathcal{P}}_{\text{out}}(R_s; m, \eta_{\text{L}}, \nu, \theta) = \frac{A(k; m, \eta_{\text{L}})}{\theta \eta_{\text{L}}^{m+k}} e^{\frac{2^{R_s}\theta}{\beta}} \sum_{j=0}^{m-1} \binom{m-1}{j} \frac{(-\frac{1}{\eta_{\text{L}}})^j}{\beta^{k+j+1}(k+j)} \sum_{t=0}^{k+j} \binom{k+j}{t} \times (-2^{R_s}\theta)^t \big[(2^{R_s}\theta)^{1-t} E_t(\tfrac{2^{R_s}\theta}{\beta}) - (2^{R_s}\theta + \beta \tfrac{\theta}{\theta-\nu})^{1-t} E_t(\tfrac{2^{R_s}\theta}{\beta} + \tfrac{\theta}{\theta-\nu}) \big]. \quad (6)$$

where $\beta \triangleq (2^{R_s}(\nu - \theta + \rho^{-1}) - \rho^{-1} + \frac{1}{\eta_{\text{L}}})$ and E is an exponential integral.

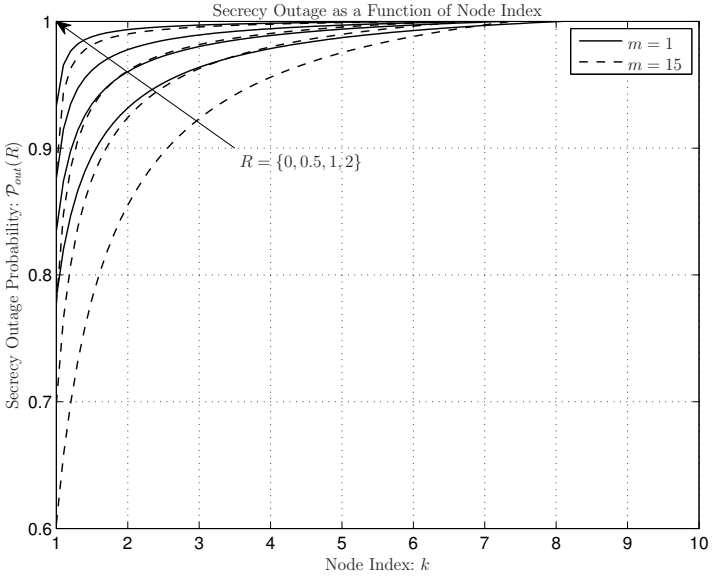


Figure 18: Secrecy outage as a function of node index in the case of Rayleigh fading ($m = 1$) and for various rates, with unitary reference SNR ($\rho = 1$).

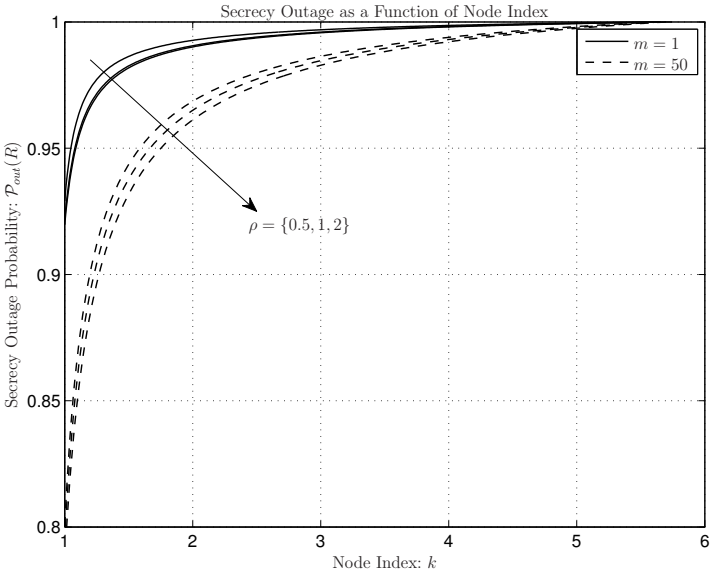


Figure 19: Secrecy outage as a function of node index in the case of Rayleigh fading ($m = 1$) and for various reference SNR's ($\rho = \{0.5, 1, 5, 25\}$), with unitary rate ($R = 1$).

Plots of the secrecy outage obtained with equation (6) are shown in Figures 18 and 19. The results in Figure 18 are intuitive, namely they indicate that it is hard to ensure a low secrecy outage for nodes further from the source and for high rates. The results in Figure 19, however, are less intuitive, as they shown that the achievable secrecy non-outage is largely **independent on the reference SNR** ρ, *even for a fixed rate*!

At a glance, equation (6) looks somewhat cumbersome, but its significance can be captured as follows. Let us consider the case where the source is able to identify which of its neighbors has

the best path loss, subsequently unicasting to that node. This case relates to the scenario studied in [23, 24], in the sense that the selection of the device with the "best channel" can either occur in terms of the "best node" at a given time thanks the quasi-stationarity of the channel – as assumed in [24] – or in terms of the "best time" – as assumed in [23]. Simplifying equation (6) accordingly, we have shown [38] that in this case the probability that a non-zero secrecy capacity between the source and the **best** node exists (in the presence of a randomly located multiple eavesdroppers) is given by

$$\tilde{\mathcal{P}}_{\rm out}(0) = \frac{\lambda_{\rm E}}{\lambda_{\rm L} + \lambda_{\rm E}}. \tag{7}$$

Again, equation (7) is highly motivating in the context of our works, as it indicates that the existence of a secrecy capacity between a pair of devices immersed in a random network, is a function of the density ratio where $\lambda_{\rm L}/\lambda_{\rm E}$ between legitimate and eavesdropping nodes! This specific result holds only for homogeneous PPP networks, but there is clear hope to obtained generalised versions for other point processes. For instance, it is evident that in the case of a non-homogeneous PPP the result would also hold, only with $\lambda_{\rm L}$ and $\lambda_{\rm E}$ replaced by **local** densities. For more general cases, other concise measures of statistical proportionality between eavesdroppers and legitimate nodes may turn out.

Correlation and Network Topology

An indirect way to generalise the assumption of uniformity that is implied by a PPP is to incorporate spatial correlation into the model. Indeed, clusterization [39] – a mechanism that is often used to approximate other point processes via PPPs [40] – can be seen a particular case of correlation, since it ultimately consists of correlation of the random distances from a source to group of clusterized nodes, which in turn as *per* equation (1) affect the legitimate and eavesdropping SNRs and thus the secrecy capacity.

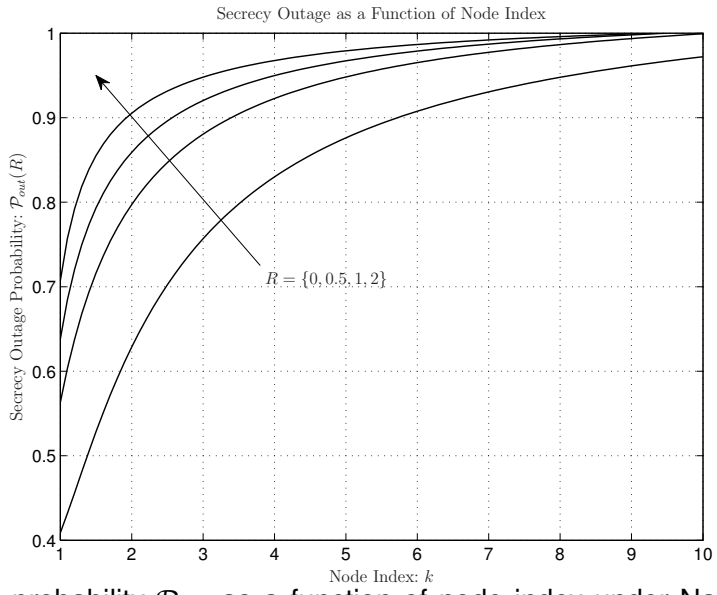


Figure 20: Secrecy outage probability $\mathcal{P}_{out}$ as a function of node index under Nakagami-m fading for various secrecy rates, with $\lambda_{\rm L} = \lambda_{\rm E} = 1$ and $m = 1$.

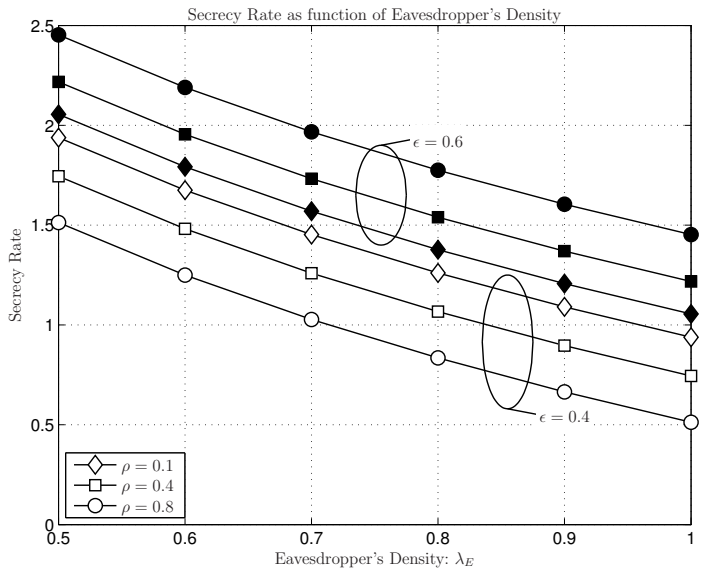


Figure 21: Secrecy transmission capacity as a function of eavesdroppers' intensity, with $\lambda_{\mathrm{L}} = 2$, $\alpha = 2$, $m = 1$, $k = 1$.

The body of work where the impact of correlation in the secrecy of random networks is comparatively small, but the issue has not entirely escaped the attention of the community. For instance, the secrecy capacity and secrecy outage capacity of point-to-point system subjected to correlated fading channels were studied in [41] and [42], respectively.

To the best of our knowledge, no previous work exists on the secrecy capacity and outage capacity of random wireless networks with channel correlation. Specifically, we have shown [43] that an **asymptotic expression** for the secrecy outage probability under correlated fading channels in the high SNR regime is

$$\mathcal{P}_{out}(R_s; m, \delta, \lambda_{\mathrm{L}}, \lambda_{\mathrm{E}}, \rho) = \frac{2^{2m-1}\Gamma(m+\frac{1}{2})(1-\rho)^m}{\sqrt{\pi}\Gamma(m)} \int_0^{\infty} \frac{z^{\delta k+m-1}(z+1)}{[(z+1)^2-4\rho z]^{m+\frac{1}{2}}} \left[\frac{1}{z^{\delta k}} - \frac{1}{(z^{\delta} + \frac{\lambda_{\mathrm{E}}}{\lambda_{\mathrm{L}}} \times 2^{R_s})^k}\right] \mathrm{d}z. \tag{8}$$

where $\delta = 2/\alpha$.

The impact of correlation on the secrecy outage of different nodes under Nakagami-m fading can be observed in Figure 20, where we show plots of $\mathcal{P}_{out}$ as a function of node index for a constant channel correlation coefficient $\rho = 0.9$. The figure displays various curves for various secrecy rates, with $\lambda_{\mathrm{L}} = \lambda_{\mathrm{E}} = 1$, $\alpha = 2$ and $m = 1$. Compared to Figure 18, no particular new insight is gained, as it is found that for any given secrecy rate R, nodes further away have higher outage than nodes closer to the source, as expected.

Consider however the minimum secrecy rate R achieved under a certain outage ϵ, as a function of the density of eavesdroppers λ_{E}, which can be obtained by numerically inverting equation (8). The results are shown above Figures. This time it can be seen that, surprisingly, the **secrecy rate increases with the correlation if the secrecy outage is higher then 50%, but decreases otherwise**. Since, as shown in Figure 20, farther nodes have higher outage, we conclude from both figures together that correlation helps farther nodes and harms nearer ones.

3.1.2.3.2 Results Related to Interference Aggregation with Secrecy

Interference Aggregation in Fading Channels

Interference is another key parameter in characterizing the network-wide secrecy throughput of large scale networks. If undesigned, interference is an aggregated sum of undesired signals due to concurrent transmissions, that may cause severe throughput degradation. Such an interference

can be modelled as a stochastic process, with the random location of interferers described by point process $\mathcal{I}$. Then a generalized model of aggregate interference can be defined as

$$\mathcal{I} = \sum_{i \in \mathcal{I}} X_i \cdot r_i^{-\alpha}, \tag{9}$$

where r_i is the distance between the receiver and the i-th interferer, α is a propagation loss coefficient and $X_i \triangleq |h_i|^2$ models the channel power.

Stochastic geometry is one of the tools that can be used to characterize the statistical behaviour of aggregate interference. A convenient way to do so is via the Laplace Transform (LT) of $\mathcal{I}$, or its characteristic function (CF), namely

$$\mathcal{L}_{\mathcal{I}}(w;\alpha) = \mathbb{E}[e^{-w\mathcal{I}}], \tag{10}$$

where the expectation is take over the distributions of X_i and r_i, and the parameters of those distributions are omitted from the notation for the sake of simplicity and generality.

In the case of the PPP model, $\mathcal{L}_{\mathcal{I}}(w;\alpha)$ can then be relatively easily evaluated via Campbell's theorem, which relying on the uniformity of the PPP yields

$$\mathcal{L}_{\mathcal{I}}(w;\alpha) = \exp\left(-2\pi\lambda \int_X \int_0^\infty [1-\exp(wxr^{-\alpha})] f_X(x) f_R(r) r \, \mathrm{d}r \mathrm{d}x\right). \tag{11}$$

Specifically, we have recently performed an analysis of the the secrecy outage in PPP networks subjected both to arbitrary Nakagami fading and Log-Normal shadowing [44]. Indeed, neither equation (11) nor its inverse Laplace transform admit closed forms except for the specific case of Rayleigh fading [33].

In order to characterize the aggregate interference under generalised fading conditions, however, we employed equation (11) to obtain closed forms of the corresponding cumulants. With the exact cumulant expression, an Edgeworth series can in principle be built, yielding an asymptotically optimum expansion that approximates the desired probability distribution in terms of its cumulants [45].

Due to the analytical intractability of Edgeworth model, however, it proves more convenient to work with other models based on the Gamma and Log-Normal distributions. With the Gamma model, for instance, we obtained the following result [44]

$$\mathcal{P}_{out}(\beta_e; r, \nu_{\mathrm{E}}, \theta_{\mathrm{E}}, \nu_i, \theta_i) = \frac{\pi\lambda_e \Gamma(\nu_{\mathrm{E}}+\nu_i)}{\Gamma(\nu_{\mathrm{E}})} \left(\frac{\theta_e}{\beta_e\theta_i}\right)^{\nu_i} \left\{ \left(\frac{\theta_e}{\beta_e\theta_i}\right)^{1-\nu_i} \Gamma\begin{pmatrix} 1+\nu_i, \nu_i-1, 1, \nu_{\mathrm{E}}+1 \\ \nu_i, \nu_{\mathrm{E}}+k+i, 2 \end{pmatrix} \right. \tag{12}$$

$$\left. \times {}_2F_2\left(1, \nu_{\mathrm{E}}+1, 2-\nu_i, \nu_i; \frac{\lambda_e\pi\theta_e}{\beta_e\theta_i}\right) + (\pi\lambda_e)^{\nu_i-1}\Gamma(1-\nu_i) {}_2F_2\left(\nu_i, \nu_{\mathrm{E}}+\nu_i, 1+\nu_i, \nu_i; \frac{\lambda_e\pi\theta_e}{\beta_e\theta_i}\right) \right\}.$$

Furthermore, we have shown that the secrecy outage probability under a combination of Nakagami-m fading and Log-Normal shadowing is well approximated by

$$\mathcal{P}_{out}(\beta_{\mathrm{E}}; r, \mu, \sigma) = -\frac{e^{-\frac{\pi\lambda_e\mu}{\beta_{\mathrm{E}}}} e^{\left(\frac{\pi\lambda_e\sigma\sqrt{2}}{2\beta_{\mathrm{E}}}\right)^2}}{2} \operatorname{erfc}\left(\frac{\pi\lambda_e\sigma}{\sqrt{2}\beta_{\mathrm{E}}} - \frac{\mu}{\sqrt{2}\sigma}\right) + \left(1+\operatorname{erf}\left(\frac{\mu}{\sqrt{2}\sigma}\right)\right)/2, \tag{13}$$

where $\operatorname{erf}(x)$ and $\operatorname{erfc}(x)$ are the Gaussian error functions.

Cooperation and Aggregation

As discussed earlier, one of the ways interference can affect the secrecy of a random network is by aggregating so as to reduce the SINR at legitimate nodes. But as argued earlier, it is somewhat naive to assume that interference aggregates randomly.

Instead, a more modern approach is to formulate optimisation problems around the aggregation of interference [46, 47], which can be in turn seen as a form of cooperation. But since cooperation is

certain to play a role, one of our recent works on the role of cooperation in network secrecy is worth of mention [32, 48].

Optimal cooperation amongst eavesdroppers is referred to as collusion, since that in light of the secrecy capacity expression given by equation (1), the optimum outcome of eavesdropping cooperation is to aggregate the power of all eavesdropping signals. The number of articles considering the impact of eavesdroppers' collusion in random networks is again comparatively small, but the issue has been occasionally considered. For instance, the secrecy capacity of a single legitimate link with length r_{L} in the presence of colluding eavesdroppers under an AWGN channel model was studied in [49], and the probability of non-zero secrecy capacity with multiple antenna transmission schemes in Rayleigh fading channels was analyzed in [50].

We contributed to this discussion by deriving in [32]. Closed-form asymptotic expressions for the secrecy rate distribution, average secrecy rate, secrecy outage probability and secrecy transmission capacity of random networks with Nakagami-m fading channel and colluding eavesdroppers.

The secrecy capacity of a unicast link in the presence of colluding eavesdroppers can be written as [49]

$$C_{s:\mathrm{L}} = \max\left\{\log_2\left(1+\varrho\frac{|h_{\mathrm{L}}|^2}{r_{\mathrm{L}}^{\alpha}}\right) - \log_2\left(1+\varrho\sum_{k=1}^{\infty}\frac{|h_{\mathrm{E}:k}|^2}{r_{\mathrm{E}:k}^{\alpha}}\right), 0\right\}, \tag{14}$$

where $h_{\mathrm{E}:k}$ and $r_{\mathrm{E}:k}$ are the fading envelope and the distance associated with the channel between the source and the k-th eavesdropper.

In turn, the connection outage probability, defined as the probability that the capacity of the legitimate channel is less than the transmission rate is given by

$$\mathcal{P}_{co} = \Pr\left\{\log_2\left(1+\varrho\frac{|h_{\mathrm{L}}|^2}{r_{\mathrm{L}}^{\alpha}}\right) \leq R_t\right\}, \tag{15}$$

For the PPP case, using also equation (2) we have shown that the secrecy outage probability for the case of **colluding** eavesdroppers is given by [32]

$$\mathcal{P}_{\mathrm{out}}(R_s; r_{\mathrm{L}}, m, \alpha, \lambda_{\mathrm{E}}) = 1 - \frac{\Gamma(\nu+m)\,{}_2F_1\left(\nu, \nu+m, 1+\nu; -\dfrac{1}{m\,\theta\, r_{\mathrm{L}}^{\alpha}\, 2^{R_s}}\right)}{(m\,\theta\, r_{\mathrm{L}}^{\alpha}\, 2^{R_s})^{\nu}\Gamma(m)\Gamma(\nu)\nu}. \tag{16}$$

Using these results we have then shown that in the case of colluding eavesdroppers, the secrecy transmission capacity of uniformly random networks under Nakagami-m fading is [32]

$$\tau = (1-\mathcal{P}_{\mathrm{co}})\lambda_{\mathrm{L}}\frac{(mr^{\alpha})^{-\nu}\Gamma(\nu+m)}{\nu^2\theta^{\nu}\Gamma(m)\Gamma(\nu)}\,{}_3F_2\left(\nu, \nu, \nu+m, 1+\nu, 1+\nu; -\frac{1}{m\,\theta\, r_{\mathrm{L}}^{\alpha}}\right). \tag{17}$$

Setting aside the impact of fading, all the results combined also indicate that information theoretical secrecy (in the Wyner sense) is only significant in random networks with colluding eavesdroppers, if a guard zone of reasonable size exists, and if the legitimate pair is within it.

3.1.2.3.3 Issues

On the Network Topology

In order to illustrate the importance of network topology on the conditions determining the secrecy capacity of corresponding network, consider the example of typical residential WiFi networks. From a security standpoint, these networks are characterised by the clusterization on a home-by-home basis, $i.e.$, devices within the house are typically considered legitimate users, while devices outside premises play the role of possible eavesdroppers. Likewise, in cellular networks, the distribution of base stations (and consequently of users) plays follow terrain, regulatory (city-plan), demand and

space availability conditions, and therefore are far from uniform. Devices in urban areas served by pico and femto cells may be clustered together, for instance, while devices in less populated areas served by macro cells are more sparsely located.

Such conditions are clearly distinct form the random and uniformly distributed network assumptions – $i.e.$, Poisson Point Process (PPP) model – commonly adopted in current literature [24, 26, 33, 34, 51]. In response to the limitations of the usual PPP model, recent work has appeared which focuses on the impact of topological models onto the accuracy of analytical results obtained for random networks [52].

In future works, we will **investigate** various parameters of interest such as the **node degree** of secrecy graphs, the **secrecy outage probability**, the **unicast secrecy capacity** and the **secrecy transmission capacity** of **random networks of various topological characteristics**, employing emerging stochastic geometric models **beyond the PPP model**, as well as alternative techniques **beyond stochastic geometry** itself.

On Interference and Secrecy

Besides topology, interference is another key parameter in characterising the performance of random network. To some extent interference is related to the network topology [53], in the sense that modifications of the latter lead to variations in the former.

But interference is far from being determined solely by topological features. Due both to its relationship with topological models, and the various methods to "design" interference in wireless systems, it is clear that studying the impact of interference in the wireless secrecy is of the highest importance. In future works we will **analyse** both **the impact of the more accurate topological models** described above **onto the interference of random networks in its relation with secrecy**, as well as **develop new interference design techniques** to improve secrecy in random networks **beyond the PPP model**.

3.1.2.4 A Secret Key Exchange Scheme for Short Range Communication

Short-Range Communication (SRC) is a growing field in wireless communications [54], as these technologies allow devices in close proximity to establish high-speed communication without physical contact. Typical examples of SRC systems are the Near Field Communication (NFC) [55], Dedicated Short-Range Communication (DSRC) [56] and ZigBee [57] standards, which are in great demand as underlying technologies to enable the deployment of the Internet of Things (IoT) in the near future.

Various techniques have been introduced in the existing standards of SRC systems. This approaches are, however, computationally demanding and power hungry, which goes against the desirable features of SRC devices. In recent years, on the other hand, significant advances have been made on physical layer security alternatives [58], which have the advantages of being more autonomous and consequently more scalable then traditional methods. Amongst these novel alternatives is the notion of dynamic secret key generation (SKG) schemes, designed to enable communicating pairs to locally produce secret keys which can then be used to seed their embedded cryptography systems [59].

However, the vast majority of SKG schemes proposed thus far relies on the assumption that a reciprocal source of entropy can be found in the channel between the communicating pair [59–67], which is an assumption clearly incompatible with SRC conditions. Specifically, SRC applications typically involve brief message exchanges between devices in close proximity – thus benefiting from high Signal-to-Noise Ratios (SNRs) – and in absence of mobility – thus not subjected no fading.

Motivated by the limitation of existing methods [59–67], and inspired by the classic information-theorerical results on secrecy capacity, we propose a new phase-based, dynamic SKG mechanism suitable to SRC systems. The key idea is to exploit the Additive White Gaussian Noise (AWGN)

conditions faced by SRC systems to build geometric secrecy regions within which eavesdroppers cannot acquire phases exchanged between the legitimate pair.

3.1.2.4.1 Secret Key Construction

System Model

Consider the setting depicted in Figure 22 in which a pair of devices, hereafter referred to as Alice and Bob, exchange phase signals chosen at random with uniform distribution, in the presence of an eavesdropper, hereafter Eve. Assume that the channel between Alice and Bob is reciprocal and constant during the exchange, such that the phases φ_A (transmitted from Alice to Bob) and φ_B (transmitted from Bob to Alice), are respectively received at their corresponding peers as $\hat{\varphi}_A\big|_B \triangleq \varphi_A + \varphi_C + \varepsilon_B$ and $\hat{\varphi}_B\big|_A \triangleq \varphi_B + \varphi_C + \varepsilon_A$, where φ_C is an unknown phase rotation introduced by the channel, while ε_A and ε_B are estimation errors due to thermal noise.

In presence of such exchange, Eve can obtain the estimates $\hat{\varphi}_A\big|_E \triangleq \varphi_A + \varphi_{C_{AE}} + \varepsilon_{AE}$ and $\hat{\varphi}_B\big|_E \triangleq \varphi_B + \varphi_{C_{BE}} + \varepsilon_{BE}$, where $\varphi_{C_{AE}}$ and $\varphi_{C_{BE}}$ are phase rotations introduced by channels between Alice or Bob and Eve, respective, while ε_{AE} and ε_{BE} are errors due to noise.

Next, consider that Alice and Bob combine their corresponding receive and transmit signals so as to collaboratively construct the uniformly distributed random number $\theta = \varphi_A + \varphi_B + \varphi_C$. Then, Alice and Bob would respectively obtain the estimates $\hat{\theta}\big|_A = \hat{\varphi}_B\big|_A + \varphi_A = \varphi_A + \varphi_B + \varphi_C + \varepsilon_A$ and $\hat{\theta}\big|_B = \hat{\varphi}_A\big|_B + \varphi_B = \varphi_A + \varphi_B + \varphi_C + \varepsilon_B$, which differ from one another only by the noise disturbances ε_A and ε_B.

Alice and Bob can then extract a binary codeword corresponding to θ by separating the signal space into a predetermined number of quantization levels, which in order to account for the presence of thermal noise disturbances may also make use of rejection zones, as illustrated in Figure 23. Mathematically, let L denote the number of quantization levels and ψ denote the width of angular guard zones at the edges of each quantization slot. Then, Alice and Bob calculate the quantities

$$s \triangleq \left\lfloor \frac{\hat{\theta} \cdot L}{2\pi} \right\rfloor,\ s^{+} \triangleq \left\lfloor \frac{(\hat{\theta}+\psi) \cdot L}{2\pi} \right\rfloor \text{ and } s^{-} \triangleq \left\lfloor \frac{(\hat{\theta}-\psi) \cdot L}{2\pi} \right\rfloor, \tag{18}$$

where $\lfloor x \rfloor$ denotes the largest integer not exceeding x.

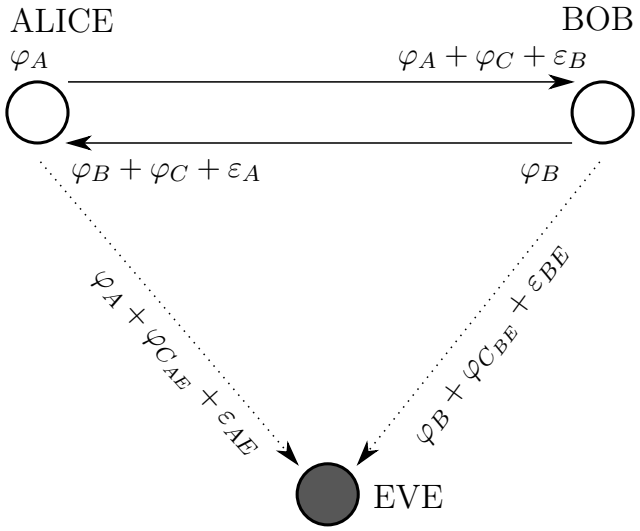


Figure 22: Illustration of the phase-based SKG scheme between Alice and Bob, in the presence of Eve.

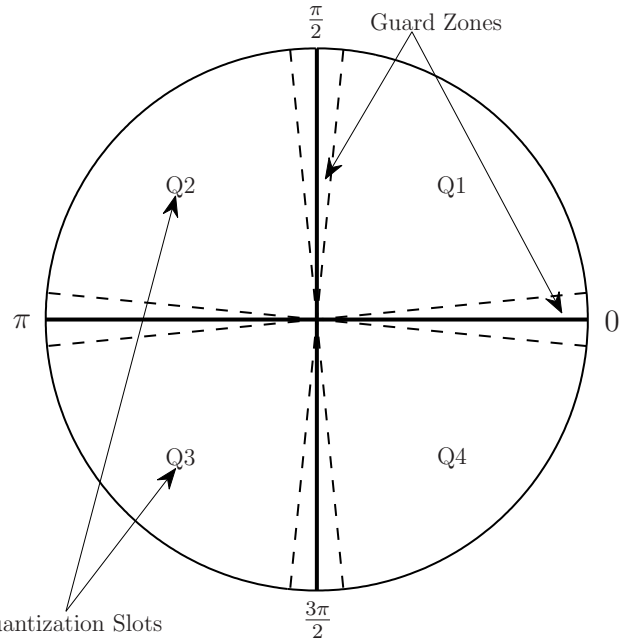


Figure 23: Signal space quantization with $L = 4$ levels and rejection zones of width $\psi = \pi/32$.

In possession of these quantities, Alice and Bob decide on a binary sequence (*symbol*) corresponding to $\hat{\theta}$ via

$$c = \begin{cases} \mathcal{B}(s) & \text{if } s = s^{+} = s^{-} \\ \emptyset & \text{otherwise,} \end{cases} \tag{19}$$

where $\mathcal{B}(s)$ yields the binary representation of s and $c = \emptyset$ indicates that the sample has been rejected.

For reasons that will be soon clarified, the scheme outlined above is implemented over a set of carefully selected frequencies $\mathcal{F} \triangleq \{f_1, f_2, \cdots, f_M\}$, such that for each channel use not one but a set of symbols $\mathcal{C} \triangleq \{c_1, c_2, \cdots, c_M\}$ are simultaneously obtained, which in turn are combined into a single component of the secret key (*codeword*)

$$k = c_1 \otimes c_2 \otimes \cdots \otimes c_M. \tag{20}$$

Assuming that Eve has full knowledge of the SKG scheme described in the previous subsection, she can attempt to build its own copy of the secret key. To this end Eve can construct her own estimate of θ by combining her observations, namely $\hat{\theta}\big|_E = \hat{\varphi}_A\big|_E + \hat{\varphi}_B\big|_E = \varphi_A + \varphi_B + \varphi_{C_{AE}} + \varphi_{C_{BE}} + \varepsilon_{AE} + \varepsilon_{BE}$, which not only is subject to two noise terms, but also a different contribution due to phase rotations, namely $\varphi_{C_{AE}} + \varphi_{C_{BE}}$.

In order to analyze the eavesdropping threat from Eve's best-possible conditions let us temporarily ignore the effect of noise, which benefits Eve in so far as the quality of her estimate of θ becomes then dependent only on the quantity $\Delta\varphi \triangleq |\varphi_C - (\varphi_{C_{AE}} + \varphi_{C_{BE}})| \mod 2\pi$, hereafter referred to as the *phase coherence* of the channels amongst Alice, Bob and Eve.

The conditions under which such criterion can be optimized by Eve lead to the notion of *geometric secrecy* and are discussed in the sequel.

Geometric Secrecy

Given the system model described above, it is clear that in rich multipath environments [62], the phase contributions φ_C, $\varphi_{C_{AE}}$ and $\varphi_{C_{BE}}$ are all uncorrelated, leaving Eve without any strategy to minimize the phase coherence $\Delta\varphi$. Even in the absence of multipath, under free-space propagation conditions and with all channels in Line of Sight (LoS), Alice and Bob could still make use of multiple

antennas to perform random TX/RX beamforming during their exchange [68], again leaving Eve without any viable strategy to decrease $\Delta\varphi$.

The best case for Eve is, therefore, when Alice and Bob utilize single antennas and attempt the random phase exchange in free-space over LoS channels. In this case, the phase rotation introduced by the channel at a given frequency f is deterministic and given by

$$\varphi_C = 2\pi \left[\frac{d_{AB}}{\lambda} - \left\lfloor \frac{d_{AB}}{\lambda} \right\rfloor \right], \tag{21}$$

where d_{AB} is the distance between Alice and Bob and λ is the wavelength corresponding to f.

It is assumed that Eve cannot know exactly the distance d_{AB}, but could use a large number of receivers distributed around Alice and Bob so as to attempt to find a location such that

$$|\varphi_C - (\varphi_{C_{\mathsf{AE}}} + \varphi_{C_{\mathsf{BE}}})| \mod 2\pi < \frac{2\pi}{L}, \tag{22}$$

where, in favor of Eve, we ignore the fact that the introduction of guard zones[2].

Given the above, and denoting the distances from Eve to Alice and Bob respectively by d_{AE} and d_{BE}, the secrecy of the system employing a single frequency is vulnerable to eavesdropping if Eve places herself at a location that satisfies the following inequality

$$d_{AB} + \lambda n - \lambda/L \leq d_E \leq d_{AB} + \lambda n + \lambda/L, \tag{23}$$

where $d_E \triangleq d_{AE} + d_{BE}$ and $n \in \mathbb{N}$.

Notice that inequality (23) defines, for each integer n and each wavelength λ, an elliptical ring around Alice and Bob with internal and external traverse diameters respectively given by $t_n^- \triangleq d_{AB} + \lambda n - \lambda/L$ and $t_n^+ \triangleq d_{AB} + \lambda n + \lambda/L$, as depicted in Figure 24. The areas within each of such rings – marked in grey in Figure 24 – define regions of the space where Eve may be able to obtain a copy of θ, while the areas in between the rings are regions where Eve is unable to do so. These two distinct regions will be referred to as *vulnerability* and *secrecy* regions, respectively.

Inspecting inequality (23), it can be learned that:

- For any λ, with $n = 0$, the ring degenerates to a full ellipse encircling Alice and Bob, which will be hereafter referred to as the *primary* ellipse;
- The thickness of the vulnerability regions is given by $2\lambda/L$, and therefore increases with lower frequencies, but decreases with more quantization levels;
- If $\lambda \to \infty$ (very low frequencies) there is no secrecy region, since the first ring has a negative inner diameter and a large outer diameter;
- If $\lambda \to 0$ (very high frequencies) the secrecy region are sparse, there is, composed by a large number of thin elliptical rings.

These characteristics imply that the geometry of the vulnerability/secrecy regions can be manipulated by the selection of an appropriate set of frequencies. This principle is what we refer to as *geometric secrecy*, and its optimization is the subject of the next subsection.

Optimization of Geometric Secrecy

In order to optimize the geometric secrecy that results from the utilization of multiple frequencies in the phase-based SKG scheme, one must seek the best set of M frequencies $\mathcal{F} \triangleq \{f_1, f_2, \cdots, f_M\}$ that enlarges the first secrecy region to the maximum.

[2]The use of guard zones further enhance security as it requires Eve not only obtain the same quantized estimates of θ but also to know which were rejected by Alice and Bob

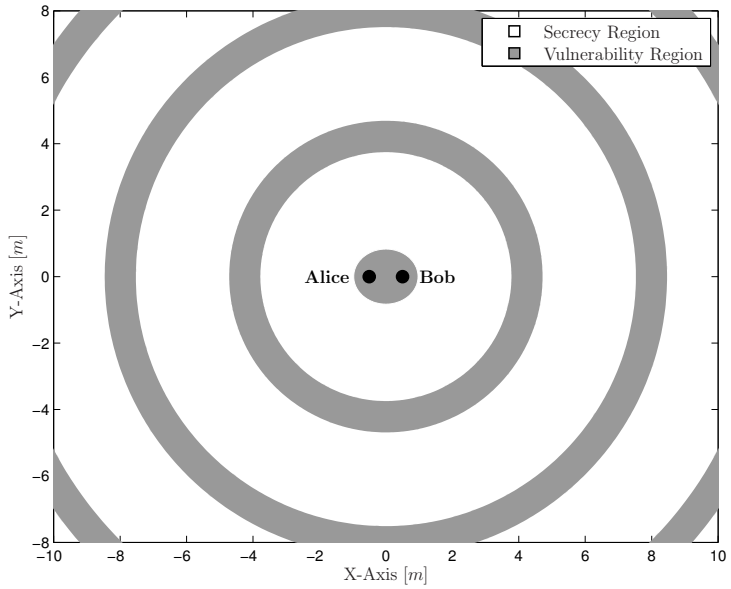


Figure 24: Example of vulnerability and secrecy regions corresponding to a phase-based SKG scheme with $L = 8$ quantization levels and a single frequency $f = 40$Mz.

Referring to Figure 24, start by noticing that in the immediate vicinity of Alice and Bob there is an unavoidable vulnerability region defined by the area common to all the primary ellipses. This implies that the set of optimal frequencies must contain a sufficiently high frequency, whose effect is to degenerate the first vulnerability region to the line[3] between Alice and Bob.

Secondly, recognize that since the SNR experienced by Eve decreases with her distance to Alice and Bob, it is sufficient for optimization purposes to consider only the first elliptical ring besides the primary ellipse.

Therefore, for a given set $\mathcal{F} \triangleq \{f_1, f_2, \cdots, f_M\}$, the optimization of geometric secrecy requires the knowledge of the inner boundary of the first elliptical ring, which is fully determined by the smallest value of d_E associated with a location of Eve, and equivalently defines the first secrecy region. To this end, let $\mathcal{N} \triangleq \{n_1, \cdots, n_M\}$ be a set of non-negative integers and, without loss of generality, order the set $\mathcal{F}$ in ascending order. Then the first secrecy region is determined by the d_E that results from the solution of the following optimization problem

$$\underset{\mathcal{N}\in\mathbb{N}^M}{\text{minimize}} \; d_E \;\; \text{subject to} \left[\begin{array}{c} |d_{AB} + \lambda_1 n_1 - d_E| \leq \lambda_1/L \\ \vdots \\ |d_{AB} + \lambda_M n_M - d_E| \leq \lambda_M/L \end{array}\right], \quad n_M > 0. \tag{24}$$

Due to the fact that the space of solutions is the M-dimensional integer set $\mathbb{N}^M$, this problem does not admit a closed form nor can it be convexized. The problem can, however, be solved using mixed integer optimization tools.

Notice also that the purpose of the constrain $n_M > 0$, which could equivalently be written as $d_E > d_{AB} + \lambda_M/L$, is incorporated in order to exclude the trivial solution associated with the primary ellipse.

With the ability to find the minimum d_E for a given set of frequencies, the optimization of geometric secrecy finally amounts to finding the set $\mathcal{F}$ that maximizes the latter, $i.e.$,

$$\underset{\substack{\mathcal{F}\triangleq\{f_1,\cdots,f_M\} \\ f_{\min}\geq f_m \geq f_{\max}}}{\text{maximize}} \;\; \underset{\mathcal{N}\in\mathbb{N}^M}{\text{minimize}} \; d_E \;\; \text{subject to} \left[\begin{array}{c} |d_{AB} + \lambda_1 n_1 - d_E| \leq \lambda_1/L \\ \vdots \\ |d_{AB} + \lambda_M n_M - d_E| \leq \lambda_M/L \end{array}\right], \quad n_M > 0. \tag{25}$$

[3]For obvious practical reasons, it is assumed that Eve cannot place itself anywhere in between Alice and Bob.

Numerical Example

In this subsection we provide numerical examples that illustrate the efficacy of the optimized geometric secrecy resulting from the phase-based SKG method described above. First, however, let us remark that the *relative* bandwidth of the system has great impact on the optimization of geometric secrecy, as can be inferred from the discussions in subsections 3.1.2.4.1 and 3.1.2.4.1. Specifically, it is desirable that the optimum set $\mathcal{F}$ contains both low and high frequencies simultaneously.

This requirement could be inconvenient if the system is implemented directly. Fortunately however, the linearity of the formulation ensures that, instead, the scheme can be implemented in a differential form, emulating base-band frequencies while operating in usual frequency bands.

To this end, given a set $\mathcal{F} = \{f_0, f_1, \cdots, f_M\}$, consider its baseband equivalent $\mathcal{F}' \triangleq \{f_1 - f_0, \cdots, f_M - f_0\} = \{f'_1, \cdots, f'_M\}$ and let the phases $\Theta \triangleq \{\theta_0, \theta_1, \cdots, \theta_M\}$ measured at the frequencies in $\mathcal{F}$ be correspondingly translated to $\Theta' \triangleq \{\theta_1 - \theta_0, \cdots, \theta_M - \theta_0\} = \{\theta'_1, \cdots, \theta'_M\}$. Then, all the above holds by simply replacing λ_i's with the corresponding equivalent wavelengths $\lambda'_i \triangleq \frac{c}{\lambda_i - \lambda_0}$.

Proceeding this way, and performing the optimization described in equation (25) over the most common the Industrial, Scientific and Medical (ISM) bands, and for $d_{AB} = 1$m, the results shown in Table 5 were obtained. In these examples, the different M optimum frequencies are taken from sets of equi-spaced frequencies with separation Δf.

The results show that as the number and the granularity of the frequencies employed in the system increases, the farther away Eve is pushed. In fact, in the case of differentially implemented systems, it can easily be shown that the optimum d_E^* can be upper-bounded by $d_{AB} + \lambda_M$, which for the examples considered is given by $d_E^* \leq 151$ for $\Delta f = 2$, $d_E^* \leq 301$ for $\Delta f = 1$, and $d_E^* \leq 601$ for $\Delta f = 0.5$. These bounds are indeed very close to those obtained and shown in Table 5.

Table 5: Optimized Geometric Secrecy in ISM Bands

ISM Band (GHz)	Δf (MHz)	L	M	d_E^* (m)
0.9020-0.9280	2	16	2	150.2789
0.9020-0.9280	1	8	3	299.5577
0.9020-0.9280	0.5	4	6	598.1154
2.4000-2.4835	2	16	3	150.7713
2.4000-2.4835	1	8	4	300.5482
2.4000-2.4835	0.5	4	7	600.1018
5.7200-5.8750	2	16	3	150.8750
5.7200-5.8750	1	8	4	300.7500
5.7200-5.8750	0.5	4	7	600.5000

3.1.2.4.2 Performance Analysis

In this section we derive analytically the performance of the proposed system: having previously focused on the techniques to make Eve unable to correctly decode the secret key, now we are interested in estimating the theoretical secrecy achievable between Alice and Bob.

Codeword Error Probability

Recalling the basic mechanism depicted in eq. (19), any transmission has three possible outcomes: i) agreement between Alice and Bob over the binary symbols, ii) disagreement and iii) rejection of the symbol. Any of these possible outcomes, for a generic symbol, is characterized by a probability, namely *primary probability of agreement* $\mathbf{P}'_a$, *primary probability of disagreement* $\mathbf{P}'_d$ and *probability of rejection* $\mathbf{P}'_r$, with $\mathbf{P}'_a + \mathbf{P}'_d + \mathbf{P}'_r = 1$. Consequently the *symbol agreement probability* $\mathbf{P}_a$,

(the agreement likelihood for a single symbol, computed only over the valid transmissions) can be computed as follows:

$$\mathbf{P}_a = \frac{\mathbf{P}_a^{'}}{\mathbf{P}_a^{'} + \mathbf{P}_d^{'}}. \tag{26}$$

The same equation, with $\mathbf{P}_d^{'}$ in place of $\mathbf{P}_a^{'}$ at the numerator, can be used to derive the *symbol error probability* $\mathbf{P}_d$. Alice and Bob in order to reach an agreement on the codeword k, must agree on all the M symbols. The *codeword agreement probability* is consequently determined by

$$\mathbf{P}_{\mathsf{A}} = (\mathbf{P}_a)^M . \tag{27}$$

and the *codeword error probability* is obviously its complementary $\mathbf{P}_{\mathsf{E}} = 1 - \mathbf{P}_{\mathsf{A}}$.

Primary Probabilities for Medium and Low SNR

For medium and low SNR we model the estimated phase θ as follows:

$$\hat{\theta} = \theta + \varepsilon, \tag{28}$$

where the estimation error ε and consequently the $\hat{\theta}$'s are Tikhonov-distributed [69] random variables with mean θ and their Probability Distribution Function (PDF) is given by:

$$f_t(\hat{\theta}; \theta, \gamma) = \frac{\exp(\gamma \cos(\hat{\theta} - \theta))}{2\pi I_0(\gamma)}. \tag{29}$$

where γ expresses the SNR of the system and $I_j(\cdot)$ is the j-th order Bessel function of the first kind.

It is now important to determine the probability $D_t(\theta, \phi_1, \phi_2, \gamma)$ that, given the transmission of θ at the SNR of γ, its estimate $\hat{\theta}$ falls within the interval $[\phi_1, \phi_2]$. Unfortunately the Cumulative Density Function (CDF) associated to the Tikhonov distribution is not analytic, but it can be expressed starting from the following indefinite integral of its PDF:

$$\int f_t(t; \theta, \gamma)\, dt = \frac{1}{2\pi}\left(t + \frac{2}{I_0(\gamma)} \sum_{j=1}^{\infty} I_j(\gamma) \frac{\sin[j(t-\theta)]}{j} \right), \tag{30}$$

We can now obtain:

$$\begin{aligned} D_t(\theta, \phi_1, \phi_2, \gamma) &= \int_{\phi_1}^{\infty} f_t(t; \theta, \gamma)\, \mathrm{d}t - \int_{\phi_2}^{\infty} f_t(t; \theta, \gamma)\, \mathrm{d}t \\ &= \frac{\phi_2 - \phi_1}{2\pi} + \frac{4}{I_0(\gamma)} \sum_{j=1}^{\infty} \frac{1}{j2\pi} I_j(\gamma) \\ &\quad \cdot \cos\left(\frac{j}{2}(\phi_1 + \phi_2 - 2\theta)\right) \sin\left(\frac{j}{2}(\phi_2 - \phi_1)\right). \end{aligned} \tag{31}$$

Symbol agreement happens when the quantization levels selected by Alice and Bob coincide, i.e. when $\boldsymbol{Q}(\hat{\theta}\big|_A) = \boldsymbol{Q}(\hat{\theta}\big|_B)$. This probability, for a given pair (θ,γ), is:

$$\sum_{i=1}^{L} \mathsf{Pr}\{\boldsymbol{Q}(\hat{\theta}\big|_A) = i \,|\, \theta\} \cdot \mathsf{Pr}\{\boldsymbol{Q}(\hat{\theta}\big|_B) = i \,|\, \theta\}. \tag{32}$$

It is possible to insert eq. (31) into the previous expression and average over θ, assumed uniformly distributed in the signal space, to get the primary probability of agreement:

$$\mathbf{P}_a^{'} = \frac{1}{2\pi} \sum_{i=1}^{L} \int_0^{2\pi} D_t^2(\theta, (i-1)\frac{2\pi}{L} + \psi, i\frac{2\pi}{L} + \psi, \gamma)\, \mathrm{d}\theta. \tag{33}$$

Similarly, the primary symbol disagreement probability $\mathbf{P}_d^{'}$, i.e. the probability that $\boldsymbol{Q}(\hat{\theta}\big|_A) \neq \boldsymbol{Q}(\hat{\theta}\big|_B)$, is

$$\mathbf{P}_d^{'} = \frac{1}{2\pi} \sum_{i=1}^{L} \sum_{\substack{j=1 \\ j \neq i}}^{L} \int_0^{2\pi} D_t(\theta, (i-1)\frac{2\pi}{L} + \psi, i\frac{2\pi}{L} - \psi, \gamma) \cdot D_t(\theta, (j-1)\frac{2\pi}{L} + \psi, j\frac{2\pi}{L} - \psi, \gamma)\, \mathrm{d}\theta. \tag{34}$$

Closed-form solution of eq. (33) and eq. (34) are available but not presented here for sake of brevity.

High SNR Derivation

For high values of SNR the Tikhonov distribution can be accurately approximated by a Normal distribution [70, 71]:

$$f_t(\hat{\theta}; \theta, \gamma) \approx \frac{1}{\sigma(\gamma)\sqrt{2\pi}}\, e^{-\dfrac{\hat{\theta}-\theta}{2\sigma^2(\gamma)}}, \tag{35}$$

where the variance $\sigma(\gamma)$ is function of the SNR:

$$\sigma(\gamma) = \sqrt{2(1 - I_1(\gamma)/I_0(\gamma))}. \tag{36}$$

Given the following generic tight bounds on the ratio of Bessel functions [72],

$$\frac{\gamma}{i + \frac{1}{2} + \sqrt{\gamma^2 + (i+\frac{3}{2})^2}} \leq \frac{I_{i+1}(\gamma)}{I_i(\gamma)} \leq \frac{\gamma}{i + \frac{1}{2} + \sqrt{\gamma^2 + (i+\frac{1}{2})^2}}, \tag{37}$$

for high SNR ($\gamma \gg 1$) it is possible to obtain for eq. (36) a very accurate closed-form approximation.

$$\sigma(\gamma) \approx \sqrt{\frac{2}{2\gamma + 1}}. \tag{38}$$

Eq. (31) now is:

$$\begin{aligned} D_t(\theta, \phi_1, \phi_2, \gamma) \quad &= \tfrac{1}{2} \operatorname{erfc}\left(\tfrac{\phi_1 - \theta}{\sqrt{2\sigma^2(\gamma)}}\right) - \tfrac{1}{2} \operatorname{erfc}\left(\tfrac{\phi_2 - \theta}{\sqrt{2\sigma^2(\gamma)}}\right) \\ &+ \tfrac{1}{2} \operatorname{erfc}\left(\tfrac{\phi_2 - \theta - 2\pi}{\sqrt{2\sigma^2(\gamma)}}\right) - \tfrac{1}{2} \operatorname{erfc}\left(\tfrac{\phi_1 - \theta - 2\pi}{\sqrt{2\sigma^2(\gamma)}}\right), \end{aligned} \tag{39}$$

with $\operatorname{erfc}(\cdot)$ denoting the complementary error function, while the expressions for $\mathbf{P}_a^{'}$ and $\mathbf{P}_d^{'}$ can be still obtained using eq. (33) and eq. (34) respectively.

Performance

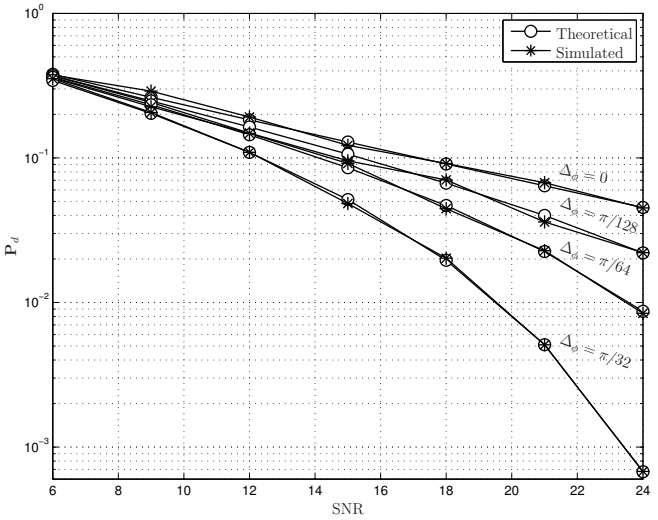
In this section we validate with simulations the system performance, in term of symbol error probability $\mathbf{P}_d$ and codeword error probability $\mathbf{P}_{\mathsf{E}}$, of the secret key generation between Alice and Bob.

In Figures 25(a), 25(b) and 25(c), we analyzed $\mathbf{P}_d$ for different widths of the guard zone and for different quantization levels ($L = \{4, 8, 16\}$). The simulation curves have been plotted for 1000 errors (i.e. disagreement between Alice and Bob) over the k-th symbol construction, always assuming a Tikhonov phase estimation error on both sides. The theoretical curves have been computed using the Tikhonov error distribution, as in eq. (31), for SNR≤ 20 dB, while for higher values its Gaussian approximation, as in eq. (39), has been employed.

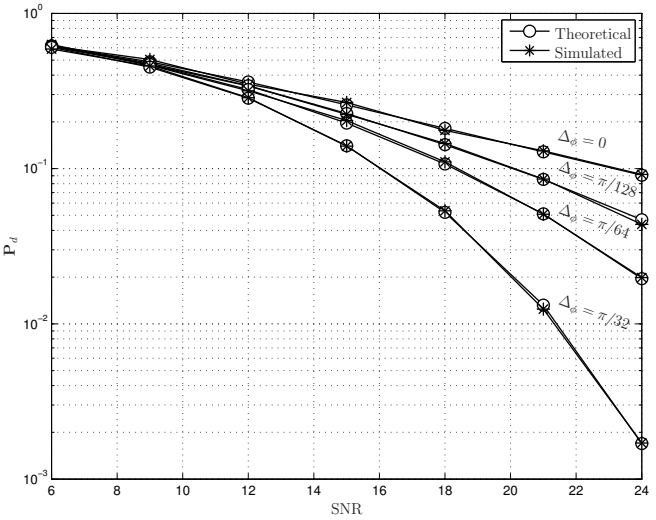
It can be observed how the choice of both L and ψ is conditioned by the SNR. For low high values of SNR the width of guard zones is not enough to prevent symbol disagreement, so their role is negligible. On the contrary, under such circumstances, it is important to use low L to minimise $\mathbf{P}_d$. Furthermore, using many levels for high values of SNR can be helpful, but as a consequence the number of transmission due to rejection can dramatically increase.

If we consider, as example, the case of transmission at 24 dB, it is possible to note that the combination $L = 8$ and $\psi = \pi/128$ has almost the same performance than for $L = 16$ and $\psi = \pi/64$, with $\mathbf{P}_d \approx 0.05$, but the scenario with only 8 levels has a rejection probability $\mathbf{P}_r$ of 4 time less (0.1 versus 0.4) than the case with $L = 16$.

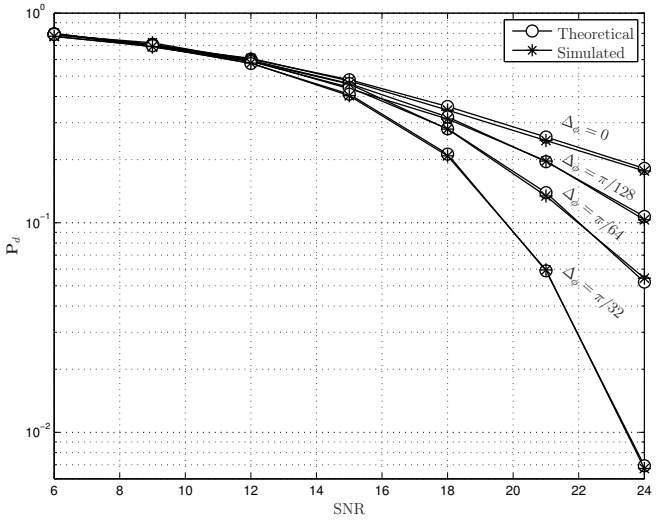
The results illustrated in Figure 25 can finally be combined with the examples depicted in table 5, to evaluate the impact of each parameter (SNR, frequency spacing and number of levels) both on the size of the geometric secrecy and on the codeword error probability $\mathbf{P}_\mathsf{E}$.



(a) Theorerical and simulated $\mathbf{P}_d$ for $L = 4$ as a function of the SNR.



(b) Theorerical and simulated $\mathbf{P}_d$ for $L = 8$ as a function of the SNR.



(c) Theorerical and simulated $\mathbf{P}_d$ for $L = 16$ as a function of the SNR.

Figure 25: System performance compared to theoretical model in terms of symbol error probability for different number of quantization levels L.

In summary, we have addressed the issue of creating dynamically a secret key to encrypt data transmission for NFC and SRC devices. Since the current key exchange schemes, to the best of our knowledge, exploit all the entropy of the communication channel to extract the secret bits, this approach is not anymore suitable for the scenario under consideration, characterized by very short and LoS links between devices.

We therefore have proposed a new, alternative scheme, based on symbol phase exchange, that geometrically determine the region where fully secrecy is assured, irrespective of the number of bits exchanged. Consequently we have extended and validated this model for both low and high SNRs.

3.1.2.5 Secret Key Generation from Sampled IR-UWB Channels

The transmission of sensitive information between Smart Objects requires a secure infrastructure that provides confidentiality, authentication and data integrity. Cryptographic materials such as symmetric secret keys are necessary in order to solve some of these challenges. Classical symmetric key distribution schemes can require specific infrastructure (e.g. certification entities, online servers for key generation and initial secret keys etc.). Decentralized symmetric key generation algorithms using physical layer metrics as a source of common information can be an alternative to such methods [73], [74]. In this section, we present an analysis of the random aspect [75] and secrecy properties [76] of keys generated from sampled IR-UWB signals. The IR-UWB technology, one of the candidates for WSN physical layer, is particularly suitable for physical layer key generation due to its high multipath resolution capabilities, which give access to entropy-rich signals or to precise estimations of the Time of Flight between two nodes.

3.1.2.5.1 Signal model

The IR-UWB received signal at a node can be modeled as: $y(t) = (h * x)(t) + w(t)$ where $h(t)$ is the reciprocal channel impulse response, $x(t)$ is the transmitted pulse and $w(t)$ is the additive white Gaussian noise. The received signals are uniformly sampled at sampling frequency F_s and N samples are collected in an observation window whose overall time duration is $W = (N-1)/F_s$.

3.1.2.5.2 State of the art: single-link key extraction algorithm (POS)

The key generation method proposed in [73] uses adaptive threshold quantization for bit extraction. The phases of the algorithm are the following:

- estimation of the noise level (N_{lev}) used as a stopping rule for the extraction;
- channel probing, filtering and sampling;
- iterative bit extraction from the sampled signal $y[n]$ at each participating node (A or B):
  - compute the first thresholds ($i = 0$) for quantization: $L_0^+ = \max(y[n]) > 0$ and $L_0^- = \min(y[n]) < 0$;
  - at iteration i, apply the operator $pos_i(y[n]) = \begin{cases} 1 & \text{if } y[n] \geq L_i^+ \\ 0 & \text{if } y[n] \leq L_i^- \end{cases}$ to the samples that cross the thresholds, memorize the extracted time indexes n (i.e. temporal positions) in a table P and the corresponding iteration step i in a table I;
  - update thresholds $L_{i+1}^+ = L_i^+ - L_0^+/\delta$ and $L_{i+1}^- = L_i^- - L_0^-/\delta$, with δ a protocol parameter;
  - repeat the last two steps until the noise level is approached within a guard interval depending on δ;
- public discussion involving the exchange of the index tables P_A, P_B followed by the selection of the common indexes and their corresponding bits;

- key correction using a Reed-Solomon code to fix mismatching bits;
- verification of the final keys using a method similar to challenge-response protocols;
- key renewal, which can be done if the key establishment fails, when the key is too short, or periodically for security;

The described bit extraction method has the main advantage of being adaptive with respect to the SNR. If the SNR is relatively high, a larger δ can be used. Nonetheless, the quantization is based only on the polarity of the input signal with respect to the last threshold, independently for each sample/position. This simple polarity-based encoding applied to a deterministic waveform could possibly lead to regular patterns in the extracted bits. An extension of this bit extraction method, which takes into consideration the relationship between the amplitudes of different samples in order to break the pattern regularities, is presented in the next subsection (HIST extension). Separately, we have investigated if it is possible to limit or encode the public information in order to guarantee better secrecy properties with respect to a nearby eavesdropper (BIN and POS$_{ToF}$ extensions). This second study highlights the benefit that can be drawn from location-dependent information, such as the Time of Flight, between the two legitimate peers.

3.1.2.5.3 Proposed improvements of the single-link key extraction algorithm

First, the HIST extension of the POS algorithm [73] concerns the bit extraction phase. The rest of the key generation protocol remains the same. The observation window is split into N_b bins of a given sample lengthΔ_{bin}. These two parameters are considered as public information. The auxiliary bit operations of HIST are defined by $hist_i(pos_{i_k}(y[n])) = \begin{cases} pos_{i_k} & \text{if } k = odd \\ \neg pos_{i_k} & \text{if } k = even \end{cases}, i_k \in I_b$, where $\neg$ is the negation operator and $I_b = \{i_1, i_2, ...\}$ is the set of ordered iteration steps (indexed by k) that correspond to the extracted bits in bin b. The result of applying this kind of post-processing is equivalent to a dynamic encoding of the negative and positive amplitudes as a function of the absolute value and of the sample time index or bin. With POS, the negative amplitudes are always encoded as “0” and the positive amplitudes as “1”, but with HIST the encoding convention varies across the observation window.

Secondly, for the public discussion phase, we propose two alternatives: one based on the limitation of the public information (BIN extension) and another based on the masking of the public information using another reciprocal physical-layer metric, i.e., the Time of Flight (ToF), which can be measured in the context of ranging protocols (POS$_{ToF}$ extension). The bit extraction algorithm is the same as in POS.

A bin is an interval of several samples which is defined beforehand for both sides of the legitimate link (i.e. N_b bins of length Δ_{bin}). Instead of exchanging tables containing the indexes of the extracted samples, A and B exchange tables containing the occupancy of each bin (i.e. the number of samples above the final detection threshold in each bin).

As the previous solution is expected to work in rather high SNR conditions, we suggest another approach for sending information over the public channel by encoding it with the equivalent ToF values (expressed in samples as N_{ToF}). These values are measured independently by each node. During the public discussion phase, the exchanged encoded tables are: $EP_i = (P_i + N^i_{ToF}) \mod W$, with i representing A or B. In order to recover the original index table of the other user, A and B try to decode EP using their own position table and N_{ToF}. We consider two decoding strategies: symmetric (POS^s_{ToF}) when A and B assume that $N^A_{ToF} = N^B_{ToF}$ and asymmetric (POS^a_{ToF}) when A tries to infer N^B_{ToF} based on maximizing the number of common temporal positions.

3.1.2.5.4 Simulation context

The advantages in terms of key length and randomness properties of POS have been highlighted in [73] where the authors used experimental traces from an indoor UWB measurement campaign. In order to extend these results to a typical indoor channel and analyze the random aspect of generated keys, we consider additional signals generated from statistical channel models (CM1 and CM2 according to the IEEE802.15.4a standard [77]). The public discussion strategies are tested on ray tracing signals [78] that account for the spatial correlation that could potentially harm the secrecy of the key generation scheme: an attacker C in the vicinity of B measures the channel between A and himself ($y_C(t)$) and the ToF on the same link. C then obtains a bit sequence from the quantization of $y_C[n]$ that he uses along with the public information to guess the key generated between A and B. In case of a missing position in his table generated from $y_C[n]$, C will randomly guess the corresponding bit.

Also, in these studies, a more realistic synchronization method for the signal acquisition step is considered. Instead of using a trigger signal as in [73], the observation window starts when the received signal crosses a certain fixed threshold. The measured key agreement and bit agreement rates will take into account the resulting synchronization errors accordingly. Finally, the simulated sampling frequency is of the same order as that of the experimental tests in [73] (i.e. 20 GHz) and the observation window is fixed at 100 ns.

3.1.2.5.5 Simulation results (HIST)

For reference purposes, the algorithms POS and HIST have been tested firstly in a noiseless scenario, in which generated bits at A and B are always identical (i.e. with a key agreement rate of 1). The goal is to evaluate their intrinsic randomness properties, which depend mostly on the transmitted waveform and on the multipath channel. Next, the performances in terms of reciprocity are discussed for various SNR values.

The NIST statistical suite citeNIST allowed the investigation of various key characteristics. It is important to keep in mind that such tests cannot say whether a sequence is random in the absolute. They can only show defects in the random nature by pointing out when certain keys are prone to a deterministic behavior. A key is said to pass the test when its computed p-value (a value between 0 and 1) is higher than 0.01. Nonetheless, a given set of keys pass a certain test if a very large proportion of the keys pass the test (pass rate) and if the p-values of the keys are uniformly distributed in the (0,1) interval. From Table 6, we can observe that POS passes the frequency and random walk tests, but fails the oscillations and pattern tests for almost all the keys. It also performs badly from the point of view of the p-value uniformity in all the tests. On the contrary, HIST shows an improvement in the uniformity of p-value over all the tests and manages to pass them with a high ratio.

Table 6: NIST results on POS and HIST algorithms

Test	POS		HIST	
	Pass ratio	p-value variance	Pass ratio	p-value variance
Block frequency	100%	3	98%	80
Random walks (cum. sums)	100%	10	97%	89
Oscillations (runs test)	1%	NA	90%	90
Patterns (serial test)	0%	NA	92%	79

In the case of noisy signals, the level of reciprocity measured by the key agreement rate (i.e. the proportion of identical keys after reconciliation) and by the mean bit agreement rate (i.e. the ratio of common bits between the two parties before reconciliation) is expected to be degraded for HIST

because of a more complex bit encoding technique. Figure 26 shows the limitations of HIST in terms of generating successful minimum 64-bit keys especially for high δ and Δ_{bin} values. POS does not experience the same phenomenon because it only encodes the polarity of the waveform, making it more robust to noise. A possible direct solution for HIST would be to increase the correction capacities of the Reed-Solomon code with the trade-off of secrecy with respect to a potential attacker. However, it can be observed that HIST maintains a reasonably high bit agreement ratio (Figure 27) while providing a dynamic amplitude encoding. The number of differently encoded bits between POS and HIST, employed to measure the diversity offered by HIST, remains relatively high even for small Δ_{bin} values (e.g. between 20 and 85 for Δ_{bin} = 4 samples).

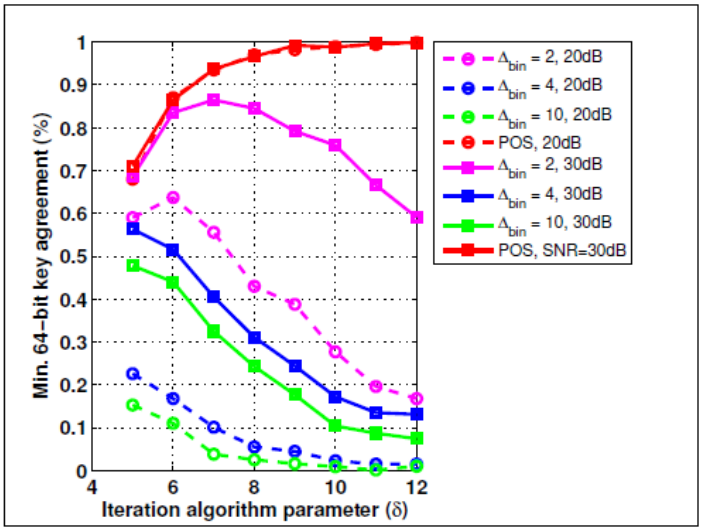


Figure 26: Key agreement for POS and HIST

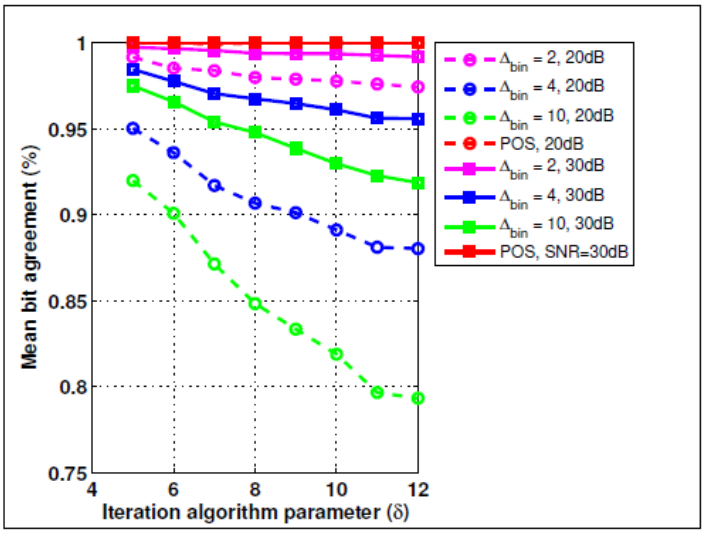


Figure 27: Mean bit agreement for POS and HIST

3.1.2.5.6 Simulation results (BIN and POS$_{ToF}$)

Preliminary tests on the mean bit agreement between legitimate users indicate that the POS$^{a}_{ToF}$ strategy is not a reliable key generation procedure: maximizing the number of common temporal

positions is not a good criterion to estimate the exact ToF of the other legitimate user because it leads to different key lengths.

Additionally, in order to investigate the secrecy of the key generation algorithms, we compute the mean illegal bit agreement ratio between the bit sequences generated at B and C. The ideal value would be 0.5 which corresponds to a random guess of the attacker. As shown in Figure 28, both BIN and POS^s_{ToF} improve the mean illegal bit agreement with respect to the conventional POS method. Moreover, POS^s_{ToF} achieves a mean illegal bit agreement of 0.5 over all parameter values and at different SNR [76]. However, Figure 29 shows that the key agreement rate is degraded for our proposals (though not significantly for POS^s_{ToF}). This phenomenon can be seen as the trade-off to be paid in return for better secrecy properties.

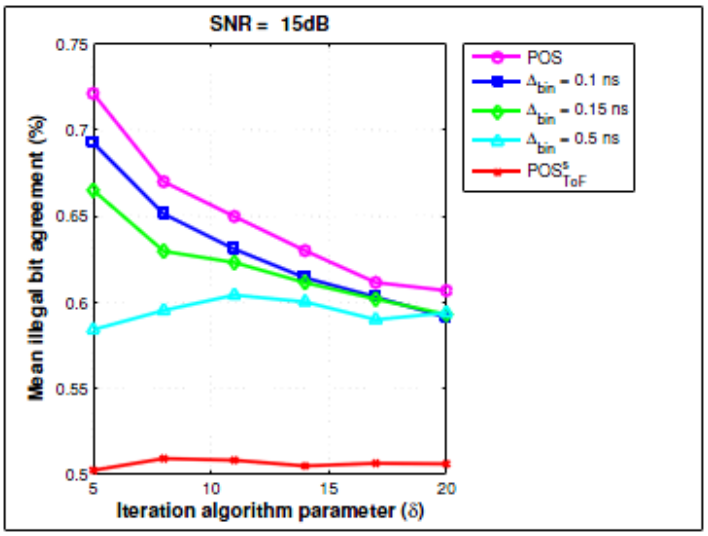


Figure 28: Mean illegal bit agreement for POS, BIN and POS_{ToF}

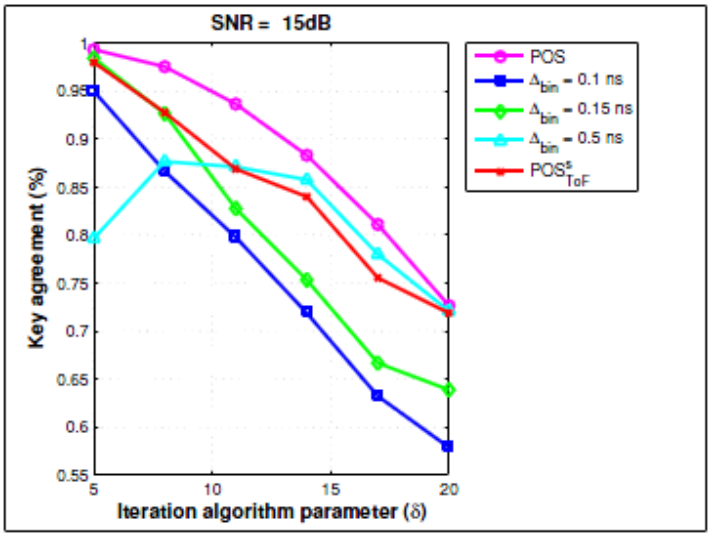


Figure 29: Key agreement for POS, BIN and POS_{ToF}

3.1.2.5.7 Conclusion

In the previous sub-sections, we have presented several extensions and improvements of an existing key generation scheme. Among the proposed reconciliation methods, BIN can be easily

adapted to the SNR conditions by varying the bin size but it is less efficient at large bin sizes because it implies dropping bits by large groups. POS^{s}_{ToF} proves to be efficient since it only hides the publicly exchanged information by using the side ToF information as a mask.

The main advantage of HIST is to improve the random patterns of the keys suffering from deterministic characteristics of the input signal, such as the pulse waveform that contributes to the correlation between adjacent samples and therefore between bits of the key. The drawback is that it can be effectively used in rather high SNR regimes.

3.1.2.6 Testing end-to-end Security at Application Level

This section describes some test instantiations of the Security Framework. Each subsection describes in detail the use cases and gives issues/problems that have been highlighted during the experimentation.

3.1.2.6.1 Gemalto Resource Consumer Test Application

A part of the BUTLER Security Framework, Gemalto has developped a Generic Resource Consumer application for testing any resource provider that are reachable on the INTERNET. The application is implemented as a web server using a web browser (user agent). The application web server accesses directly the resource provider. The access to the resource is performed by the web server which is a client of the Resource Provider. In this application, the session keys are never sent to the user agent (the web browser) but kept at the web server. The application web server is registered at the Trust Manager. Thanks to the Trust Manager administrator for providing the application authentication credentials. The application displays a (decrypted) resource result and assumes the result is a ASCII string - possibly a JSON string. The application does not assume any data semantic.

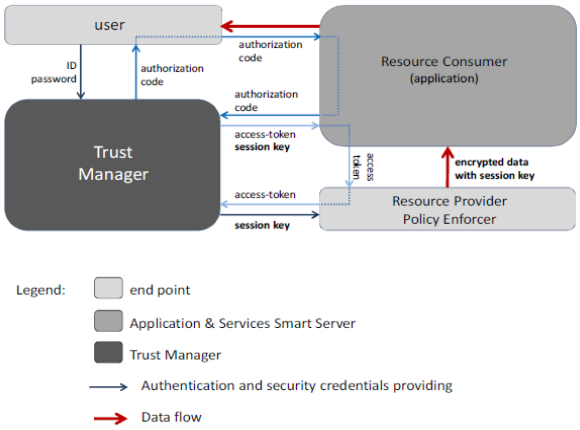


Figure 30: Generic Application and its Environment.

The Figure 30 presents the application in its environment and the exchanged messages. The Generic Application highlights the main features of the Trust Manager according the application point of view.

1. Federation of identities. The Trust Manager (TM) supports user authentication. On return, the TM generates an opaque handle for the application. The application can use such handle to identify the TM authenticated user in the application environment. The opaque handle does not include any user attributes. The Trust Manager needs user consent for providing the opaque handle to the application.

2. Authorization. The Trust Manager supports generation of access token and security material for the application. The Trust Manager exposes two APIs. The first one needs user consent for providing the access token to the application. The second one uses the user authentication credentials. This test application presents the user consent feature.
3. Security. The access token is associated to security material. The security material is used by the application for accessing the resource. The application uses the security material:
   (a) to authenticate to the resource. Using this feature, the resource can check that the calling application is the one which received the access token.
   (b) to encrypt and sign command data. The command data shall be signed and encrypted using the related keys of the security material.
   (c) to check signature and decrypt resource result. The result shall be signed and decrypted using the related keys of the security material.

Application User Interface

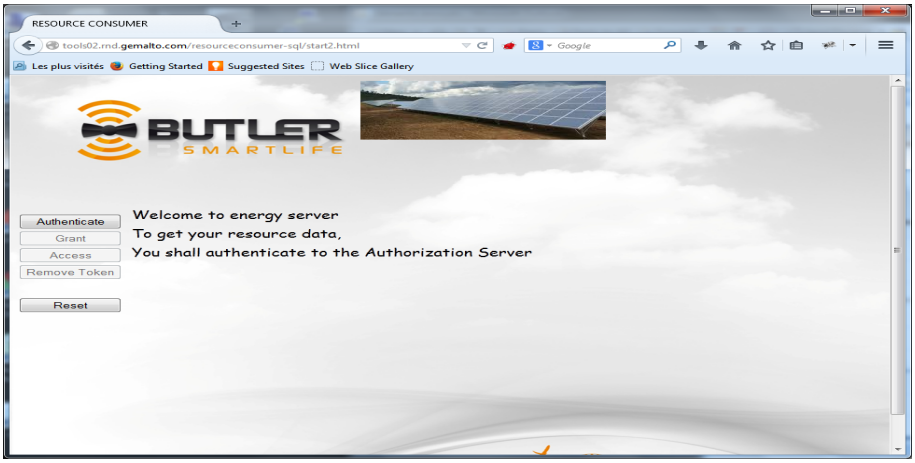


Figure 31: User Interface.

The Figure 31 presents the user interface of the application. The user interface consists on a menu frame including Authentication, Grant, Access and Remove Token buttons.

1. **Authentication.** The user shall authenticate to the application. This application does not implement any user authentication, but relies on the Trust Manager. Once the user is authenticated, the Trust Manager return on opaque handle. This opaque handle does not give any information on the user but uniquely identifies the user in the application domain (thanks to the Trust Manager / Authentication end point). Using the opaque handle, the application implements a user profile that encompasses only the last used resources and the related access token. As presented in Figure 32 and in Figure 33 the Trust Manager authenticates the user and requests user consent to return the opaque handle (Authentication Token) to the application.
2. **Grant.** In Figure 34, the user enters the required resource url and the actions he wants to perform. If the user previously entered a resource url, the application proposes the last used resource url of authenticated user. On Grant, the application authenticates to the Trust Manager (thanks to the application credentials), the Trust Manager checks that the user is allowed to retrieve an access token. If not allowed, the Trust Manager returns an error. If allowed the Trust Manager requires user consent for returning an access token to the application; the

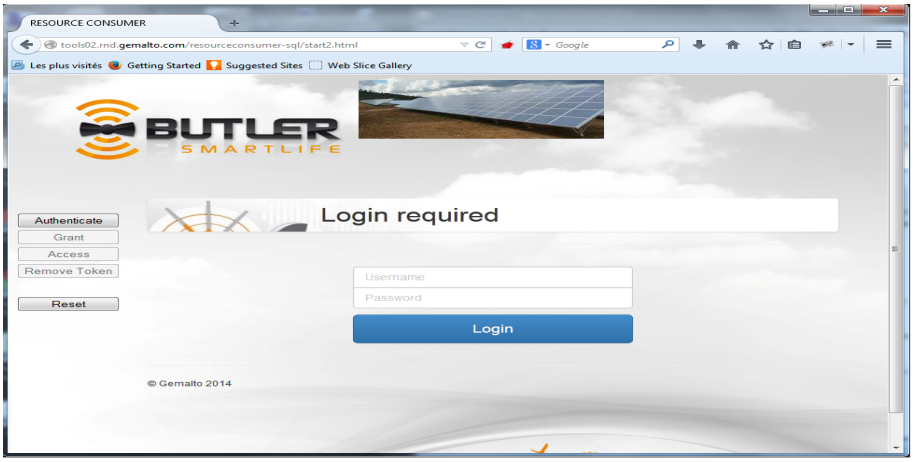


Figure 32: Trust Manager Login Required.

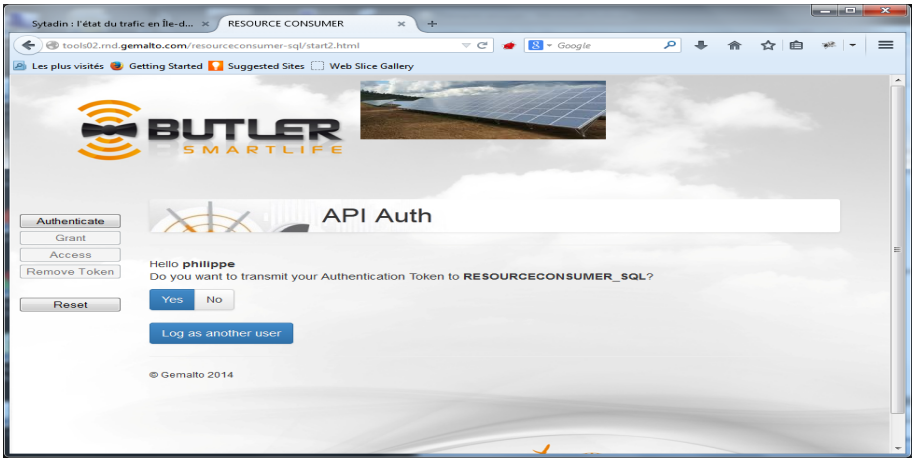


Figure 33: Trust Manager User Consent for Authentication Token/ Opaque Handle.

corresponding user interface is presented in Figure 35. The user authorizes the generation of the access token, the application retreives the Access Token (and security material). The Security materials are kept at the application web server.

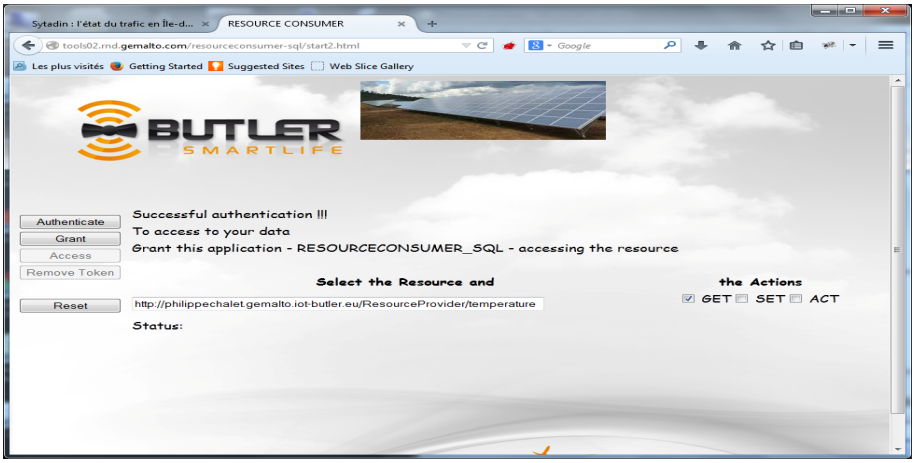


Figure 34: Resource Selection.

3. **Access.** The access button is validated. The user interface displays:
   (a) The first part of the Access Token (for information).

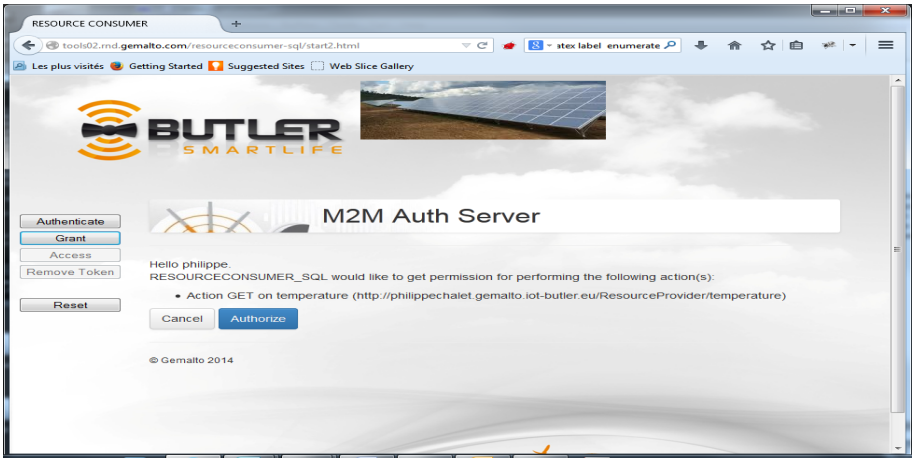


Figure 35: Trust Manager User Consent for Access Token.

(b) The HTTP method to be used to access the resource. The allowed methods depend on the resource API.

(c) the content type. The allowed content types depend on the resource API.

(d) parameters. It is a string compliant with the content type. The application does not check the input syntax (the checking shall be performed by the resource. The parameter can be empty. If not empty, the parameter will be encrypted by the application web server.

(e) Result. Upon access, the application displays the decrypted response. In Figure 36, the resource “http://philippechalet.gemalto.iot-butler.eu/ResourceProvider/temperature” returns the value of the parameter “p1” , for information the value of the “application-private-identifier” (it is Security Protocol related data that can be retrieved by resource), the “body” of the content received by the resource (after decryption). It has to be noted that, in this example, the resource can be accessed using HTTP. The security of the access is implemented by the Security Protocol - this feature allows security without relying on server certificate which can be difficult (or too costly) to deploy in the resource software.

Issues hightligthed by this demonstration

A potential issue concerns the architecture of the demonstration: the application is implemented as a web server, it means that the user implicitly trusts the application for running on its behalf. The demonstration application (web server) does not exploit any user or resource data but only displays the results. The access tokens and associated security material are registered in application web server database, and therefore the application may act as a fraudulent application, for instance selling clear data to another server which is not under control of the user. Nevertheless, it has to be noted that the application does not have the user credentials (in this case, the user credentials are user name and password). At application level, the user is identified with its opaque handle; in consequence, the application cannot send/sell clear data as user data and the privacy requirement is respected.

The application uses the main feature of the Security Framework. The main issues are related to the user interface (also the look & Feel but it is out of scope this study) which is complex. It is due to the implementation of the user consent.

1. User consent for providing the user authentication opaque handle. see example in Figure 33
2. User consent for providing the access token and security material. see example in Figure 35

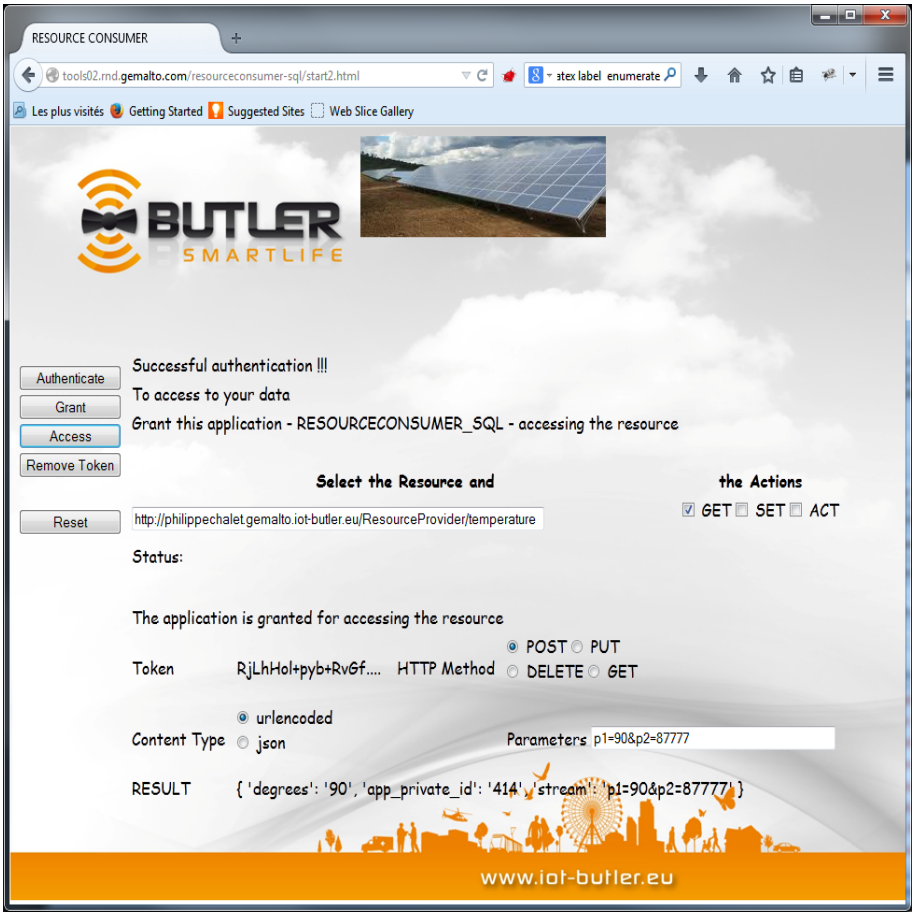


Figure 36: Accessing Resource.

For privacy reasons, these user consents are useful to perform user control. In case the user fully trusts the application web server, the user can provide the user credentials to the application web server. The user consents are not implemented at the user interface, the user trusts the application web server which, in consequence, has ability to retrieve access token for any user's owned resources (or resources on which the user has the required permissions). Nevertheless, at application level, the user is identified with his/her user credentials; in consequence, the application may act as a fraudulent application by sending/selling clear user data as user data and, in consequence, the privacy requirement is not respected.

3.1.2.6.2 Using Gemalto Communication Module Device

For the BUTLER project, Gemalto has developed a BUTLER device software running in Gemalto communication module. Gemalto M2M modules cover a spectrum of wireless technology standards including CDMA, GSM and UMTS with LTE forthcoming. A wide range of features, capabilities and innovative technologies such as Java, GPS and SIM Access Profile are available to meet specific needs and provide greater integration flexibility.

The modules are designed to be embedded in customer terminal/gateway fulfilling customer requirements. Currently, Gemalto offer includes some generic simple terminals that can fulfill large customer requirements but does not address specific customer requirements for instance none of Gemalto M2M Terminals support Personal Area Wireless Network.

For the project, we used a Terminal (see Figure 37) embedding a SIM card. The Terminal device embeds a temperature sensor and a light (on/off) sensor. The device can initiate data communica-

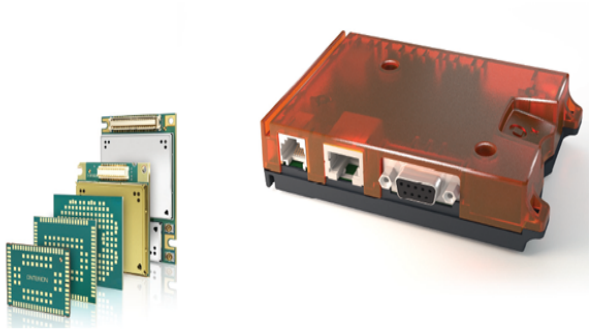


Figure 37: Gemalto Communication Device

tion on the Internet and receives SMS. With our Mobile Network Operator, as usual, the terminal does not have public IP address and therefore is not directly accessible from the Internet.

For the purpose of the demonstration, the terminal is configured as a Resource Provider. It exposes the following resources:

1. Get-Temperature. on request the terminal functionally shall check the received token and return the temperature of the room in JSON format; {“degree”:“value” }.
2. Configure as Resource Consumer. The device can play the role of application to give the temperature of the room to an entity that will use the information. For this purpose, it needs to be dynamically configured. On secure request, the terminal securely retrieve the Resource Consumer configuration data and the Service Platform URL to be used for sending data.
   {
   {“serviceplatform”: { “url”: “url of the service platform”, “security-type”: “type”},
   {“resourceconsumer”: { “client-id”: “xxxxx”, “client-password”: “yyyy”, “user-id”: “uuuuu”, “user-password”: “ppppp”, “application-private-identifier”: “.....” },
   { “resourceprovider”: { “url of the resource provider”: “....”, “period-in-seconds”: “....nseconds” }
   }.
   (a) Service Platform. The configuration data consists of the “url of the plafform” and the security type.
      i. Security user/password. The device shall authenticate to the plaform by sending its user-id and user-password (see below).
      ii. BUTLER Authorization. The device shall consider the plaform as a Resource Provider and retrieve a token to send data to the service platform - thanks to the Resource Consumer configuration.
   (b) Resource Consumer configuration.
      i. client-id. the identifier of the application registered in the Trust Manager.
      ii. client-password. the password of the application registered in the Trust Manager.
      iii. user-id. the identifier of the user - the device will act on behalf of the user identified by user-id.
      iv. user-password. The user credential.
      v. application-private-identifier. The private identifier to be used for retrieving an access token. This value can be used to identify the device for all resource provider using the device data. The application-private-identifer can be user when the device (resource consumer) and the resource provider runs in the same eco-system.
      vi. Use this Resource Provider. The device shall push every n-seconds the temperature for this specified resource provider.

3. Remove this Resource Provider. The device at the Trust Manager provides an API to remove the sending of data to the resource provider. The RP is specified in JSON: { “resourceprovider”: “the url of the resource provider” }
4. Add/Update this Resource Provider. The device provides an API to update (or add) a destination resource provider. The device shall push every n seconds the temperature for this specified resource provider. The resource provider is specified as follow.
   { “resourceprovider”: { “url of the resource provider”: “....”, “period-in-seconds”: “....nseconds” }

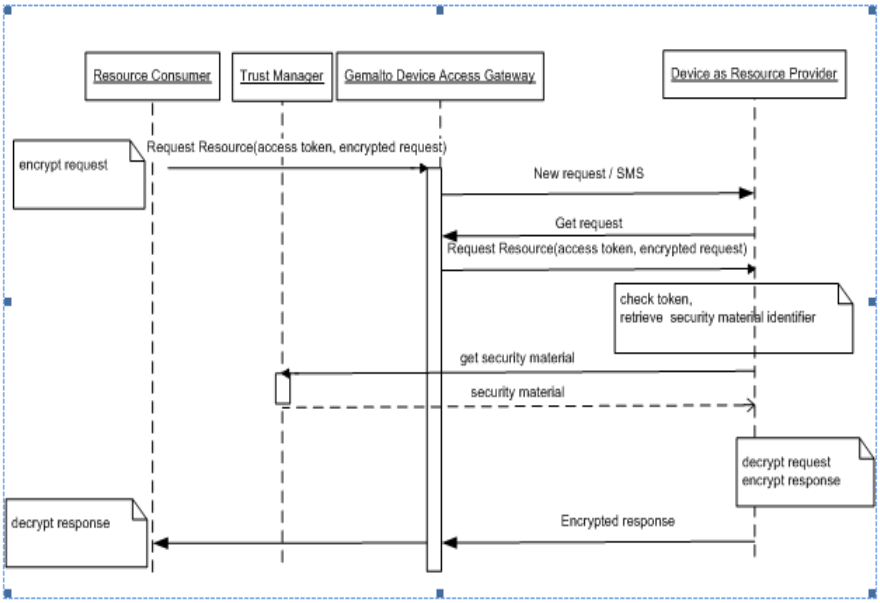


Figure 38: Gemalto Communication Device Demonstration Flow

Accessing the Device as Resource Provider.

The Figure 38 presents the message flow for accessing the device as Resource Provider.

1. Building the request. The application previously got an access token to be used to access the resource. It builds the request following the security protocol requirements.
2. Sending the request. The device is not accessible directly on the Internet. The application connects to the “Gemalto Device Access Gateway” with the following parameter:
   (a) Phone Number of the device.
   (b) Access token and encrypted data.

   The Gateway could be protected as a Resource Provider providing the Gateway API. For simplicity purpose, the use case is usual and does need to be presented here. On reception, the Gemalto Device Access gateway generates a random session number and sends a SMS to the device. The SMS includes the URL to be used by the device to retrieve the access token and encrypted data.
3. Receiving SMS. The device receives the SMS and connects to the url of the gateway. The gateway returns the access token and the encrypted data.
4. Checking access token and retrieving security material. On reception of the access token and the encrypted data, the device checks the access token (using access token credential), on success, it authenticates to the Trust Manager to retrieve the security material associated to the token. It decrypts the request, encrypts the response and returns the response to the gateway.

5. Finally. The gateway returns the response to the application.

Issue - Accessing the Device as Resource Provider.

The message flow works properly but is not very efficient. The gateway uses SMS to reach the device which could be costly and may spend too much time. We may rely on “push” mechanism but the “push” mechanism is also costly in term of data connection. “Accessing the device as Resource Provider” is useful for performing secure actuation of the device.

Device running as Resource Consumer

The device can run as resource consumer. In the Security Framework parlance, the device runs an application which can use some resources. In many case, the device sends resource data to entity which used it. For instance, the position of the device can be used to localize the device. In this sense the resource that uses the device data is an actuactor - it acts on device data.

The Figure 4, summarizes the architecture. For providing end-to-end security, Daisy exposes a resource at the Trust Manager.

Donald’s application authenticates to the Trust Manager to retrieve a token for accessing the Daisy’s exposed resource. Donald’s device sends secure device data to the Service Platform - for instance the User Profile Smart Server or the Context Manager Smart Server. As usual, the secure data consists of the access token, the application authentication data and the business encrypted data. Daisy mobile connects to Service Platform for reading the secure data. On completion, the mobile checks the access tokien (thanks to the access token and the authentication data), connects to the Trust Manager for retreiving the security material and decrypts the device data.

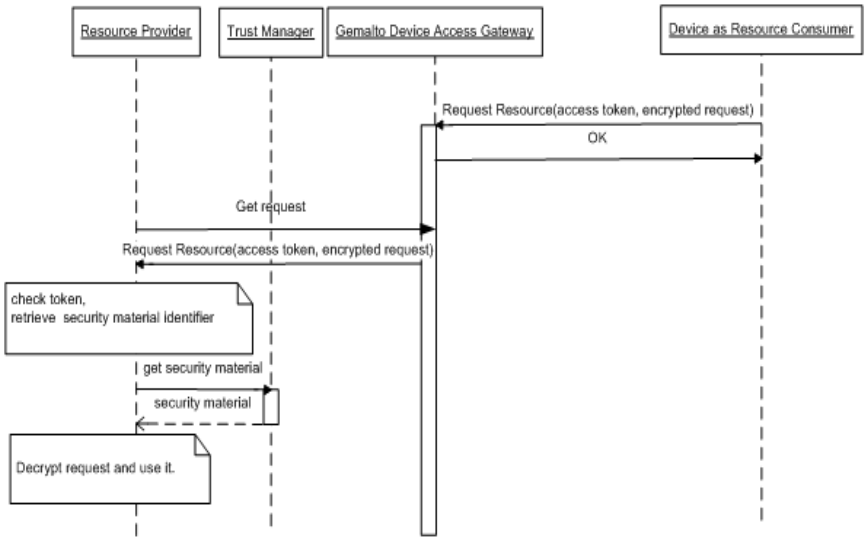


Figure 39: End To End security with Device As Resource Consumer

When the device runs as Resource Consumer, it shall obtain a token to access the specified (destination) Resource Provider. For this purpose the device uses the password schema for retrieving the token. See “Deliverable 5.1 Security Framework API - 2.10.4.5. Password Grant”. The device uses the “SET” scope for obtaining the access token.

The Figure 39 shows the message flow. The schema assumes that the device is already connected to the Trust Manager to obtain the access token and related security material. It has to be noted that the flow is asynchronuous, the device (resource consumer) pushes secure data to the Service Platform and does not receive return value but only a status information; the resource provider (the actuator) asynchronuously uses the data. The point-to-point security between the device and the service platform and the point-to-point security between the resource provider are out-of-scope this description and can rely on BUTLER security framework or other point-to-point security solution as CoAP/DTLS or HTTPS.

Issue - Device running as Resource Consumer

Running as Resource Consumer is usefull for monitoring solution. The device regularly sends data for further application (implementing the role of Resource Provider). This way, the end-to-end security is implemented from the device to the resource. In consequence, the Service Platform has no way to get clear data and therefore cannot market the data but only the data access service and possibly other services related to device management.

3.1.3 Issues Highlighted in the Experimentations

o

The section summarizes the issues that have been found during the implementation of the prototypes and experimentations related to security.

3.1.3.1 Summary of Low-Level security issues

Several security schemes have been tested for IoT devices running on TinyOS (e.g. TinySec, AES Encryption of CC2420, MiniSec, Relic and TinyECC) or Contiki (e.g. ContikiSec, Contiki-TLS-DTLS, Contiki IPsec, CoAPs: COAP over DTLS/TLS). In order to flash sensors with correct applications, a development environment (computer) should be configured in order to support the data transfer towards all motes of the wireless sensor network. A Linux environment has been selected because few communication issues raised during the experimentations, due to missing drivers (for TinyOS). We also used as Instant Contiki that is a virtual machine containing the complete environment to set up the Contiki operating system an a sensor. The main issue met during the experimentation is the limited space of the sensor's ROM. The operating systems such as Contiki and TinyOS can quickly consume the complete ROM/RAM of each mote. We used TelosB platforms that have 48kb of programmable flash and 10kb of data RAM. OS, application(temperature sensing) and communication schemes consumed the large part of the memory. As a consequence, the addition of a lightweight cryptography often became impossible due to the lack of space. The solution is to use other motes with larger memory. Interoperability was also another issue. For instance, the same application has been set up on different sensors (TelosB and Zolertia Z1) but the communication between theses motes was impossible.

The following issues were found in the IKEv2 implementation [11] as described in 3.1.2.1.3. To enable IKEv2 in low-power platforms, programmable flash memory of more than 157 KB is required. Since, only Arago's WISMOTE and ST microelectronics' STM32W108CC platform currently provide enough flash memory (256 KB), this implementation was developed for WISMOTE platform and it needs to be ported to the latter. Therefore, this directly restricts platform interoperability. This implementation was developed and tested with the Cooja simulator provided by Contiki-OS. In addition, software encryption modules were used for performing cryptographic functions which

increase packet processing time, thus limiting the node's ability to process many packets simultaneously. Sometimes, the node crashes when it receives too many packets to be processed at the same time. Therefore, this solution is not stable for production environment. Few lines of the code had to be modified to run IKEv2 application on real WISMOTE platform nodes and as a result, key-exchange was successful, but problems with communications arouse. IPSEC's IKEv2 has Security Policy Database (SPD) and Security Association Database (SAD) which check all inbound and outbound traffic with rules specified within SAD and SPD. We noticed that even after successful IKEv2 negotiation between a computer host (Border Router) and a IPSEC node, the IPSEC node had corrupted its SAD and dropped all outbound packet to the computer host. Therefore, communication cannot occur and to fix this problem, core IPSEC and IKEv2 modules have to be debugged.

CEA has provided a key management scheme compliant with the ZigBee protocol for Wireless Sensor Network (WSN). However, the keys are managed at high level from a secure server and security credentials are stored into a secure database. Each node of the WSN is provided with its dedicated keys and seeds before deployment. This manipulation is tedious when the nodes are numerous and should be performed by an administrator owning the highest access right in the network. The future work will focus on bootstrapping techniques and protocols enabling an easy and cheap deployment of the constrained nodes and that can be performed by anyone.

Implementation of theoretical secrecy framework is not attempted yet. Although we have found some issues when we consider the practical systems like WiFi and cellular networks. First consider the example of typical residential WiFi networks. From a security standpoint, these networks are characterized by the clusterization on a home-by-home basis, i.e, devices within the house are typically considered legitimate users, while devices outside premises play the role of possible eavesdroppers. As a result, the assumption of a random but statistically uniform spatial distribution of base stations and users (as implied by a PPP model) is unrealistic, motivating studies conducted under a more general model . Likewise, in cellular networks, the distribution of base stations follow terrain, as well as regulatory (city-plan), demand and space availability conditions, and therefore are far from (statistically) uniform. Furthermore, devices in urban areas served by pico and femto cells may be clustered together, for instance in and around shopping malls and train-stations), while devices in less populated areas served by macro cells are more sparsely located. Such conditions are clearly distinct from the random and uniformly distributed network assumptions that lead to a Poison number of nodes per unit area – i.e., the PPP model. In response to the limitations, we will consider more generalized models in future works.

1. Implementation of the algorithm on real hardware - it requires hardware capable of quite accurate phase estimation.
2. Performance evaluation after key reconciliation procedures is another issue.
3. Performance evaluation of the system described at point 2 within a real encryption system.

In the general context of IoT and Smart environments, emerging device-centric wireless networks are expected to be truly pervasive. Accordingly they tend to favor peer-to-peer transactions under opportunistic connectivity conditions. They also afford very low embedded computational complexity. Decentralized symmetric secret keys generation and sharing at the physical layer level thus represent appealing alternatives to more conventional key (pre-)distribution and centralized key management strategies. Among the possible radio technologies and standards supporting these novel physical-layer functionalities, the Impulse Radio Ultra Wide-Band (IR-UWB) technology, which is already considered in several low data rate standards (e.g., IEEE 802.15.4a for ranging-enabled Wireless Sensor Networks (WSN), IEEE 802.15.4f for Real Time Location Systems or

IEEE 802.15.6 for Wireless Body Area Networks (WBAN)), enjoys beneficial resolution capabilities for direct bit extraction out of received multipath channels. However the theoretical single-link IR-UWB key generation approaches investigated so far (and recently improved in terms of key randomness, key diversity, legitimate reciprocity, and synchronization/quantization feasibility...) must still be evaluated using real integrated radio devices in time-varying environments, such as that commercialized by BeSpoon http://spoonphone.com/en/. They must also be scaled further to assess practical cooperative scenarios within small (ad hoc) groups of devices/users, which clearly represent the next frontier to enable local “bubbles of trust” in future IoT applications.

3.1.3.2 Summary of Application-Level security issues

At the application level, the main problems concern the bootstrapping of the security and the privacy and the usability of the applications. The next subsection discusses the issues about the Bootstrapping of the security.

About the privacy, the current IoT solutions are vertical solutions where the data are stored in clear at Service Plaform and can be retrieved later by applications. There is a security link from resource to Service Platform and another security link between application and Service Platform, in consequence there is no end-to-end security between the application and the resource, the Service Plaform may perform data analysis but without user consent and therefore this poses a problem of privacy.

Integrating the concept of authorization allows the user to grant access permission to applications. Anyway, for non expert or user who are not concerned by privacy aspects, the required process can be difficult to understand. The user must understand what is a resource, must declare a resource at the Trust Manager, configure such resource, optionally grant permission to another user. For applications, the user shall dynamically allow application to obtain an access token. For that, it must authenticate to the Trust Manager and add its user consent. In consequence, for all users the usability of the Authorization model can be seen as difficult. This usability problem is very important to tackle properly for final user acceptance.

3.1.3.3 Bootstrapping of Security

“IoT is a network of networks consisting of static or mobile devices owning an identity, digital entities and physical objects able to interact and retrieve, store, transfer and process data exchanged via the Internet Protocol (IP)”. “Things” are often located in a Local Area Network (LAN) connected to the global internet via a “gateway” device connected to both the Local Area and the Wide Area Networks (WAN). IP communication protocols are indeed being used after the gateway device, but their adoption on the LAN part, composed of low-power and constrained devices is not always possible. Fortunately, the emergence of solutions and protocols designed for low profile devices such as 6LoWPAN adaptation layer, CoAP and DTLS facilitates the adoption of IP communication protocols up the leaf devices.

Sensing platforms are often used to help “things” communicate with remote applications. These sensing platforms must provide security services properly integrated into the system architecture to secure end-to-end communications between heterogeneous elements.

To achieve such a challenging issue, IP-based communications offer many benefits to develop a secure framework supporting the whole network from the data sources to the final user [73].Internet security protocols provide solutions to secure data transport transparently for the applications [74].

The security solutions used for IoT are usually designed for dedicated scenario requirements without considering generic interoperability with the Internet Security Protocols (ISP) stack. Protocol

translation is done by the gateway device at the boundary between the LAN and the WAN domains. However such translation is the main obstacle to achieve end-to-end security between IoT constrained devices and remote applications [16], [17], [18].

A typical way to proceed in this case is to implement “hop-by-hop” security, involving independently securing with different credentials the WAN and the LAN parts of the communication. This type of solution requires a rekeying operation at the level of the gateway and means that data is always available in clear in this device. Furthermore it leads to a concentration of credentials in the gateway increasing significantly the risk associated to a compromised gateway device.

A classification of the security aspects to consider in the IoT is provided in [79]: The security architecture refers to the system, the interactions between the elements that belong to the system and the security management involving the “things”. A security scheme at the node level enables the node to retrieve its security materials and to process security operations and storage. The bootstrapping phase addresses the secure addition of a new object to the IoT network thanks to trust operations including authorization and authentication. The bootstrapping is not specific to any MAC or PHY layers. It concerns each component that would like to communicate with other components without any previous knowledge of the one with the others. Attacks must be prevented thanks to the threat analysis of the whole network. The resulting system should guarantee that only trusted instances of applications can run over the IoT network and the infrastructure. Moreover, the life of a thing begins at manufacture. The device identity and the secret keys used during the running cycle are provided during the bootstrapping phase. Once deployed, the device is under the control of its owner.

An architecture built around a security management platform is commonly adopted. The proposed security framework handles all the security aspects detailed in [79]. It could be implemented in three steps:

- The Bootstrapping at low-level (LAN): consists in securing the “small” data between the object and the gateway / proxy / modem linked to the WAN. It may consist in crediting the components of the LAN with a shared session key that we can call “local” session key.
- The Bootstrapping at “high-level” (WAN): consists in distributing the access rights under the form of access-token and cryptographic keys to the components of the WAN - connected to internet - authorized to exchange information or to access to resources.
- The Session establishment: addresses the problem of distribution ephemeral session credentials from the object to the user in order to implement “hop-by-hop” or “end-to-end” security.

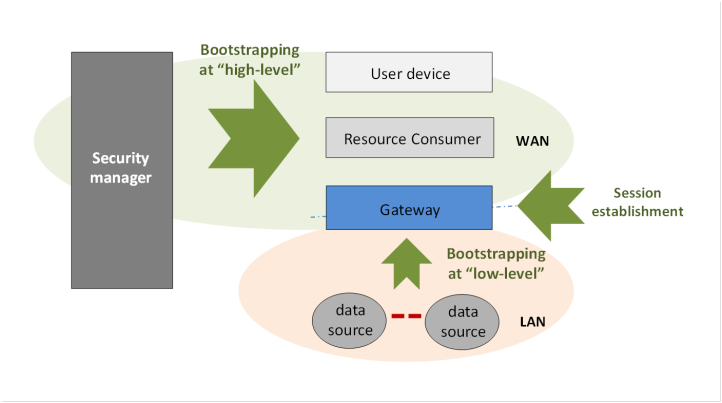


Figure 40: The three phases of security deployment.

The bootstrapping at “high level” in the WAN can be performed thanks to protocol as OAuth2.0. OAuth 2.0 [80] was recently introduced by Google to provide third-party applications access to resources via an application. It introduces an authorization server distinct from the data application

path. The authorization server provides an access-token to the application to access to the remote resources whatever the credentials of the end-user. However, this scheme runs on the WAN domain. The remaining challenge is to authorize the access to a resource located on a constrained device communicating wirelessly with a low-power standard in the LAN domain.

Nowadays, there is no standard to perform the bootstrapping of the objects located in the LAN to the network. A variety of techniques co-exist. The bootstrapping technique most familiar to the general public is to connect via WiFi a computer, tablet or smartphone to the modem / box at home. To bootstrap, simply note the key value that can be read on the modem and enter this value using the keyboard when the device tries to connect to private and secure WiFi network. In this case, all devices connected to the same modem share the same key.

The problem is more complex with the connected objects:

- The objects are headless and/or physically difficult to access,
- The confidentiality should be ensured between different nodes which involve managing the group concept,
- The objects are inexpensive, purchased bare, resource constrained and communicate via a low-power radio standard as IEEE 802.15.4.

Several techniques have been envisaged to secure a pair of nodes, starting from nothing - an insecure channel - and leading to a secure channel:

1. Channel estimation based: Based on the assumption of the channel reciprocity, the channel estimation by both devices of a peer-to-peer communication gives temporal signals that are “mirrors” the one of the other. These two signals could be considered as two correlated random sources and a common secret can be forged secretly between the two devices.

   This technique implies that the two devices are close enough (i.e. within the communication range), that the communication channel is stationary during the time of one symbol and that little interference perturbs the channel meanwhile. For instance, relying advantageously on impulse radio - ultra wideband signals over a given communication link, each peer should be able to send sequences of impulse waveforms (on the transmitter side) and to perform channel estimation (on the receiver side), so as to compute the channel impulse response and benefit from multipath components accordingly. As such capabilities are still challenging to embed in low power integrated devices, this technology and the related standards, e.g. IEEE 802.15.4a/4r have been emerging for the last past years and the transfer to industry has just started, mostly for localization applications [81] [82].
2. In-band pairing: No additional interface or hardware is needed, but the applicability with IEEE 802.15.4 standard is questioned. As the devices communicating with low-power IEEE 802.15.4 on UDP are usually resource constrained, the management of certificates and the implementation of a PKI is too costly. Lightweight handshake could be envisaged at the application layer thanks to DTLS but the session key establishment remains an issue. At the network layer, compression schemes for IPsec have been proposed in order to fit into 127-bits length frames. IPsec AH or IPsec ESP may be available in transport mode only for low-power communications. No security association is proposed leaving aside the issue of bootstrapping.
3. Secure storage for private key needed: PUF, TPM-like. To physically secure secret credentials as cryptographic private key, a secure element - TPM-like - could be used. This hardware is now provided inside numerous devices as computer or tablet, but remains costly for resource constrained devices.

   The PUF - as Physical Unclonable Function - enables to forge a singular secret inside a device based on its singular characteristics provided at manufacture. From this secret, cryptographic keys or identity can be derived before initiating a handshake with another device.

4. Based on the assumption that the public key or the certificate of the service provider is already deployed inside the node memory, and that the node is able to generate its own cryptographic features - thanks to an embedded “true” random number or a PUF for example - lightweight handshake protocols can be launched to establish a secure channel. This implies that lightweight asymmetric cryptography technique is embedded into the node. This solution is a compromise that allows the node mobility and needs an appropriate management of lightweight certificates. The actual standards - as X.509 - leads to certificates too big to be embedded into resource constrained nodes.
5. Out-of-band: A different channel could be used to share a secret key between two devices while ensuring that the information is broadcast on the air without confidentiality. Carelessly, everybody located in the coverage area of the transmission is able to eavesdrop the secret. An USB link, a button-to-button or a LED transmission can be used to share the secret and/or the cryptographic material before the deployment of the node. This technique could require an additional PHY/MAC layer embedded into the node only used for the bootstrapping operation. The maintenance of the keys may be delicate if the secret is lost and this situation must be avoided. This solution may be easy to use and “plug and play”.

The choice of a bootstrapping technique must involve the architecture of the network, the technology of the nodes, the supported uses cases or applications running over the network, the criticality of the data exchanged. A threat analysis should be launched that considers the vulnerabilities of the system and to determine the risks that we want to cover.

All of these techniques aim to secure a connection between two components. In the Internet of Things, the ultimate goal is to protect the resource emitted by the object and hide the identity of its owner. This requires rethinking these techniques by introducing the concept of groups and managing cryptographic features with flexibility for groups of individuals and / or objects. A thorough analysis of the different security protocols currently existing for IEEE 802.15.4 frames is detailed in [83]. This analysis provides a valuable point of departure in order to consider the security at low level with the goal to provide secure bootstrapping from link to application layer thanks to Plug & Play techniques.

3.2 Localization

Wireless localization is a fairly mature area of research, with a vast literature [84] and various contributions in Deliverable 2.2 [85]. It is somehow paradoxical that despite the formidable effort put into the problem, wireless positioning is still shy of its potential as a truly ubiquitous and real-time locating technology [86–88]. Ubiquity requires the technology to be available in every environment, while real-time location implies automatic identification and tracking of location within an environment.

It is well-known that wireless localization systems are still inaccurate and unreliable in places such as urban cities and indoors, which are characterized by high multipath propagation and scarcity of Line of Sight (LoS) conditions. Therefore we seek to present algorithms to mitiage against mulitpath and Non Line of Sight (NLoS) conditions for accurate indoor wireless localization.

In this section, we provide an accurate ranging algorithm using superresolution techniques over phase measurements for distance estimation between devices. This algorithm has already been implemented in ZIGPOS Localization Architecture. We also presented the possiblity of performing multipoint ranging via these algorithm using orthogonal set of Golomb rulers. With distance estimation discussed, we provided accurate, robust and efficient positioning algorithms for target localization using algebraic confidence via circular interval scaling and an hybrid cooperative algorithm. To improve the performance of the positioning algorithms, therein target coordinates, a cooperative technique for detecting NLoS measurements was also presented.

3.2.1 Ranging Algorithms

To qualify the above statements, consider the case of Angle of Arrival (AoA) positioning were a good number of AoA localization, and estimation algorithms [89, 90] exists. Of particular relevance is the fact that simultaneous estimation of the AoA of multiple signals/sources is relatively easy to perform, which is of fundamental importance to reduce latency in indoor applications where the concentration of users is typically large.

Yet, AoA-based indoor positioning is not common today because: a) AoA-based localization algorithms are highly susceptible to NLoS conditions, such that accurate and robust AoA input is needed; and b) accurate and robust AoA estimation requires expensive multi-antenna systems and high computational capabilities, which are incompatible with typical indoor requirements of small, low-cost, low-power devices.

The limitations of the AoA-based approaches partially explain the predominance of range-based indoor localization systems proposed by academia. Indeed, various accurate robust distance-based localization algorithms exist, and distance estimates are relatively inexpensive to obtain from radio signals without requiring multiple antennas or significant additional RF circuitry. But again the deployment of this technology is short of its potential, which arguably is a result of the fact that since ranging quality is severely degraded by interference, positioning systems are required to carefully schedule the collection of ranging information, leading to low refreshing ratios and communication costs.

On one hand, many excellent AoA estimation algorithms exist, which however are not typically utilised for indoor positioning as multi-antenna systems are too expensive. On the other, many excellent distance-based localization algorithms exist [84], which however can only be effectively employed for indoor positioning, if ranging information can be collected efficiently from multiple sources so as to reduce latency.

The work is as follows, in Paragraphs 3.2.1.1 and 3.2.1.2, we briefly revise the Phase-Difference of Arrival (PDoA)-based ranging techniques and their integration with superresolution algorithms for multipoint ranging. In Paragraph 3.2.1.3, the Golomb-optimized scheme which yields the possibility of extending the technique to perform simultaneous multipoint ranging to reduce latency is discussed. The performance of the algorithm is also discussed in Paragraph 3.2.2, which include comparisons against the corresponding Cramér-Rao lower bounds (CRLBs) for ranging and Localization. Afterwards, conclusions are drawn.

3.2.1.1 PDoA Continuous Wave Radar Ranging

We shall focus on PDoA to estimate the distance between a pair of wireless devices using their signals. Consider the problem of estimating the distance d between an anchor A and a target T based on the phases of the signals exchanged between the devices. As illustrated in Figure 41, an anchor A emits a continuous sinusoidal wave of frequency f with a known phase φ_{TX} and the target T acts as an active reflector, such that A can measure the phase φ_{RX} of the returned signal [91]. In this case, the roundtrip distance $2d$ and the phases φ_{TX} and φ_{RX} are related by

$$\varphi = \varphi_{\mathsf{RX}} - \varphi_{\mathsf{TX}} = \frac{4\pi d}{c} f - 2\pi N, \tag{40}$$

where N is the integer number of complete cycles of the sinusoidal over the distance $2d$.

Obviously the distance d cannot be estimated directly based on equation (40) since the quantity N is unknown. However, taking the derivative of (40) with respect to f one obtains

$$\frac{\mathrm{d}\varphi}{\mathrm{d}f} = \frac{4\pi d}{c}. \tag{41}$$

Let there be a set of equi-spaced frequencies $\mathbb{F} = \{f_0, \cdots, f_K\}$ such that $\Delta f = f_{k+1} - f_k$ for all $0 \leq k < K$, and assume the roundtrip phases φ_k for all f_k are measured.

From the linear relationship between f and d in (41), we have

$$\Delta\varphi_k = \frac{4\pi\Delta f d}{c} k = \omega_d k, \tag{42}$$

where $\Delta\varphi_k \triangleq \varphi_k - \varphi_0$ for all $1 \leq k < K$.

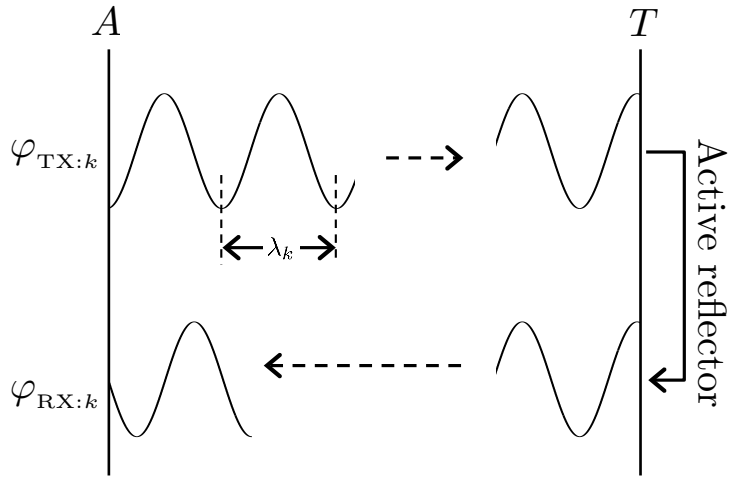


Figure 41: Illustration of PDoA ranging mechanism, Multipoint-point ranging can be performed by allocating different sources to different orthogonal carriers.

Notice that due to the linearity of this relationship, we have, for any pair of integers (k, q), with $k > q$,

$$\Delta\varphi_k - \Delta\varphi_q = \omega_d \cdot (k - q) = \Delta\varphi_{k-q}. \tag{43}$$

This simple property has a remarkable consequence. Indeed, consider an ascending sequence of non-negative integers $\mathcal{N} = \{n_1, \cdots, n_K\}$ and the associated set of input measurements $\mathbb{\Delta}\varphi_{\mathcal{N}} = \{\Delta\varphi_{n_1}, \cdots, \Delta\varphi_{n_K}\}$. By virtue of (43), the set $\mathbb{\Delta}\varphi_{\mathcal{N}}$ can be expanded into $\mathbb{\Delta}\varphi_{\mathcal{V}} = \{\Delta\varphi_{\nu_1}, \cdots, \Delta\varphi_{\nu_M}\}$, where the cardinality M of $\mathbb{\Delta}\varphi_{\mathcal{V}}$ is upper bounded by $M \leq K\frac{K-1}{2}$.

Other than the much larger cardinality, the sequences $\mathbb{\Delta}\varphi_{\mathcal{V}}$ and $\mathbb{\Delta}\varphi_{\mathcal{N}}$ have, as far as the purpose of distance estimation is concerned, fundamentally the same nature since both carry samples of the quantities $\Delta\varphi_k$. In other words, the model described above allows for large input sets of cardinality M to be obtained from a significantly smaller number K of actual measurements, by carefully designing the carrier frequencies required to perform ranging estimates. Furthermore, the linearity between the measured quantities $\Delta\varphi_k$ and the corresponding indexes k is so that such design can be considered directly in terms of the relationship between the integer sequences $\mathcal{N} \to \mathcal{V}$. Sparse sequences $\mathcal{N}$ that generate optimally expanded $\mathcal{V}$ are known as *Golomb rulers*.

3.2.1.2 Multipoint Ranging via Superresolution Algorithms

Straightforwardly, assume that a set of input measurements $\mathbb{\Delta}\varphi_{\mathcal{N}}$ is collected, from which the associated expanded set $\mathbb{\Delta}\varphi_{\mathcal{V}}$ is constructed and consider the corresponding complex vector

$$\mathbf{x} = [e^{j\Delta\varphi_{\nu_1}}, \cdots, e^{j\Delta\varphi_{\nu_M}}]^{\mathsf{T}} \equiv [e^{j\omega_d}, \cdots, e^{j\nu_M\omega_d}]^{\mathsf{T}}, \tag{44}$$

where $^{\mathsf{T}}$ denotes transposition and ν_1 is normalised to be 1.

One can immediately recognize from (44) the similarity between the vector $\mathbf{x}$ and the steering vector of a linear antenna array [90]. An estimate of the parameter of interest ω_d can therefore be recovered from the covariance matrix $\mathbf{R}_{\mathbf{x}} \triangleq \mathbb{E}[\mathbf{x} \cdot \mathbf{x}^{\mathsf{H}}]$. Specifically, under the assumption that each measurement $\Delta\varphi_{\nu_m}$ is subject to independent and identically distributed white noise with variance σ^2, the covariance matrix $\mathbf{R}_{\mathbf{x}}$ can be eigen-decomposed to

$$\mathbf{R}_{\mathbf{x}} = \mathbf{U} \cdot \mathbf{\Lambda} \cdot \mathbf{U}^{\mathsf{H}}, \tag{45}$$

with

$$\mathbf{U} = \left[\mathbf{u_x} \,\middle|\, \mathbf{U}_0,\right] \quad \text{and} \quad \mathbf{\Lambda} = \left[\begin{array}{c|c} 1+\sigma^2 & \mathbf{0} \\ \hline \mathbf{0} & \sigma^2\mathbf{I} \end{array}\right], \tag{46}$$

where $\mathbf{U}_0$ is the K-by-$(K-1)$ null-space of $\mathbf{R_x}$.

From the above properties, many superresolution algorithms such as Music, Root-Music, etc. [90, 92], can be employed to obtain ranging estimates as in [93]. Our focus mainly is to discuss the resulting opportunities to optimize resources, and analyze the corresponding implications on the achievable ranging accuracies. Whatever the specific method used to extract the distance $\hat{d}$ (embedded in $\hat{\omega}_d$) from the vectors constructed in (44), the following properties apply to superresolution algorithms.

- Superposibility: Thanks to the expansions $\mathcal{N} \to \mathcal{V}$, measurement frequencies corresponding to multiple sources can be superposed without harm. To exemplify, consider the case of two sources A and B and the measurements from both sources be collected continuously according to the sequence $\mathcal{N} = \{1, 3, 4, 5, 6, 7, 8, 10\}$, but such that the sources A and B are only active according to the orthogonal sequences $\mathcal{N}_A = \{1, 3, 6, 7\}$ and $\mathcal{N}_B = \{4, 5, 8, 10\}$. The samples in $\mathcal{N}_A$ can, however, be transformed into the sequence $\mathcal{V}_A = \{1, 2, 3, 4, 5, 6\}$, which contains 6 samples. Furthermore and likewise, $\mathcal{N}_B \to \mathcal{V}_B = \{1, 2, 3, 4, 5, 6\}$. In other words, out of only 8 jointly collected samples, 6 equivalent measurements from each source are obtained without interference.
- Unambiguity: In AoA estimation using antenna arrays, the elements of the steering vectors are complex numbers whose arguments are *periodic functions* of the desired parameter, which in turn gives rise to aliasing of multiple parameter values that lead to the same set of measurements [94]. In contrast, in the context hereby the quantities $\Delta\varphi_k$ are *linear functions* of the desired parameter d, such that no such ambiguity occurs.
- Separability: Thanks to both properties above, superresolution ranging can be carried without interference using orthogonal non-uniform sample vectors, each processed by a separate estimator. Consequently, issues such as correlation amongst multiple signals in superresolution algorithms [90], do not exist in the context hereby.

3.2.1.3 Optimization of PDoA Range Sampling via Golomb Rulers

Under the model described in Section 3.2.1.1, the optimization of resources amounts to allocating ranging frequencies to multiple sources, directly related to that of Golomb rulers [95].

Golomb rulers are sets of integer numbers that generate, by means of the difference amongst their elements, larger sets of integers, without repetition. Briefly, we will review some of the basic characteristics and features of Golomb rulers.

Basic Features of Golomb Rulers

Consider a set of ordered, non-negative integer numbers $\mathcal{N} = \{n_1, n_2, \cdots, n_K\}$, with $n_1 = 0$ and $n_K = N$. This set has *order* K, and *length* – its largest element N. The corresponding set $\mathcal{V}$ of all possible pairwise differences

$$\nu_{k\ell} = n_k - n_\ell \quad (1 \leq \ell < k \leq K). \tag{47}$$

If the differences $\nu_{k\ell}$ are such that $\nu_{k\ell} = \nu_{pq}$ if and only if $k = p$ and $\ell = q$, then the set $\mathcal{N}$ is known as a *Golomb ruler* and their elements are understood as the *marks* of a ruler, which can *measure* only the lengths indicated by any pair of marks. We henceforth refer to the set $\mathcal{V}$ as the *measures* set. It follows from the definition that the number of distinct lengths that can be measured by a Golomb ruler – the order of $\mathcal{V}$ – is $M = K\frac{K-1}{2}$. The first key feature of a Golomb ruler is that if $\mathcal{N}$ has order K, then $\mathcal{V}$ has order M.

An example of a Golomb ruler is $\mathcal{N} = \{0, 1, 4, 6\}$, which generates the Measures $\mathcal{V} = \{1, 2, 3, 4, 5, 6\}$. In this particular example, $\mathcal{V}$ is *complete*, as it contains all integers up to its length, so that the

Golomb of order 4 is said to be a *perfect* ruler as it allows for *all distances* to its length to be measured.

Unfortunately, no perfect Golomb ruler exists [96] for $K > 4$. It is therefore typical to focus on designing rulers that retain another feature of the order-4 Golomb ruler, namely, its compactness or *optimality* in the following senses: a) no ruler shorter then $N = 6$ can exist that yields $M = 6$ distinct measures; and b) no further marks can be added to the ruler, without adding redundancy. In general, these two distinct optimality criteria are defined as

a) *Length optimality*: Given a certain order K, the ruler's length N is *minimal*;

b) *Density optimality*: Given a certain length N, the ruler's order K is *maximal*.

The new genetic algorithm we presented in [97] is used to generate *orthogonal* Golomb rulers (so as to enable multipoint ranging), that also come as close as possible to satisfying the length and density of the optimality criteria described above (so as to optimise resources).

Furthermore, there are two distinguished ways the resulting Golomb rulers can be grouped together. One possibility is to group the rulers such that all have the same length N, even if with different different number of marks. This approach is motivated by the fact that the corresponding array-like vectors (see (44)) will have the same aperture, which in turn is directly related to the accuracy of the corresponding distance estimation via superresolution algorithms. This choice is referred to as Equivalent Ranging Quality (ERQ). Another possibility, however, is to group the Golomb rulers with the same cardinality K. This grouping approach is motivated by the fact that, in the context hereby, each marker in the ruler corresponds to a measurement that is taken, and therefore is referred to as Fair Resource Allocation (FRA).

Examples of Golomb rulers obtained with the algorithm described in [97] and grouped according to the ERQ and FRA criteria are listed in Table 7. It can be observed that, as desired, no two identical numbers (markers) can be found in two different rulers within the same group. It follows that all the rulers of each group can be superimposed without interference and within a maximally compact span (conventional design would be to shift each ruler by length of the later).

To clarify, thanks to the rulers displayed in Table 7, within a block of no more than 100 frequencies, multipoint ranging between a target and 5 different anchors can be carried out by taking only 50 PDoA measurements. This can be achieved either by using the ERQ or FRA group of rulers, respectively.

Table 7: Examples of Golomb Rulers with ERQ and FRA Designs.

K	Equal Ranging Quality	N	M	K	Fair Resource Allocation	N	M
9	0,1,7,10,30,41,45,63,87	87	36	10	0,1,16,21,24,49,63,75,81,85	85	45
9	2,3,6,32,37,49,56,76,89	87	36	10	2,3,11,32,45,56,60,72,78,92	90	45
10	4,5,16,20,33,42,52,66,73,91	87	45	10	5,9,15,29,42,51,68,80,91,96	91	45
11	8,9,18,21,38,46,53,72,77,93,95	87	55	10	6,13,17,19,33,43,61,62,84,93	87	45
11	12,13,17,25,31,47,68,70,79,96,99	87	55	10	12,14,22,27,28,46,66,73,77,94	82	45

Conclusion

We offered an efficient and accurate solution to the multipoint ranging problem, by adapting superresolution techniques with optimized sampling. Specifically, using Phase-Difference of Arrival (PDoA), we constructed a variation applicable with superresolution algorithms to perform distance estimation over sparse sample sets determined by Golomb rulers. The design of the mutually orthogonal sets of Golomb rulers required by the proposed method was shown to be achievable via the algorithm in [97].

3.2.2 Error Analysis and Comparisons in Wireless Localization

In this paragraph, we analyse the performance of the multipoint ranging described above for target localization. To this end, we first derive the Fisher Information Matrices and associated CRLBs corresponding to ranging and target localization and later offer comparisons with simulated results.

Start by recognising that phase difference measurements subject to errors are circular random variables. The Central Limit Theorem (CLT) over circular domains establishes that the most entropic model for circular variables with known mean and variance is the von Mises or Tikhonov distribution [98]. We therefore model the phase measurements as

$$\hat{\Delta}\varphi \sim P_{\mathcal{T}}(\theta; \Delta\varphi, \kappa) \tag{48}$$

with

$$P_{\mathcal{T}}(\theta; \Delta\varphi, \kappa) \triangleq \frac{\exp(\kappa\cos(\theta - \Delta\varphi))}{2\pi I_0(\kappa)}, \;\; -\pi \leq \theta \leq \pi, \tag{49}$$

where $I_n(\kappa)$ is the n-th order modified Bessel function of the first kind and κ is the shape parameter given by the signal-to-noise-ratio (SNR) of input signals, and relates to the error variance by

$$\sigma^2_{\Delta\varphi} = 1 - \frac{I_1(\kappa)}{I_0(\kappa)} \xrightarrow[\kappa>>1]{} \frac{2}{2\kappa+1} \approx \frac{2}{\kappa}. \tag{50}$$

Consider then that a set of K independent measurements $\{\Delta\varphi_k\}_{k\in\mathcal{N}}$ is collected according to a Golomb ruler $\mathcal{N}$, such that the samples can be expanded into and augmented set of M samples $\{\Delta\varphi_m\}_{m\in\mathcal{V}}$, with

$$\Delta\varphi_m = \Delta\varphi_k - \Delta\varphi_\ell = \omega_d(k-\ell) = \omega_d\nu_m, \tag{51}$$

where each index m corresponds to a pair (k, ℓ) with $k > \ell$ with ascending differences, and note that $\nu_m \neq m$.

Although the expanded samples $\Delta\varphi_m$ are differences of phase differences, these quantities not only preserve the linear relationship with the parameter of interest but also their independence. As a result of the double-differences, however, the SNR of $\Delta\varphi_m$ from equation (51) is twice that of $\Delta\varphi_k$ from equation (42). In light of the asymptotic relationship being twice as large, it follows that the shape parameter κ associated with $\Delta\varphi_m$'s are twice as small.

3.2.2.1 Ranging

Using the model above, and incorporating the optimized sampling via Golomb ruler, the likelihood function associated with M independent measurements as per (42) is,

$$\begin{aligned} L_{\mathcal{T}}(\hat{d}; \Delta f, \kappa) &= \prod_{m=1}^{M} P_{\mathcal{T}}(\theta; \Delta\varphi_m, \kappa) \\ &= \prod_{m=1}^{M} \frac{\exp\left[\frac{\kappa}{2}\cos\left(\alpha\cdot\nu_m(\hat{d}-d)\right)\right]}{2\pi I_0(\kappa/2))}, \end{aligned} \tag{52}$$

where α is defined as $\frac{4\pi\Delta f}{c}$, $\nu_m \in \mathcal{V}$ and have modified the notation above to emphasise the parameter of interest in $\hat{d}$.

The Fisher Information as the negated expectation of the Hessian is

$$\begin{aligned} J(\mathcal{V}; \Delta f, \kappa) &= -\mathbb{E}\left[\frac{\partial^2 \ln L_{\mathcal{T}}(\hat{d}; \Delta f, \kappa)}{\partial d^2}\right] \\ &= \frac{\alpha^2\kappa}{2}\frac{I_1(\kappa/2)}{I_0(\kappa/2)}\sum_{m=1}^{M}\nu_m^2, \end{aligned} \tag{53}$$

where the notation alludes to the fact that the key input determining the Fisher Information is the set of measures $\mathcal{V} = \{\nu_1, \cdots, \nu_M\}$.

The CRLB is obtained directly by taking its inverse, $i.e.$,

$$\text{CRLB}(\mathcal{V}; \Delta f, \kappa) = \frac{1}{J(\mathcal{V}; \Delta f, \kappa)}. \tag{54}$$

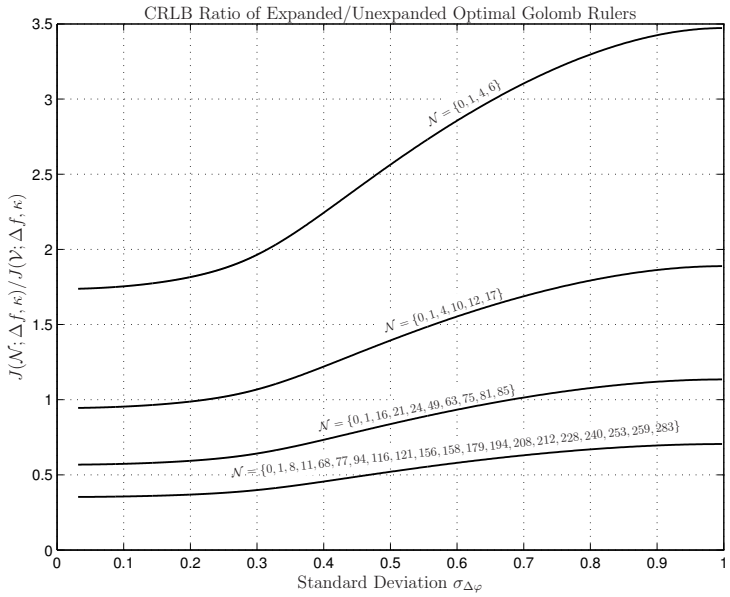


Figure 42: Evolution of CRLB ratio $J(\mathcal{N}; \Delta f, \kappa)/J(\mathcal{V}; \Delta f, \kappa)$ as a function of the phase error standard deviation $\sigma_{\Delta\varphi}$, associated with different rulers $\mathcal{N}$. A value of sigma above 0.5 signifies highly noisy phase measuremens which are already outside the region of interest.

Before proceeding, discussions on the analytical results off-ered above are in order. First, let us emphasise that given a set of phase difference measurements $\{\Delta\varphi_{n_k}\}_{k=1}^{K}$, with $n_k \in \mathcal{N}$, one always has the *option* of either exploit the properties of the Golomb ruler $\mathcal{N}$ and expand to a set of measurements $\{\Delta\varphi_{\nu_m}\}_{m=1}^{M}$, or not. In case such option is *not* adopted, the associated Fisher Information and CRLB are obtained as done above, but with κ replacing $\kappa/2$ and $\mathcal{N}$ replacing $\mathcal{V}$. That is,

$$J(\mathcal{N}; \Delta f, \kappa) = \alpha^2 \kappa \frac{I_1(\kappa)}{I_0(\kappa)} \sum_{k=1}^{K} n_k^2 \tag{55}$$

$$\text{CRLB}(\mathcal{N}; \Delta f, \kappa) = \frac{1}{J(\mathcal{N}; \Delta f, \kappa)}.$$

Comparing these expressions, it can be readily seen that the choice of adopting the Golomb approach on the one hand subjects the resulting double-phase-differences to twice the noise, but on the other hand expands the number terms in the summation. In principle, the optimum choice between these options depends on the ruler $\mathcal{N}$ and its order K, and the associated $\mathcal{V}$ and M, as well as κ. As shown in Figure 42, for instance, the ruler $\mathcal{N} = \{0, 1, 4, 6\}$ yields superior results compared to its associated measure set $\mathcal{V} = \{1, 2, 3, 4, 5, 6\}$, because the loss of 3dB (implied by $\kappa \to \kappa/2$) incurred by the latter is not compensated by the gain in the sum of squares achieved by using $\mathcal{V}$ instead of $\mathcal{N}$.

For larger rulers, however, the advantage of expanding the rulers quickly becomes significant, thanks to the geometric increase of M with respect to K. A ruler of order-6, $e.g.$, $\mathcal{N} = \{0, 1, 4, 10, 12, 17\}$, already achieves better performance expanded into $\mathcal{V} = \{1, \cdots, 17\}$ then otherwise, for $\sigma_{\Delta_\varphi} \leq 0.22$. Likewise, the expanded version of the order-10 ruler $\mathcal{N} = \{0, 1, 16, 21, 24, 49, 63, 75, 81, 85\}$ is superior up to $\sigma_{\Delta_\varphi} \leq 0.67$ – which defines essentially the entire range of interest – and finally the expanded ruler of order-20 is always superior, for any σ_{Δ_φ}. In summary, it can be said that applying the Golomb expansion leads to superior results, as long as the ruler is large enough and $\sigma_{\Delta\varphi}$ is in the region of interest.

3.2.2.2 Localization

Assume a *network* of $N = \{1, \cdots, N_a + 1\}$ devices in an 2-dimensional Euclidean Space, of which there are N_a anchors (x_i, y_i) and a target (x, y). Incorporating the multipoint ranging via orthogonal Golomb rulers for target location, the likelihood function associated with N_a anchors M independent measurements as per (42) becomes,

$$\begin{aligned} L_{\mathcal{T}}(\hat{d}; \Delta f, \kappa) &= \prod_{i=1}^{N_a} \prod_{m=1}^{M} P_{\mathcal{T}}(\theta; \Delta\varphi_{m:i}, \kappa) \\ &= \prod_{i=1}^{N_a} \prod_{m=1}^{M} \frac{\exp\left[\frac{\kappa}{2} \cos\left(\alpha \cdot \nu_{m:i}(\hat{d}_i - d_i)\right)\right]}{2\pi I_0(\kappa/2))}, \end{aligned} \tag{56}$$

where $\nu_{m:i} \in \mathcal{V}_i$ and i in the above equation used to denote each target-to-anchor link.

The Fisher Information Matrix in 2-dimensions is

$$\mathbf{J}(\boldsymbol{\mathcal{V}}; \Delta f, \kappa) \triangleq \begin{bmatrix} J_{xx} & J_{xy} \\ J_{xy} & J_{yy} \end{bmatrix}, \tag{57}$$

where

$$\begin{aligned} J_{xx}(\boldsymbol{\mathcal{V}}; \Delta f, \kappa) &= -\mathbb{E}\left[\frac{\partial^2 \ln L_{\mathcal{T}}(\hat{d}; \Delta f, \kappa)}{\partial x^2}\right] \\ &= \frac{\alpha^2 \kappa}{2} \frac{I_1(\kappa/2)}{I_0(\kappa/2)} \sum_{i=1}^{N_a} \left(\frac{(x - x_i)^2}{d_i^2} \sum_{m=1}^{M} \nu_{m:i}^2 \right), \end{aligned} \tag{58}$$

$$\begin{aligned} J_{yy}(\boldsymbol{\mathcal{V}}; \Delta f, \kappa) &= -\mathbb{E}\left[\frac{\partial^2 \ln L_{\mathcal{T}}(\hat{d}; \Delta f, \kappa)}{\partial y^2}\right] \\ &= \frac{\alpha^2 \kappa}{2} \frac{I_1(\kappa/2)}{I_0(\kappa/2)} \sum_{i=1}^{N_a} \left(\frac{(y - y_i)^2}{d_i^2} \sum_{m=1}^{M} \nu_{m:i}^2 \right), \end{aligned} \tag{59}$$

$$\begin{aligned} J_{xy}(\boldsymbol{\mathcal{V}}; \Delta f, \kappa) &= -\mathbb{E}\left[\frac{\partial^2 \ln L_{\mathcal{T}}(\hat{d}; \Delta f, \kappa)}{\partial x \partial y}\right] \\ &= \frac{\alpha^2 \kappa}{2} \frac{I_1(\kappa/2)}{I_0(\kappa/2)} \sum_{i=1}^{N_a} \left(\frac{(x - x_i)(y - y_i)}{d_i^2} \sum_{m=1}^{M} \nu_{m:i}^2 \right), \end{aligned} \tag{60}$$

where the key input determining the Fisher Information are the N_a set of measures $\boldsymbol{\mathcal{V}} = \{\mathcal{V}_i\}_{i=1}^{N_a}$.

The CRLB is obtained directly by taking the trace of its inverse, $i.e.$,

$$\text{CRLB}(\boldsymbol{\mathcal{V}}; \Delta f, \kappa) = \text{Tr}\{\frac{1}{\mathbf{J}(\boldsymbol{\mathcal{V}}; \Delta f, \kappa)}\}. \tag{61}$$

3.2.2.3 Simulations and Comparison Results

Let us finally study the performance of the proposed multipoint ranging technique by means of simulations and comparisons with the corresponding CRLBs derived above.

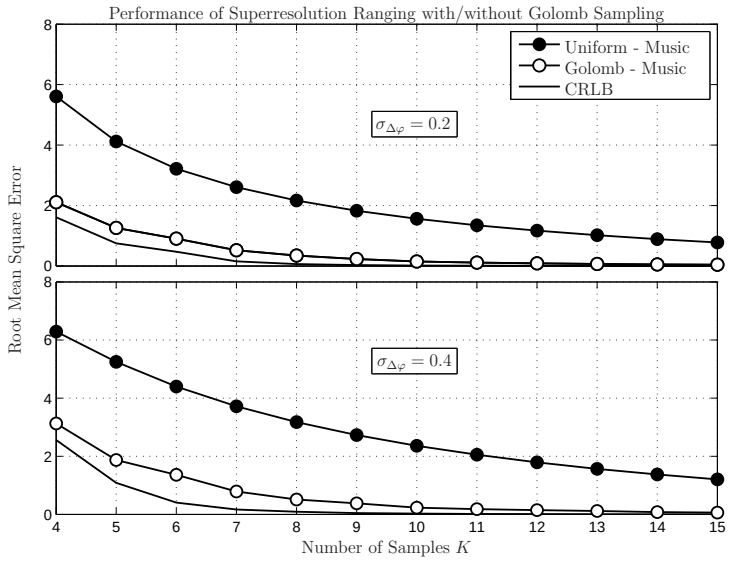


Figure 43: Performance of superresolution ranging algorithms as a function of the sample set sizes K for different $\sigma_{\Delta\varphi}$.

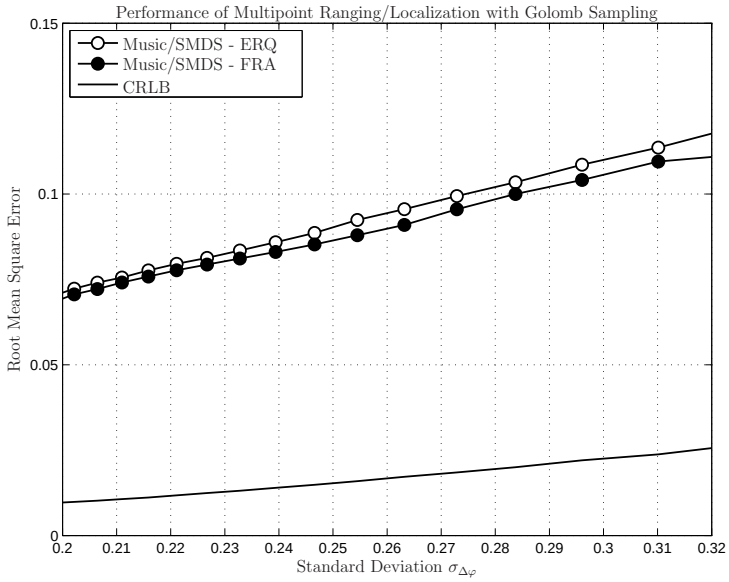


Figure 44: Performance of wireless localization with Golomb-optimized multipoint ranging with ERQ and FRA ruler allocation approaches.

One fact learned from the plots in [93], where *no Golomb ruler is used* and rather a sequence of K consecutive samples are collected for each range estimate, as in existing work [99] is that without the efficient use of samples made possible by the Golomb ruler, superresolution algorithms require a large number of samples in order to reach the CRLB, which is a problem in energy consumption and latency related directly to the number of samples collected. Although supperresolution methods do improve on a "naive" average-based estimator, that gain in itself is not that significant unless the sample size K is rather large.

The results above emphasize the significance of our contribution, by demonstrating that the efficient utilisation of samples is fundamental to reap from superresolution algorithms their true potential performance. This is further illustrated in Figure 43, where it can be seen that thanks to the Golomb sampling, superresolution algorithms with a relatively small number of samples are much closer to the CRLB.

Considered in coordination with the results of Figure 42, it can be generally said that a Golomb-optimized scheme with a total of 10 samples, taken at frequencies corresponding to a Golomb ruler $\mathcal{N}$ expanded into the associated measure set $\mathcal{V}$, followed by MUSIC estimation is an excellent choice for PDoA ranging.

In fact, as illustrated by Table 7, such a choice also allows for an easy design of various orthogonal Golomb rulers, such that multipoint ranging can be efficiently performed for target localization. But since in this case a choice needs to be made between the ERQ and FRA ruler allocation approaches, a fair question to ask in this context is what are the performances of corresponding choices in target localization.

This is addressed in Figure 44, in which we considered a network with anchors located at coordinates $[x_i, y_i] = [1, 4], [4, 8], [7, 5], [2, 2]$ and $[6, 2]$, respectively, and a target at position $[x, y] = [3, 5]$ in which the distances for each target-to-anchor link $(d_i : i = 1, \cdots, 5)$ are converted to phase measurements and Tikhononv random variables are added as phase errors. The estimated distances $\hat{d}_i$ are obtained using 5 orthogonal Golomb rulers which are rotated amongst the anchors. Thereafter, the target's location is determined by applying the Super MultiDimensional Scaling (SMDS) algorithm [100] on the estimated distances and the root mean square error (RMSE) on the estimated target's location is obtained.

The average performances of the ERQ and FRA multipoint ranging schemes employing the rulers shown in Table 7 while using the SMDS algorithm for target localization are compared against corresponding CRLBs. Figure 44 shows that in fact both approaches have similar performances relative to one another and are close to the CRLBs.

Conclusion

A Cramér-Rao lower bound (CRLB) and an error analysis of the overall optimized multipoint ranging solution for wireless localization was performed, which was compared to simulated results quantified the substantial gains achieved by this technique.

3.2.3 Positioning Algorithms

Cooperative localization can significantly improve positioning performance in terms of accuracy and availability [101], when there are not adequate measurements from anchors whose positions are known. Cooperative positioning algorithms take into account range measurements not only from anchors but also from the unknown neighbors, and exchange positional information among cooperating users. This section presents some novel cooperative positioning algorithms, which propose efficient schemes to exchange positional information and model the uncertainties of neighbors' positions.

3.2.3.1 Target Localization with Algebraic Confidence via Circular Interval Scaling (CIS)

Introduction

In many localization applications it is desired that, together with the targets' position estimates, the algorithm return a measure of the confidence of such estimate. This feature is already intrinsic in Bayesian-based solutions [101] in which the entries of the state covariance matrix associated to the target's coordinate can be used to assess the confidence on the estimate. However, Bayesian approaches can be computationally costly and they rely on a priori information on the statistic of the observations.

This can be a limit, at least in some localization scenarios in which complexity must be kept as low as possible, *e.g.* to prevent battery depletion, and especially in cases where no a priori information is made available.

In this section we propose a novel algorithm that jointly estimates and utilize the confidence region associated to each estimated position.

To achieve this we define *algebraic confidence* as the measure of belief that is provided by the algebraic algorithm without any a priori information. We further demonstrate via illustration whereby

areas of the objects, *i.e.* circles as outputs through Circular Interval Scaling (CIS) algorithm proposed in [102] are related to the Fisher uncertainty ellipses.

After having combined *Fisher Ellipses* and CIS together, in Section 3.2.3.1 is shown how the new measure of uncertainty is utilized to design the improved cost function (CIS+). It is shown how the new algorithm, using the cost function develop here, outperforms the CIS algorithm in both, computational complexity and accuracy of the position estimate. Furthermore, in Section 3.2.3.2. the CIS+ cost function is extended to work in more realistic scenarios, namely to account scenario where some of the links are subject to NLoS conditions.

Preliminaries

Throughout the article a *network* is understood as a set of N nodes whose η-dimensional coordinates are presented in matrix form as $\mathbf{P} = [\mathbf{p}_1, \ldots, \mathbf{p}_N]$ where $\mathbf{p}_i \in \mathbb{R}^{1\times\eta}$ is the vector containing the coordinates for the i-th node. Without loss of generality we restrict ourselves to the case of $\eta = 2$ from which it follows that $\mathbf{p}_i = [x_i, y_i]$.[4]

It is assumed that $n_{\mathsf{A}} > \eta$ out the N nodes in the scenario have location known *a priori*. These will be referred from now on as the set of *anchor* nodes and without loss of generality it is assumed that their coordinates will occupy the first n_A entries in $\mathbf{P}$. Correspondingly the indexes for the remaining nodes, namely the *target* nodes, are labeled as $[n_{\mathsf{A}} + 1, \ldots, N]$.

Assuming that a measure of mutual dissimilarity between all nodes in the network is known in the form of a dissimilarity matrix, then this information suffices to recover the targets' locations by solving the resulting *scaling* problem [103]. This can be solved either relying on algebraic procedures, *e.g.* using the Classical-Multidimensional Scaling (C-MDS) method [104], or by optimization of a specific cost function [105]. With reference to the latter, a popular choice is the so called STRESS cost function given by[5]

$$\mathcal{C}(\mathbf{P}) = \sum_{i<j}^{N} w_{ij} \left[\delta_{ij} - d_{ij}(\mathbf{P})\right]^2, \tag{62}$$

where δ_{ij} is the dissimilarity measure between objects i and j, w_{ij} is the corresponding non-negative coefficient weighting the confidence on the measure and $d_{ij}(\cdot)$ is the function describing the dissimilarity measure between the i-th and j-th rows of $\mathbf{P}$.

Circular Interval-Based SMACOF algorithm

Amongst the algorithm that can be used to solve STRESS, a popular choice goes under the name of Scaling by Majorizing a Complicated Function (SMACOF) [107, 108].

The Circular-based Interval SMACOF (CIS) [102] provides an extension to the STRESS cost function used in SMACOF. Specifically in the CIS framework, rather than considering a set of dissimilarities between agents in the network, *i.e.* range measurements between points, CIS extends the definition of object by describing those, rather than points, *circles* in the Euclidean space.

This is accomplished by modifying the cost function defined in equation (62) into

$$\mathcal{C}(\mathbf{P}, \mathbf{R}) = \sum_{i<j}^{N} w_{ij} \left[\delta_{ij}^{(\mathsf{L})} - d_{ij}^{(\mathsf{L})}(\mathbf{P}, \mathbf{R})\right]^2 + \sum_{i<j}^{N} w_{ij} \left[\delta_{ij}^{(\mathsf{U})} - d_{ij}^{(\mathsf{U})}(\mathbf{P}, \mathbf{R})\right]^2, \tag{63}$$

where the matrix $\mathbf{R} \in \mathbb{R}^{N\times\eta_0}$ describes the size of the N objects parameterized by variable η_0, $\delta_{ij}^{(\mathsf{L})}$ and $\delta_{ij}^{(\mathsf{U})}$ are lower and upper bounds of an specific interval[6], corresponding respectively to

[4] The generalization to higher dimensions is straightforward.

[5] Notice that in presence of white Gaussian noise $\sim \mathcal{N}(0, \sigma^2)$, the solution of STRESS is equivalent to the maximum likelihood solution [106].

[6] The case of the single measurement can be seen as the extreme case, in which the $\delta_{ij}^{(\mathsf{L})} = \delta_{ij}^{(\mathsf{U})}$.

the minimum and maximum dissimilarity measured between the objects, and $d_{ij}^{(\mathsf{L})}(\cdot)$ and $d_{ij}^{(\mathsf{U})}(\cdot)$ are functions returning the estimated lower (minimum) and upper (maximum) distance between the objects.

In [102] it was shown that by considering the objects as circles, the matrix $\mathbf{R} \in \mathbb{R}^{N\times 1}$ reduces to a vector $\mathbf{r} = [r_1, \ldots, r_N]$ containing the radius of the circles associated to the node i in our network. Additionally lower and upper distances, $d_{ij}^{(\mathrm{L})}$ and $d_{ij}^{(\mathrm{U})}$ respectively, are defined as

$$d_{ij}^{(\mathrm{L})}(\mathbf{P}, \mathbf{r}) \quad = \max\left\{0, \|\mathbf{p}_i - \mathbf{p}_j\|_{\mathsf{F}} - (r_i + r_j)\right\}, \tag{64}$$

$$d_{ij}^{(\mathrm{U})}(\mathbf{P}, \mathbf{r}) \quad = \|\mathbf{p}_i - \mathbf{p}_j\|_{\mathsf{F}} + (r_i + r_j), \tag{65}$$

where $\mathbf{p}_i$ is i-th row of the coordinate matrix $\mathbf{P}$ and r_i is i-th element of the radius vector $\mathbf{r}$.

The complete derivation of a majorization-based [109, 110] algorithm to solve equation (63) was offered in [102] with defining minimum and maximum distances as given in equations (64) and (65). This solution was proved to be efficient and advantageous in our localization scenario when compared to different choices of objects' shapes [111].

Confidence Region: Fisher Uncertainty Ellipses

From estimation theory it is known that, for a given *confidence* P_e, the Fisher uncertainty ellipses yields the area within which the maximum likelihood (ML) estimate of $\mathbf{p}_i$ is ensured to be [112].

Let $\hat{\mathbf{p}}_i$ be the estimated location for the i-th target, then let the associated covariance matrix be

$$\mathbf{\Omega}_{\mathbf{p}_i} \triangleq \mathbb{E}\left[(\hat{\mathbf{p}}_i - \mathbf{p}_i)(\hat{\mathbf{p}}_i - \mathbf{p}_i)^{\mathsf{T}}\right]. \tag{66}$$

It follows that the lower bound on the estimator is represented by the CRLB which is given by

$$\mathbf{\Omega}_{\mathbf{p}_i} \succeq \mathbf{F}_{\mathbf{p}_i}^{-1}. \tag{67}$$

The matrix $\mathbf{F}_{\mathbf{p}_i}$ is the Fisher Information Matrix (FIM) [113, 114] and which (k, l)-th element is defined as [115]

$$[\mathbf{F}_{\mathbf{p}_i}]_{k,l} = \mathrm{E}\left[\frac{\partial \log f(\mathbf{\Delta}|\mathbf{P})}{\partial p_{i,k}} \frac{\partial \log f(\mathbf{\Delta}|\mathbf{P})}{\partial p_{i,l}}\right], \tag{68}$$

where matrix $\mathbf{\Delta}$ stores dissimilarity measurements δ_{ij} and $f(\cdot|\cdot)$ is the joint conditional probability density function.

Under the assumption that the ranging error has zero mean with spatial components in the "x" and "y" dimensions, namely $\sigma_{i,\mathsf{x}}^2$, $\sigma_{i,\mathsf{y}}^2$, as well as their cross-term $\sigma_{i,\mathsf{xy}}$ known, then equation (66) can be rewritten as

$$\mathbf{\Omega}_{\mathbf{p}_i} \triangleq \begin{bmatrix} \sigma_{i,\mathsf{x}}^2 & \sigma_{i,\mathsf{xy}} \\ \sigma_{i,\mathsf{xy}} & \sigma_{i,\mathsf{y}}^2 \end{bmatrix}. \tag{69}$$

It is known that the directions of maximum *dispersion* in the space for the random vector $\hat{\mathbf{p}}_i$ are proportional, up to a factor κ_i, to the eigenvalues associated to $\mathbf{\Omega}_{\mathbf{p}_i}$ [112]. In particular, the axis of the ellipse that better describes such dispersion in the space are given by $2\sqrt{\kappa_i \lambda_{i:1}}$, $2\sqrt{\kappa_i \lambda_{i:2}}$, where

$$\lambda_{i:1} \triangleq \frac{1}{2}\left[\sigma_{i,\mathsf{x}}^2 + \sigma_{i,\mathsf{y}}^2 + \sqrt{(\sigma_{i,\mathsf{x}}^2 - \sigma_{i,\mathsf{y}}^2)^2 + 4\sigma_{i,\mathsf{xy}}^2}\right], \tag{70}$$

$$\lambda_{i:2} \triangleq \frac{1}{2}\left[\sigma_{i,\mathsf{x}}^2 + \sigma_{i,\mathsf{y}}^2 - \sqrt{(\sigma_{i,\mathsf{x}}^2 - \sigma_{i,\mathsf{y}}^2)^2 + 4\sigma_{i,\mathsf{xy}}^2}\right]. \tag{71}$$

In presence of Gaussian random vectors, the proportionality factor κ_i is related to probability P_e through

$$\kappa_i = -2\ln(1 - P_e). \tag{72}$$

It follows that the Fisher Ellipse for the i-th target is given by [112]

$$\frac{[(x-p_{i,\mathsf{x}})\cos\gamma_i+(y-p_{i,\mathsf{y}})\sin\gamma_i]^2}{\kappa_i\cdot\lambda_{i:1}}+\frac{[(x-p_{i,\mathsf{x}})\sin\gamma_i-(y-p_{i,\mathsf{y}})\cos\gamma_i]^2}{\kappa_i\cdot\lambda_{i:2}}=1, \tag{73}$$

where the *rotation angle* γ_i describes the offset between the principal axis for the ellipse and reference axis and it is defined as[7]

$$\gamma_i \triangleq \frac{1}{2}\arctan\left(\frac{2\sigma_{i,\mathsf{xy}}}{\sigma^2_{i,\mathsf{x}}-\sigma^2_{i,\mathsf{y}}}\right). \tag{74}$$

Confidence Region: Circular Interval Scaling

As discussed in previous section, $\hat{\mathbf{p}}_i$ is influenced by both the error on the measurements as well as the geometrical configuration which is captured by the derivatives with respect to the coordinate components of $\hat{\mathbf{p}}_i$ in equation (68).

To uncover the dependences of the bound from the noise and geometry characterising the scenario, similarly to [116] we rewrite the CRLB explicitly in terms of these two components. For the sake of simplicity however, in the following a source localization scenario is considered. Let N_{A} denote the number of anchors and $\hat{\mathbf{u}}_\mathbf{i}=u_i^{(x)}\hat{\mathbf{x}}+u_i^{(y)}\hat{\mathbf{y}}$ the unit vector[8] originating from the target's location and lined up towards the i-th anchor. Then $\mathcal{M}_{\mathbf{\Omega}_{\mathbf{p}_i}}$ can be written as [117]

$$\mathcal{M}_{\mathbf{\Omega}_{\mathbf{p}_i}}=\sqrt{\frac{A}{\det\{\mathbf{M}\}}}, \tag{75}$$

where $A=\sum_{i=1}^{N_{\mathsf{A}}}\left(\sigma_i^{-2}(u_i^{(x)})^2+\sigma_i^{-2}(u_i^{(y)})^2\right)$ and $\mathbf{M}^{-1}=\mathbf{\Omega}_{\mathbf{p}_i}$.

Also equation (75) can be rewritten as [116]

$$\mathcal{M}^2_{\mathbf{\Omega}_{\mathbf{p}_i}}=\frac{\sum_{i=1}^{N_{\mathsf{A}}}\sigma_i^{-2}}{\det\{\mathbf{M}\}}, \tag{76}$$

where

$$\det\{\mathbf{M}\}=\sum_{i=1}^{N_{\mathsf{A}}}\sum_{\substack{j=1\\j>i}}^{N_{\mathsf{A}}}\sigma_i^{-2}\sigma_j^{-2}|\hat{\mathbf{u}}_\mathbf{i}\times\hat{\mathbf{u}}_\mathbf{j}|^2. \tag{77}$$

By the definition of cross product and considering a polar coordinate system then the term $|\hat{\mathbf{u}}_\mathbf{i}\times\hat{\mathbf{u}}_\mathbf{j}|$ can be rewritten as $|\hat{\mathbf{u}}_\mathbf{i}\times\hat{\mathbf{u}}_\mathbf{j}|=|\hat{\mathbf{u}}_\mathbf{i}||\hat{\mathbf{u}}_\mathbf{j}||\sin(\theta_{ij})|$, where θ_{ij} is the angle between the i-th and j-th anchor as seen from the target.

Now, let denote

$$\mathcal{A}_{ij}=|\hat{\mathbf{u}}_\mathbf{i}||\hat{\mathbf{u}}_\mathbf{j}||\sin\theta_{ij}|, \tag{78}$$

and Equation (76) can be rewritten

$$\mathcal{M}^2_{\mathbf{\Omega}_{\mathbf{p}_i}}=\frac{\sum_{i=1}^{N_{\mathsf{A}}}\sigma_i^{-2}}{\sum_{i=1}^{N_{\mathsf{A}}}\sum_{\substack{j=1\\j>i}}^{N_{\mathsf{A}}}\sigma_i^{-2}\sigma_j^{-2}\mathcal{A}_{ij}^2}. \tag{79}$$

Under the assumption that $\sigma_i=\sigma^{(d)}\ \forall i\in\{1,\dots,N_{\mathsf{A}}\}$, the CRLB can be written as [116]

$$\mathcal{M}_{\mathbf{F}^{-1}_{\mathbf{p}_i}}=\sigma^{(d)}\cdot\sqrt{\frac{N_{\mathsf{A}}}{\sum_{j>i}^{N_{\mathsf{A}}}\mathcal{A}_{ij}}}, \tag{80}$$

[7] For $\sigma^2_{i,\mathsf{x}}=\sigma^2_{i,\mathsf{y}}$ then $\gamma_i=0$.

[8] $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit vectors along the x- and y axis, respectively, and $u_i^{(x)}$ and $u_i^{(y)}$ denotes the x and y component of the vector $\mathbf{u}_i$.

where $\sum_{j>i}^{N_\mathsf{A}}$ is used to indicate $\sum_{i=1}^{N_\mathsf{A}} \sum_{\substack{j=1 \\ j<i}}^{N_\mathsf{A}}$.

Furthermore, noticing that by the definition of the unit vector, *i.e.* $|\mathbf{u}| = 1$, then equation (78) simplifies to $\mathcal{A}_{ij} = |\sin(\theta_{ij})|$.

It follows that $\mathcal{M}_{\mathbf{F}_{\mathbf{p}_i}^{-1}}$ can be expressed as the product of two components, namely the error depending on the measurements' noise and a geometric term depending on the angular distribution of nodes in the scenario under investigation, *i.e.* $\{\theta_{ij}\}$ which is also known in the literature as the geometric dilution of precision (GDOP), namely

$$\mathcal{M}_{\mathbf{F}_{\mathbf{p}_i}^{-1}} = \sigma^{(d)} \cdot \mathrm{GDOP}_{(i)}, \tag{81}$$

with

$$\mathrm{GDOP}_{(i)} = \sqrt{\frac{N_\mathsf{A}}{\sum_{j>i}^{N_\mathsf{A}} |\sin(\theta_{ij})|}}. \tag{82}$$

It follows that an ideal measure of *confidence* for the position estimate is given by $\rho_{\mathbf{F}_{\mathbf{p}_i}^{-1}} = 1/\mathcal{M}_{\mathbf{F}_{\mathbf{p}_i}^{-1}}$.

Improved Circular-Confidence -Based Scaling (CIS+)

The idea of the CIS+ framework is to develop an algorithm that, similarly to the error ellipses associated to the CRLB [112], quantifies the confidence region associated to each location estimate as one of the output of localization process and uses it as a priori information in subsequent computations. To achieve this, we use the CIS algorithm previously introduced in [102] to exploit the uncertainty measure provided by the $\mathbf{r}$ parameter in the computation of the targets' locations. Furthermore it was shown in [102] that the CIS implementation is convenient due to its low complexity and its sure convergence to local minima.

The core of the CIS algorithm is the extension to the standard STRESS cost function commonly utilized in localization problems to allow targets to be described by, rather than points, *areas* in the space [111]. To do so, in [102] it was proposed to generalize the STRESS cost function to include a matrix $\mathbf{R} \in \mathbb{R}^{N \times \eta_o}$ accounting for the *uncertainty* of the N objects. In particular, following the arguments in [118], the CIS cost function[9] can be modified to account the uncertainty information resulted from the algorithm, yielding to the CIS+ cost function. This is done by adding a term that exploits eventual a priori information on the targets estimates, namely

$$\mathcal{C}^{(+)}(\mathbf{P}, \mathbf{r}) = \sum_{i<j}^{N} w_{ij} \left[\tilde{d}_{ij}^{(\mathsf{L})} - d_{ij}^{(\mathsf{L})}(\mathbf{P}, \mathbf{r})\right]^2 + \sum_{i<j}^{N} w_{ij} \left[\tilde{d}_{ij}^{(\mathsf{U})} - d_{ij}^{(\mathsf{U})}(\mathbf{P}, \mathbf{r})\right]^2 + \sum_{i=1}^{N} \rho_i \|\mathbf{p}_i - \bar{\mathbf{p}}_i\|_\mathsf{F}^2; \tag{83}$$

where ρ_i represent a confidence measure over an initial target estimate $\bar{\mathbf{p}}_i$, $d_{ij}^{(\mathsf{L})}(\cdot)$ and $d_{ij}^{(\mathsf{U})}(\cdot)$ are functions returning the estimated minimum and maximum distance between the objects and $\tilde{d}_{ij}^{(\mathsf{L})}$ and $\tilde{d}_{ij}^{(\mathsf{U})}$ the corresponding variables used in the optimization.

Complexity

The advantage of the CIS+ algorithm is that, it reduces drastically the total computational complexity. Although, the CIS+ solution has an additional sum component with the matrix multiplication, it should be noted that the concerning matrix has only non-zero elements on its diagonal and thus the number of multiplications added top of CIS algorithm is very small.

The comparison of complexity costs is presented in Table 9. From the table it can be read that, depending on the scenario in the question, the number of iteration CIS+ requires to reach the same level of convergence is from 4 to 8 times less.

[9] In CIS algorithm objects are defined as circles, thus $\mathbf{R}$ is suppressed to the vector $\mathbf{r} \in \mathbb{R}^{N \times 1}$.

Table 9

Number of iterations[a]						
	CIS			CIS+		
$\sigma =$	0.1	0.2	0.4	0.1	0.2	0.4
Single[b]	75.4	52.2	42.3	9.3	9.8	10.5
Multi[b]	179	161	145	20.7	20.7	20.4
Single[c]	70.0	56.5	41.4	9.4	10.4	11.2
Multi[c]	133	119	112	22.5	22.8	24.9

[a] Number of iterations required from the both algorithms to reach same level of convergence.
[b] For the Scenarios where σ is constant.
[c] For the Scenarios where σ is varying from link to link.

Position Confidence with the CIS Algorithm

From the previous subsection we saw how the confidence depends on an error term and a geometric term. Following we first uncover the relationship between the sets of range measurements $\{d_{ij}^{(\mathsf{L})}, d_{ij}^{(\mathsf{U})}\}$ and $\mathbf{r}$, showing how the latter is related to the uncertainty associated to the relative target estimates $\hat{\mathbf{P}}$. This is then used together with the notion of GDOP defined in equation (82) inside the proposed *geometric* confidence.

Statistical Analysis of $\mathbf{r}$

Let the ij-th range measurements be $\tilde{d}_{ij} = d_{ij} + n_{ij}$, where $n_{ij} \sim \mathcal{N}(0, \sigma_d^2)$ is the random variable modeling the noise process for the ij-th link with the such that all n_{ij} in the network are independent and identically distributed (i.i.d.). Also let assume two observations are available for each one of the measured links, such that

$$\begin{aligned} \tilde{\delta}_{ij}^{(\mathsf{L})} &= \min\left\{\tilde{d}_{ij}^{(1)},\, \tilde{d}_{ij}^{(2)}\right\}, \\ \tilde{\delta}_{ij}^{(\mathsf{U})} &= \max\left\{\tilde{d}_{ij}^{(1)},\, \tilde{d}_{ij}^{(2)}\right\}. \end{aligned} \tag{84}$$

Let r_i be the i-th entry of $\mathbf{r}$, namely the radius referred to the set of measured ranges between the i-th target and the j-th sensors connected to it. To retrieve the distribution of r_i, the first problem is to characterize the distributions of $\tilde{\delta}_{ij}^{(\mathsf{L})}$ and $\tilde{\delta}_{ij}^{(\mathsf{U})}$ defined in equation (84). Due to the equivalence of analysis of these two cases, we restrict ourselves to the study $\tilde{\delta}_{ij}^{(\mathsf{U})}$, keeping in mind that similar steps can be done for $\tilde{\delta}_{ij}^{(\mathsf{L})}$.

To begin with, consider the anchor-to-target measurements case, *i.e.* the source localization problems in which $\tilde{\delta}_i$ refers to the i-th anchor from which the observation is generated.[10]

Let X_1, X_2 denote two Gaussian random variables with means μ_1, μ_2, standard deviations σ_1, σ_2 and correlation coefficient ρ. Using order statistics notation [119] let $r^{(2)}$ define the random variable resulting from $\max(X_1, X_2)$.[11]

[10] Namely there is no uncertainty associated with the anchors' locations.
[11] Similarly $r^{(1)}$ will be used to denote the random variable corresponding to $\min(X_1, X_2)$.

Then the distribution of $r_i^{(2)}$ for to the i-th link is [120]

$$p_{r_i^{(2)}}(x;\mu_1,\mu_2,\sigma_1,\sigma_2,\rho) = \frac{1}{\sigma_1}\phi\left(\frac{\mu_1-x}{\sigma_1}\right)\Phi\left(\frac{\rho(\mu_1-x)}{\sigma_1\sqrt{1-\rho^2}}-\frac{\mu_2-x}{\sigma_2\sqrt{1-\rho^2}}\right) + \frac{1}{\sigma_2}\phi\left(\frac{\mu_2-x}{\sigma_2}\right)\Phi\left(\frac{\rho(\mu_2-x)}{\sigma_2\sqrt{1-\rho^2}}-\frac{\mu_1-x}{\sigma_1\sqrt{1-\rho^2}}\right), \tag{85}$$

where $\phi(\cdot)$ and $\Phi(\cdot)$ are respectively the PDF and the CDF of the standard Gaussian distribution.[12]

While equation (85) provides the distribution of the single $\tilde{\delta}_{ij}^{(\mathsf{U})}$, from the equation (83) it is evident that r corresponds to the linear combination of the two independent random variables $r^{(1)}$ and $r^{(2)}$, thus from equation (83) it follows that the distribution of r is given by

$$p_r(x;\cdot) = p_{r^{(1)}}(x;\cdot) \star p_{r^{(2)}}(x;\cdot), \tag{86}$$

where $\star$ denotes the convolution operator.

Also notice that, as indicated in equation (83), the random variable $r^{(2)}$ is also obtained averaging over all the anchor-to-target links, *i.e.* $\{\tilde{\delta}_i^{(\mathsf{U})}\}\ \forall i \in \{1,\ldots,N_{\mathsf{A}}\}$, which yields

$$p_{r^{(2)}}(x;\cdot) = p_{r_1^{(2)}}(x;\cdot) \star \ldots \star p_{r_{N_{\mathsf{A}}}^{(2)}}(x;\cdot). \tag{87}$$

Let the scaled and shifted *skew*-normal distribution [121] be

$$p(x;\mu,\sigma,\alpha) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\alpha\left(\frac{x-\mu}{\sigma}\right)\right), \tag{88}$$

and its excess kurtosis γ be

$$\gamma = 2(\pi-3)\frac{(\varrho\sqrt{2/\pi})^4}{(1-2\varrho^2/\pi)^2}, \tag{89}$$

where $\varrho = \alpha/\sqrt{1+\alpha^2}$ with shape parameter α.

Since the noise process n over all links is independent and identically distributed (i.i.d.), then $p_{r_i^{(2)}}(x;\cdot)$ defined in equation (85) is exactly a skew normal distribution with shape parameter $\alpha = 1$, location parameter $\mu = d_{ij}$ and scale parameter $\sigma = \sigma_d$. Moreover, being $\gamma = 0.0617$, then $p_{r_i^{(2)}}(x;\cdot)$ can be well approximated by a Gaussian distribution which first two moments are

$$\mu_i^{(2)} = \mathsf{E}[r_i^{(2)}] = \mu + \sigma\varrho\sqrt{\frac{2}{\pi}}, \tag{90}$$

$$\sigma_i^{(2)} = \mathsf{E}[(r_i^{(2)} - \mathsf{E}[r_i^{(2)}])^2] = \sigma^2\left(1-\frac{2\varrho^2}{\pi}\right). \tag{91}$$

As consequence of this approximation, both expressions in equations (86) and (87) reduces to the sum of Gaussian random variables, namely the PDF of r defined in equation (86) is approximately $r_* \sim \mathcal{N}(\mu_n^{(r)}, \sigma_n^{(r)})$ where

$$\mu_n^{(r)} = \frac{1}{2N_{\mathsf{A}}}\sum_{i=1}^{N_{\mathsf{A}}}(\mu_i^{(2)} - \mu_i^{(1)}), \tag{92}$$

$$\sigma_n^{(r)} = \frac{1}{2N_{\mathsf{A}}}\Big(\sum_{i=1}^{N_{\mathsf{A}}}\left((\sigma_i^{(1)})^2 + (\sigma_i^{(2)})^2\right)\Big)^{1/2}. \tag{93}$$

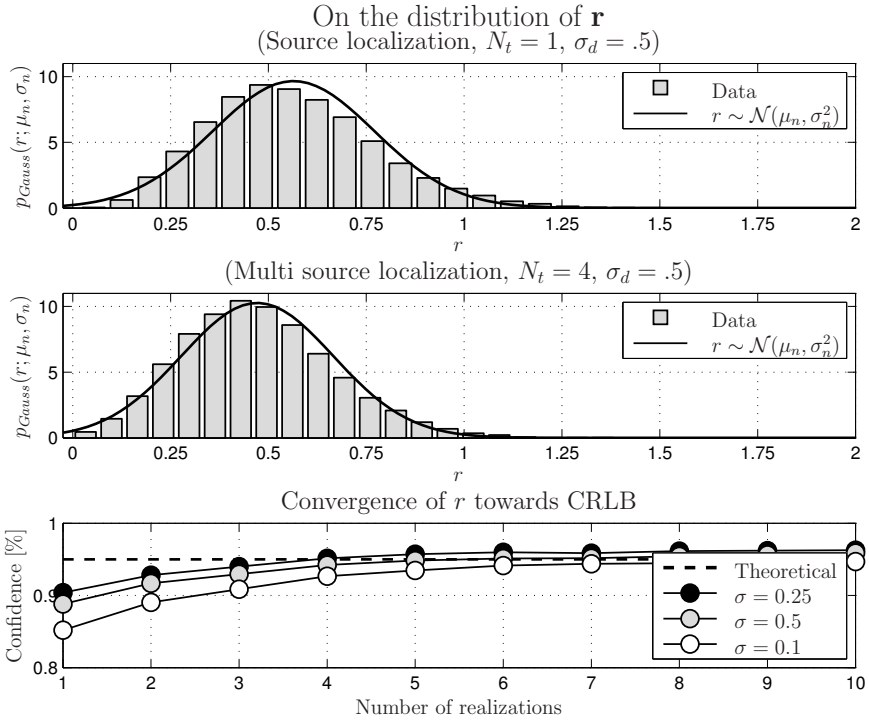


Figure 45: Distribution of $\mathbf{r}$ parameter derived from algorithm in single target scenario.

The distribution of r using the approximation above is compared versus the proposed approximation and data generated by the CIS algorithm in the first two subplots in Figure 45.

The simulations were carried out in the scenario, where 4 anchor nodes are located at corners of the 5×5 square and the target nodes are located in the fixed positions, as is illustrated in Figure 46. In the first subplot, the distribution of $\mathbf{r}$ is derived in the source location scenario with Gaussian noise affecting to the range measurement characterized by $\sigma_d = .5$. The second subplot shows the results from a multi target scenario with 4 targets in fixed positions with all like perturbed by Gaussian noise with $\sigma_d = .5$.[13] From these subplots it can be appreciated that the distribution of $\mathbf{r}$ in both scenarios is well approximated by the derived estimates, which supports the validity of the analysis. The third subplot of Figure 45 shows, as a function of number of measurements per link, the consistency of r to match the confidence $P_e = 95\%$, where P_e represents the certainty that an estimate is enclosed within an error ellipse [112]. In particular it can be appreciated how, even for a single pair of measurements per link, the confidence levels are high, suggesting that a lot of the uncertainty over the $\hat{\mathbf{P}}$ is captured by $\mathbf{r}$.

Algebraic Confidence from CIS algorithm

Given a set of range measurements $\mathcal{S} = \{\{d_{ij}^{(L)}\}, \{d_{ij}^{(U)}\}\}$, from previous subsection showed that the distribution of $r_i^{(k)}$ is well approximated by Gaussian distribution whose moments are given in equation (92). Therefore a measure of confidence $\hat{\rho}_i$ for the i-th target estimate $\hat{\mathbf{p}}_i$ be expressed as the inverse of $\mathcal{M}_{(r_i, \hat{\mathbf{P}}^{(-1)})}$ where

$$\mathcal{M}_{(r_i, \hat{\mathbf{P}}^{(-1)})} = \mu_n^{(i)} \cdot \tilde{\mathrm{GDOP}}_i, \tag{94}$$

[12] The PDF and the CDF for the standard Gaussian are defined as $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ and $\Phi(x) = \frac{1}{2}\left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}}\right)\right]$ respectively.

[13] In presence of the i-th link as a target-to-target measurement the statistic of r is modified to account for the noise term σ_i as equally shared amongst the two targets' radiuses.

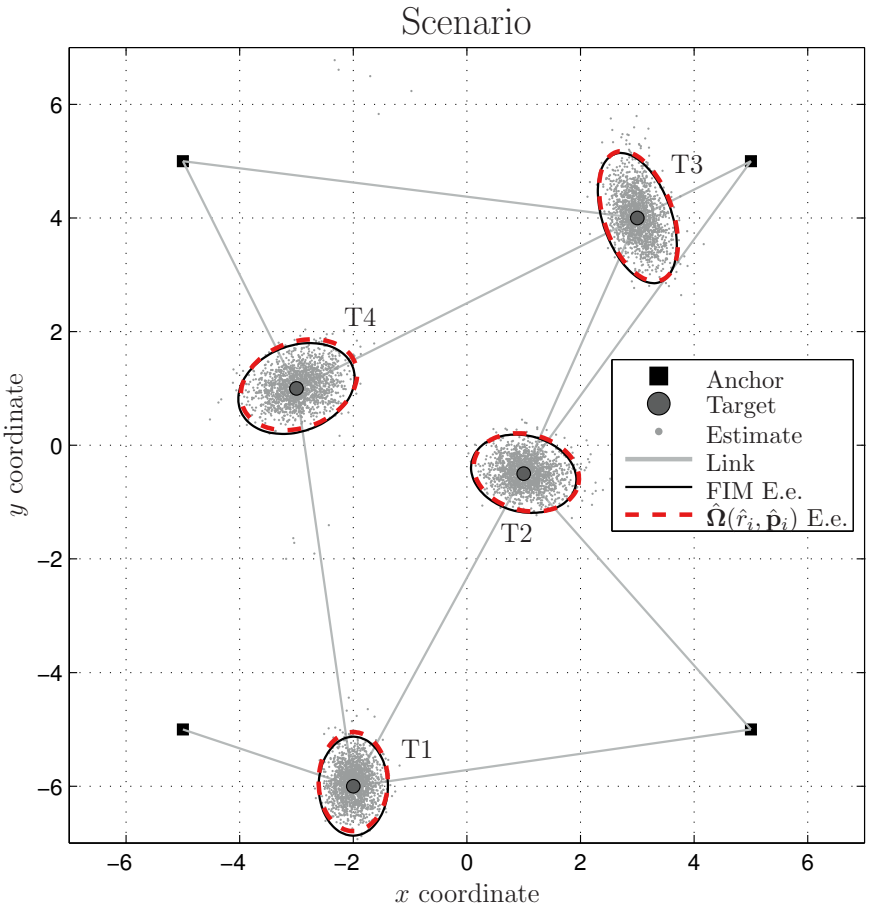


Figure 46: Illustration of simulation scenario having 4 targets with position estimate results from CIS+ algorithm as well as Fisher error ellipses and error ellipses derived from CIS+ algorithm.

where $\mu_n^{(i)}$ refers to the equation (92) for i-th target, and $\tilde{\text{GDOP}}_i$ is the approximation of equation (82) computed on the basis of $\bar{\mathbf{P}}$ using the estimated coordinates for the nodes connected to it, to compute the geometric factor.

In order to evaluate the performance of the proposed algorithms, we compare those against SMACOF algorithm [107] in the scenario depicted in Figure 46. More specifically, two different versions of SMACOF algorithm are considered, namely Smacof-2 and Smacof-4, which differs in such a way that the first and the latter are averaging over set of 2 and 4 measurements per link, respectively. The CIS algorithm uses set of 2 measurements as shown in equation (84), while CIS+ algorithm uses in addition 2 measurements per link to gain a priori knowledge of $\mathbf{P}$.

The limited connectivity of targets is illustrated in Figure 46, while anchors are assumed to have full connectivity amongst each others. All algorithms are fed with the same initial starting point. Furthermore, CIS$+^*$ uses the perfect confidence measure in a sense that it is derived via CRLB analysis using estimated points, thus providing a lower bound for that information.

Figure 47 presents results from the multi target localization problem, in which CIS+ as well as CIS algorithms outperforms both SMACOF solutions. In addition, CIS+ solution is shown to be close to the ideal solution.

Conclusion

It was shown that using the Circular Interval-Scaling (CIS) framework it is possible to quantify the confidence of each location estimate in a way consistent with the error ellipses associated to the Cramér-Rao Lower Bound (CRLB). This information can then be included in the optimization by extending the cost function resulting in the Improved Circular-Confidence -based Scaling (CIS+)

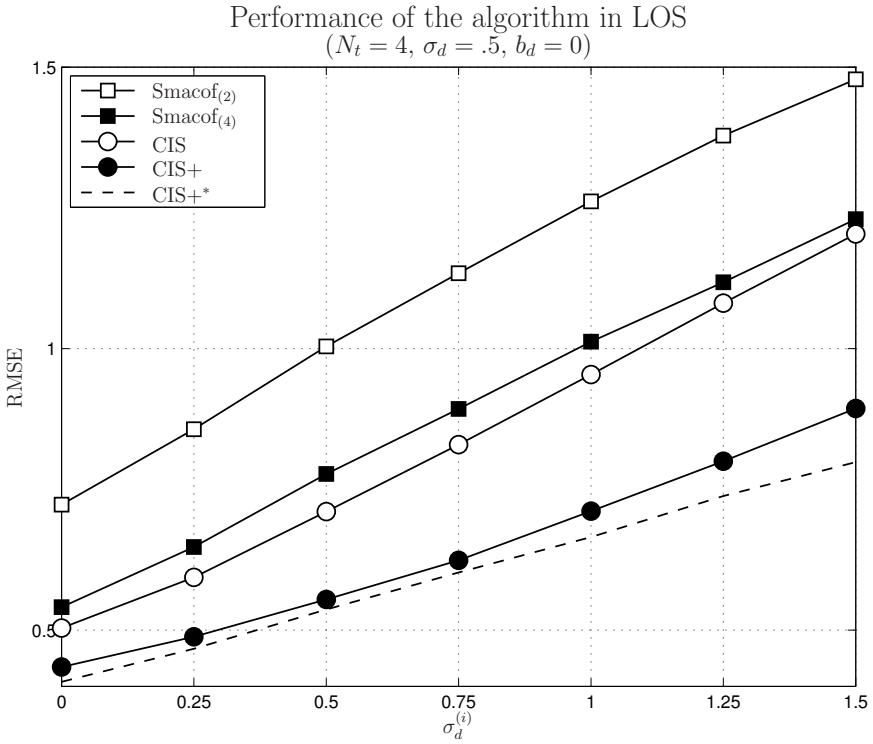


Figure 47: Comparison of CIS and extended CIS algorithms in scenario with 4 Anchors nodes and 4 Target nodes, where 1 selected target is affected by higher noise.

algorithm which still benefits of the low computational complexity and the convergence properties of the CIS solution.

3.2.3.2 Robust Positioning with CIS+

In realistic scenario the possibility that some of the links are subject to NLoS conditions must be accounted for [101]. To address this problem, in the following the least-square terms in the CIS+ cost function defined in equation (83) are replaced with terms favoring sparsity. In particular the Huber function [122] was chosen due to its well known properties to reject outliers.

The Huber function, which can be seen as an interpolation between ℓ_1 -norm and ℓ_2 -norm minimizations, is defined as [123]

$$\mathcal{H}_\xi(x) = \begin{cases} \frac{1}{2}\,x^2 & , \quad |x| < \xi \\ \xi\ |x| - \frac{1}{2}\xi^2 & , \quad |x| \geq \xi \end{cases}, \tag{95}$$

where ξ is given Huber threshold.[14]

Replacing (95) into equation (83) yields to

$$\mathcal{C}^{(\mathsf{H+})}(\mathbf{P},\mathbf{r}) = \sum_{i<j}^{N} w_{ij}\mathcal{H}_\sigma\left(\left[\delta_{ij}^{(\mathsf{L})} - d_{ij}^{(\mathsf{L})}(\mathbf{P},\mathbf{r})\right]\right) \quad + \sum_{i<j}^{N} w_{ij}\mathcal{H}_\sigma\left(\left[\delta_{ij}^{(\mathsf{U})} - d_{ij}^{(\mathsf{U})}(\mathbf{P},\mathbf{r})\right]\right) + \sum_{i=1}^{N}\rho_i\|\mathbf{P}_i - \bar{\mathbf{P}}_i\|_{\mathsf{F}}^2. \tag{96}$$

The cost function above is referred to as the *Robust-CIS+* (R-CIS+) due to its intrinsic ability to reject biased range measurements.

To compare the performance of R-CIS+ versus CIS+ we considered scenarios in which some of the range measurements are modeled as $\tilde{d}_i = d_i + n_i + b_i$, where $b_i \sim \mathcal{U}(0, b_d)$.

[14] Within simulation result presented here, $\xi = 0.01$ was used.

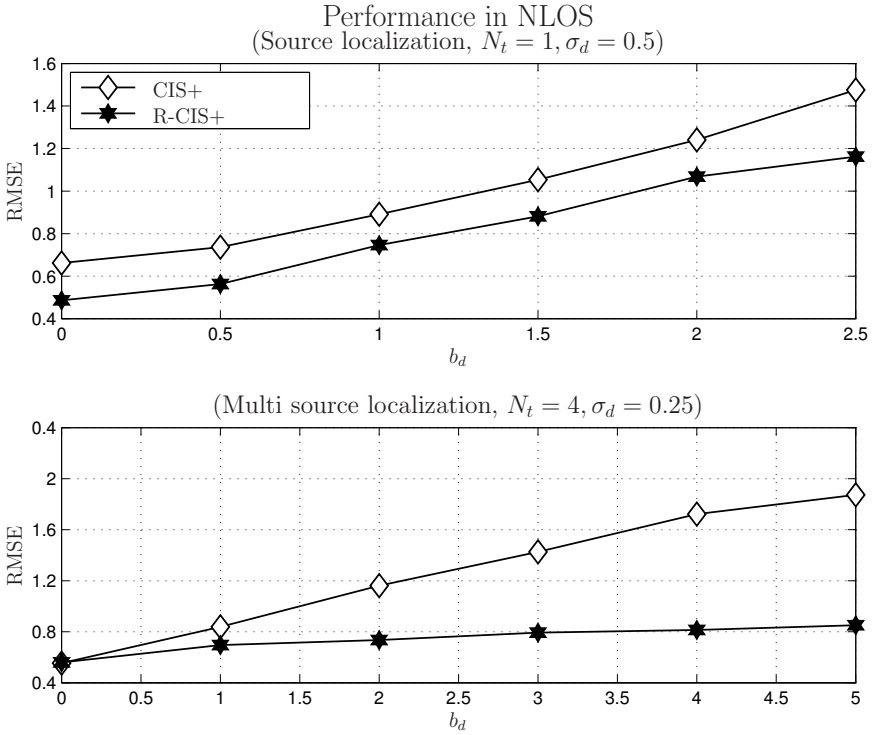


Figure 48: Comparison of CIS+ and R-CIS+ algorithms in a source scenario with 1 link affected by biased range observations and the multi target scenario depicted in Figure 46 with 3 randomly selected links as biased.

Figure 48 offers an insight of the performance of the R-CIS+ algorithm against CIS+ algorithm as function of b_d. In the first subplot refers to a source localization problem with 4 anchors placed at the corner of a square and in which all 4 observed measurements are perturbed by Gaussian noise with $\sigma_d = .5$ and one distance is affected by an additive bias.

In the second subplot the scenario shown in Figure 46 was considered. In particular all the links were perturbed with Gaussian noise ($\sigma_d = .25$) amongst them, to 3 randomly selected links were added a bias term.

Results in both scenarios show that the R-CIS+ always outperforms the CIS+ solution, especially in presence of more complex topologies.

Conclusion
By changing the least-square terms in the Circular Interval-Scaling (CIS) cost function proposed in 3.2.3.1 with norm favoring sparsity implemented using the Huber function, we extended the framework to handle range-based positioning problems in mesh networks affected by biased measurements, *i.e.* Non-Line-of-Sight (NLoS) conditions.

3.2.3.3 Cooperative NLoS Detection for Positioning Improvement

NLoS propagation of Radio Frequency (RF) signal has proven to be challenging for the localization of unknown nodes in wireless networks. In particular, the blockage of RF signal can heavily affect the accuracy of range measurements performed between nodes and in turn may cause the position estimation diverging. Based on contextual geo-localization principle defined in [124], this subsection proposes a cooperative NLoS identification scheme as well as a cooperative positioning algorithm based on belief propagation. The proposed algorithm is fully distributed and does not need to know the state of NLoS range measurements [125].

3.2.3.3.1 Measurement Modeling

Concerning range measurement models, this work adopted models presented in [126] as they have been extracted from experimental measurements by using UWB modules [127].

Range measurements in LoS condition are assumed as Gaussian distributed:

$$\tilde{r} = d + n_{\text{los}}, \tag{97}$$

where d is the exact distance between the two nodes involved in the measurement, and n_{los} is a Gaussian distributed noise, $n_{\text{los}} \sim \mathcal{N}(0, \sigma^2)$, with zero mean and standard deviation $\sigma = 0.25$ m.

Range measurements in NLoS condition are modeled as:

$$\tilde{r} = d + n_{\text{nlos}}. \tag{98}$$

where n_{nlos} is the measurement noise supposed to be exponentially distributed, $p_{n_{\text{nlos}}}(x) = \lambda \exp(-\lambda x)$ when $x \geq 0$, with rate parameter $\lambda = 0.38$ m^{-1}.

A generic range measurement can be performed either in LoS condition with probability P or in NLoS condition with probability $1 - P$. [15] Let $s_{n\to m}$ be the state associated to the range measurement $\tilde{r}_{n\to m}$ from neighbor n to mobile m. The state $s_{n\to m}$ is defined as 0 if the corresponding range measurement is performed in LoS condition or as 1 in NLoS condition. As a consequence, $P(s_{n\to m} = 0) + P(s_{n\to m} = 1) = 1$.

Based on the above definitions, the likelihood function of the range measurement could be simply expressed as the weighted sum on the state:

$$p(\tilde{r}_{n\to m}|\mathbf{x}_m, \mathbf{x}_n) = \sum_{i=0}^{1} P(s_{n\to m} = i)\, p(\tilde{r}_{n\to m}|\mathbf{x}_m, \mathbf{x}_n, s_{n\to m}), \tag{99}$$

where $\mathbf{x}_m = [x_m, y_m]$ is the position of the mobile m and $\mathbf{x}_n = [x_n, y_n]$ the position of the neighbor n. [16] Note that the likelihood function could be either a normal distribution or an exponential one depending on the link condition:

$$p(\tilde{r}_{n\to m}|\mathbf{x}_m, \mathbf{x}_n, s_{n\to m}) = \begin{cases} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\tilde{r}_{n\to m} - \|\mathbf{x}_n - \mathbf{x}_m\|)^2}{2\sigma^2}\right), & \text{if } s_{n\to m} = 0 \\ \lambda \exp\left(-\lambda\left(\tilde{r}_{n\to m} - \|\mathbf{x}_n - \mathbf{x}_m\|\right)\right). & \text{if } s_{n\to m} = 1 \end{cases} \tag{100}$$

where $\|\cdot\|$ denotes the Euclidean distance.

Some NLoS identification techniques presented in literature are based on the processing of the received signal [127, 128], but they are so complex and that it may be infeasible to be implemented on cheap devices. Since range measurements are correlated with the position of the mobile, it would be an efficient way to proceed in parallel both the localization and NLoS estimation for all the involved range measurements.

3.2.3.3.2 Message Passing Algorithm

Since there is no prior information about the state of each range measurement, the basic idea would be to use range measurements to infer first the mobile's position, then the state of range measurements. Alternatively, in order to improve positioning accuracy, both mobiles' positions and links' states can be estimated in parallel through some iterations of the Belief Propagation (BP)

[15] Only two states of measurements are considered here and it is straightforward for the extension to more states.

[16] Here only 2D localization is considered and the extension to 3D case is clear.

algorithm. However, this approach has some drawbacks. One is the network traffic generated by the cooperation packets (note that the size of communication messages depends on the number of particles used to approximate the distribution of mobile's position). Another drawback is the computational effort required to calculate the integral of neighbor's belief. The proposed algorithm assumes that the belief of mobile's position is Gaussian distributed, thus the mobile just needs to send to its neighbors the estimated position and the corresponding uncertainty. This approach known as Expectation Propagation (EP) [129] is an approximation of the BP algorithm. Based on this assumption, we propose a cooperative NLoS identification and positioning algorithm, namely cooperative NLoS identification and positioning algorithm. In the following sections, the message passing for a generic mobile m is introduced.

Incoming Messages

The localization approach is based on the factor graph depicted in Fig. 49. In particular, the joint posterior distribution can be factorized by messages from anchor nodes and mobile neighbors as

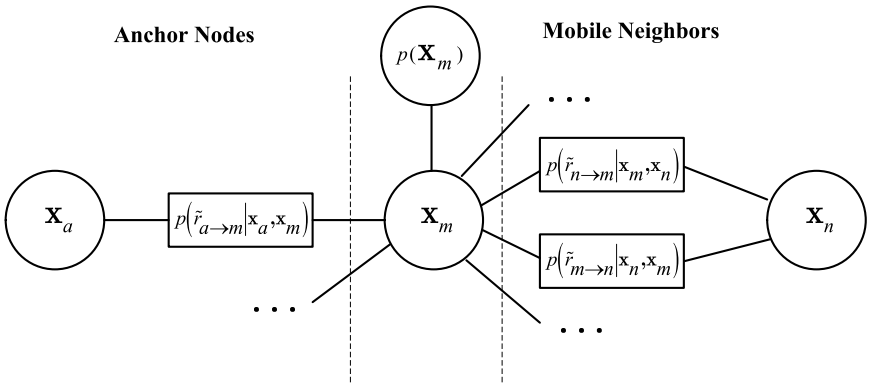


Figure 49: Factor graph for cooperative positioning.

Message from Anchor: The incoming message from an anchor $a \in \mathcal{A}_m$ is proportional to the integral of the multiplication between the likelihood function and the belief of the anchor that is a Dirac delta function centered on $\mathbf{x}_a$, *i.e.*, $b(\mathbf{x}_a) = \delta(\mathbf{x} - \mathbf{x}_a)$:

$$\begin{aligned} \mu_{a\to m} &\propto \int p(\tilde{r}_{a\to m}|\mathbf{x}_m, \mathbf{x}_a) b(\mathbf{x}_a) d\mathbf{x}_a \\ &= p(\tilde{r}_{a\to m}|\mathbf{x}_m, \mathbf{x}_a), \end{aligned} \tag{101}$$

When referring to more than one state, the likelihood function can be calculated by using (99), thus $p(\tilde{r}_a|\mathbf{x}_m, \mathbf{x}_a)$ becomes

$$p(\tilde{r}_{a\to m}|\mathbf{x}_m, \mathbf{x}_a) = \sum_{i=0}^{1} P\left(s_{a\to m} = i\right) p(\tilde{r}_{a\to m}|\mathbf{x}_m, \mathbf{x}_a, s_{a\to m}), \tag{102}$$

Message from Mobile Neighbor: Similarly, the incoming message from a mobile neighbor can be expressed as:

$$\mu_{n\to m} \propto \int p(\tilde{r}_{n\to m}|\mathbf{x}_m, \mathbf{x}_n) b(\mathbf{x}_n) d\mathbf{x}_n, \tag{103}$$

Since the mobile neighbor's position x_n has some uncertainties, the belief $b(\mathbf{x}_n)$ is not a Dirac delta function. In principle, the belief can be approximated by the samples (*e.g.*, particle filter [130]). If so, the calculation of (103) may be too complex to be performed in a mobile node. In order to simplify that calculation, some approaches, presented in [126], assume that $b(\mathbf{x}_n)$ is Gaussian distributed.

In order to further reduce the complexity, the belief of the mobile neighbor n is approximated as a Dirac delta function (*i.e.*, the mobile neighbor is considered as an anchor), $b(\mathbf{x}_n) \approx \delta(\mathbf{x} - \hat{\mathbf{x}}_n)$. To compensate this approximation, the position uncertainty associated to neighbor n is considered as an additional noise to be evaluated on the range measurement $\tilde{r}_{n\to m}$. More specifically, the variance associated to a range (given by σ^2 for LoS measurements or $1/\lambda^2$ for NLoS measurements) is increased by the uncertainty of the mobile's neighbor. For simplicity, this uncertainty is calculated as the trace of the estimated covariance matrix [130], *i.e.*, trace$(\mathbf{P}_n)$. As a consequence, the new parameters σ_{nm} and λ_{nm} to be used in the likelihood function are given by (104) and (105), respectively.

$$\sigma_{nm} = \sqrt{\sigma^2 + \text{trace}(\mathbf{P}_n)}, \tag{104}$$

$$\lambda_{nm} = \frac{\lambda}{\sqrt{1 + \lambda^2 \text{trace}(\mathbf{P}_n)}}, \tag{105}$$

To sum up, by using the above approximation, the incoming message from mobile neighbor is

$$\mu_{n\to m} \propto p(\tilde{r}_{n\to m}|\mathbf{x}_m, \hat{\mathbf{x}}_n). \tag{106}$$

where $p(\tilde{r}_{n\to m}|\mathbf{x}_m, \hat{\mathbf{x}}_n)$ is the likelihood function evaluated by using the new modified parameters σ_{nm} and λ_{nm} that take into account the uncertainty of mobile neighbor n.

Position Estimate

The mobile node can calculate its belief $b(\mathbf{x}_m)$ as the factorization all the incoming messages and the *a priori* PDF $p(\mathbf{x}_m)$:

$$b(\mathbf{x}_m) \propto p(\mathbf{x}_m) \prod_{a\in\mathcal{A}_m} \mu_{a\to m}(\mathbf{x}_m) \times \prod_{n\in\mathcal{M}_m} \mu_{n\to m}(\mathbf{x}_m). \tag{107}$$

where $\mu_{a\to m}(\mathbf{x}_m)$ and $\mu_{n\to m}(\mathbf{x}_m)$ are calculated by using (101) and (103), respectively. After that, the estimated position is calculated as the average value of the belief distribution while the estimated covariance matrix $\mathbf{P}_m$ calculated by using the set of particles as reported in [130]. Therefore, the belief is approximated with a Gaussian distribution and the related parameters, *i.e.*, the mean and the trace of $\mathbf{P}_m$, are broadcast to neighbors.

Outgoing Messages

The outgoing message is simply proportional to the belief dividing the incoming message from a specific factor node.

Messages to Anchor: The message from mobile to anchor node is

$$\mu_{m\to a}(\mathbf{x}_m) \propto \frac{b(\mathbf{x}_m)}{\mu_{a\to m}(\mathbf{x}_m)}, \tag{108}$$

The state probability is defined as the integration of multiplication of likelihood and message from the mobile:

$$\tilde{P}(s_{a\to m}) = \int p(\tilde{r}_{a\to m}|\mathbf{x}_m, \mathbf{x}_a, s_{a\to m})\, \mu_{m\to a}(\mathbf{x}_m)\, d\mathbf{x}_m, \tag{109}$$

By applying the approximation that $b(\mathbf{x}_m)$ is a delta function, the previous equation can be simplified as:

$$\tilde{P}(s_{a\to m}) \approx \frac{p(\tilde{r}_{a\to m}|\hat{\mathbf{x}}_m, \mathbf{x}_a, s_{a\to m})}{\mu_{a\to m}(\hat{\mathbf{x}}_m)}, \tag{110}$$

Since the probability of one range measurement should be normalized, the link state probability can be furthermore simplified as

$$\begin{aligned} P(s_{a\to m}) &= \frac{\tilde{P}(s_{a\to m})}{\sum_{i=0}^{1} \tilde{P}(s_{a\to m} = i)} \\ &= \frac{p(\tilde{r}_{a\to m}|\hat{\mathbf{x}}_m, \mathbf{x}_a, s_{a\to m})}{\sum_{i=0}^{1} p(\tilde{r}_{a\to m}|\hat{\mathbf{x}}_m, \mathbf{x}_a, s_{a\to m} = i)}. \end{aligned} \tag{111}$$

Note that the incoming message $\mu_{a\to m}(\hat{\mathbf{x}}_m)$ is eliminated in the above normalization. Hence, the calculation of outgoing messages to anchors can be avoided, that is, it is not necessary to compute equation (108).

Messages to Mobile: The outgoing message to mobile $\mu_{m\to n}$ is the similar to the one to anchor, but it can be canceled out when calculating the state probability. Therefore, it is not calculated in the implementation of the algorithm. Similarly, the link state probability is given by

$$P(s_{n\to m}) = \frac{p(\tilde{r}_{n\to m}|\hat{\mathbf{x}}_m, \hat{\mathbf{x}}_n, s_{n\to m})}{\sum_{i=0}^{1} p(\tilde{r}_{n\to m}|\hat{\mathbf{x}}_m, \hat{\mathbf{x}}_n, s_{n\to m} = i)}. \tag{112}$$

Finally, hard decision is made when the algorithm converges. For a given range measurement, if $P(s_{n\to m} = 1)$ is larger than 0.5, the link is assumed in NLoS state, otherwise in LoS state.

Given the mobile belief approximation, the computational complexity and network traffic can be greatly reduced, making the proposed algorithm suitable to be implemented in mobile devices with low computational capability.

Conclusion

This paragraph proposed a cooperative NLoS identification and positioning algorithm. The proposed algorithm is fully distributed with low complexity and low network traffic and does not require prior information of NLoS state. The proposed algorithm was able to detect NLoS range measurements and improved positioning accuracy in NLoS conditions.

3.2.3.4 Hybrid Cooperative Localization Algorithms

Nowadays, Global Navigation Satellite Systems (GNSS) are widely employed in navigation, mapping, environment protection, etc. However, there are still some limitations of GNSS-only localization systems. In some scenarios, for example, urban canyons, dense foliage and indoors, the GNSS receiver usually fail to see enough satellites, due to the blockage of the GNSS signals.

Based on the contextual fusion and acquisition defined in [85], hybrid cooperative positioning approaches are proposed in order to cope with these limitations. The focus of this subsection is to enhance the concept of hybrid cooperative positioning [131], where a set of nodes in a mesh network adopt cooperative Peer-to-Peer (P2P) aiding approach between receivers, with the aim to increase positioning accuracy and availability. The paradigm of this P2P cooperative localization relies on the existence of direct communication links among the nodes (GNSS receivers) in the P2P network, where each node voluntarily shares its positioning information to enhance the positioning performance of the network. Consider the simple example scenario depicted in Fig. 50, neither of the two unknown devices can see four satellites to estimate their positions, however, by performing ranging between them and exchanging their positioning information, an unambiguous position fix could be obtained.

From a probabilistic graphical model point of view, the hybrid and cooperative approach can be viewed as a mapping of the factorization onto a Bayesian network [132], like the one depicted in Fig. 51 whose topology matches the wireless connectivity in the network of receivers. Nodes in the graph represent the state vectors of each receiver and edges link their conditional dependencies due to the ranging between neighbor nodes or the information known at previous timestep and propagated to the current one. In particular,

- $f_m(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_m^{(k-1)}) \equiv p(\boldsymbol{x}_m^{(k)}|\boldsymbol{x}_m^{(k-1)})$ represents mobility.
- $g_{s,m}(\boldsymbol{x}_m^{(k)}) \equiv p(d_{s\to m}^{(k)}|\boldsymbol{x}_m^{(k)})$ represents the likelihood of pseudorange difference $s \in \mathcal{S}_m^{(k)} \backslash S_m$, given the state of node m.
- $e_{a,m}(\boldsymbol{x}_m^{(k)}) \equiv p(r_{a\to m}^{(k)}|\boldsymbol{x}_m^{(k)})$ represents the likelihood of range measurements from anchor $a \in \mathcal{A}_m^{(k)}$, given the state of node m.

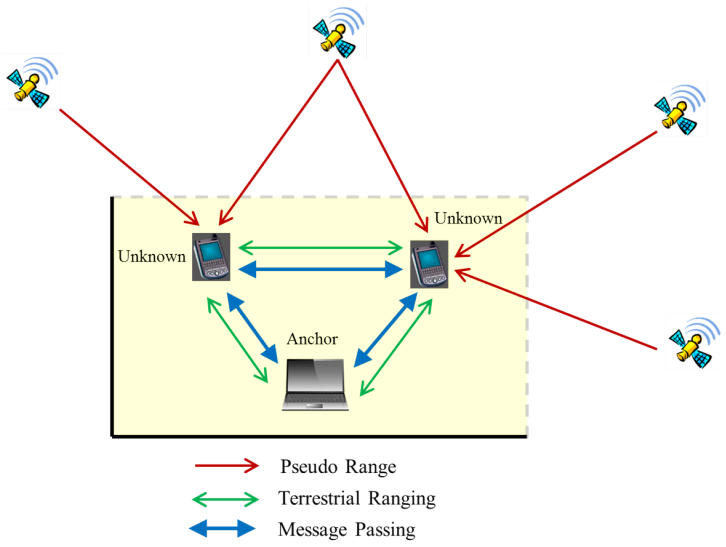


Figure 50: Hybrid-cooperative paradigm: scarce GNSS availability (red arrows) could be coped with P2P terrestrial ranging (green arrows) and information exchange (blue arrows). This rectangular shows a typical office environment: the dark lines represent walls that block GNSS signals, and the dash lines represent the windows that allows GNSS signals to pass through.

- $h_{n,m}(\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_n^{(k)}) \equiv p(r_{n\to m}^{(k)}|\boldsymbol{x}_m^{(k)}, \boldsymbol{x}_n^{(k)})$ represents the likelihood of range measurement from mobile neighbor $n \in \mathcal{M}_m^{(k)}$, given the positions of nodes m and n.

The marginals are computed by each node taking into account all the incoming messages and current measurements. State distribution estimation (belief update) and exchange (message passing) can be performed by applying loopy BP on the graph.

Numerically, these operations can be performed by particle filtering. Particle Filters (PFs) are a category of Monte Carlo (MC) methods [133] that approximate the probability density function of a generic state vector $\boldsymbol{x}$ by an *empirical distribution* given a set of N particles and normalized weights $\{\boldsymbol{\chi}_i, w_i\}$, such that $\sum_i^N w_i = 1$ and

$$p(\boldsymbol{x}) \approx \sum_{i=1}^{N} w_i \delta(\boldsymbol{x} - \boldsymbol{\chi}_i), \tag{113}$$

where $\delta(\cdot)$ the a Dirac-delta function ($\delta(0) = +\infty$ zero elsewhere).

Major details on PF fundamentals and some of its several variants, as well as a discussion on theoretical and practical issues can be found on [134] and [135]. The following subsections will describe the three main operations (prediction, update and cooperation) performed by each node in order to run the PF-based Sum-Product Algorithm (SPA) in a distributed fashion. The proposed PFs fuse different types of range measurements while enables cooperation among peers, and are extensions of Hybrid-Cooperative Particle Filter (HC-PF) presented in [130].

Prediction

Given the outcome of the previous time step inference process $p(\boldsymbol{x}_m^{(k-1)}|_{\mathbb{Z}_{\mathcal{M}}^{(1:k-1)}})$ (or the initial distribution $p(\boldsymbol{x}_m^{(0)})$ if it is the first step), prediction (downward message from previous to current time step nodes in the graph) is given by

$$p\left(\boldsymbol{x}_m^{(k)}\middle|\mathbb{z}_{\mathcal{M}}^{(1:k-1)}\right) = \int p\left(\boldsymbol{x}_m^{(k)}\middle|\boldsymbol{x}_m^{(k-1)}\right) \cdot p\left(\boldsymbol{x}_m^{(k-1)}\middle|\mathbb{z}_{\mathcal{M}}^{(1:k-1)}\right) d\boldsymbol{x}_m^{(k-1)}, \tag{114}$$

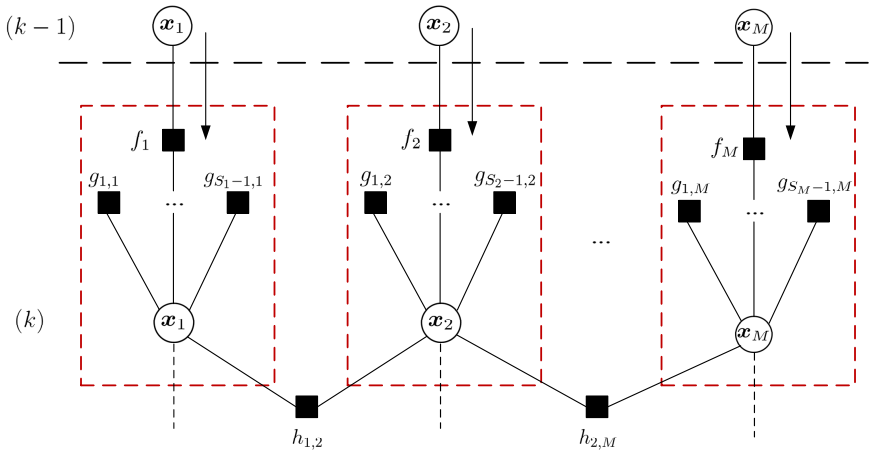


Figure 51: Bayesian network graph for the hybrid-cooperative positioning. Upper variable nodes contain the state vector estimates (beliefs) at previous time step $k-1$, while central variable nodes contain the state vectors at current time step k. Downward edges carries temporal messages or predictions, while internode edges exist when receivers perform ranging and exchange their position estimates (message passing).

and approximated with importance sampling, by drawing N state samples $\boldsymbol{\chi}_{i,m}^{(k-1)}$

$$\boldsymbol{\chi}_{i,m}^{(k-1)} \sim p\left(\boldsymbol{x}_m^{(k-1)} \middle| \mathbb{Z}_{\mathcal{M}}^{(1:k-1)}\right), \tag{115}$$

which will be propagated with a *state transition function* $f(\cdot)$ (either a Position-Time (PT), Position-Velocity-Time (PVT) or any suitable mobility model)

$$\boldsymbol{\chi}_{i,m}^{(k)} = f\left(\boldsymbol{\chi}_{i,m}^{(k-1)}, \boldsymbol{\omega}_{i,m}^{(k)}\right). \tag{116}$$

where $\boldsymbol{\omega}_{i,m}^{(k)} \sim \mathcal{N}\left(0, \mathbf{Q}_m^{(k)}\right)$ is the *process noise vector*, which takes into account non-linearities and perturbations on the system between time steps $k-1$ and k. It is chosen according to the dynamics of the system, typically modeled by a vector of random noise (not necessarily stationary) normally distributed with zero mean and covariance matrix $\mathbf{Q}_m^{(k)}$. In the prediction, all particles have the same weight, that is, $w_{i,m}^{(k|k-1)} = \frac{1}{N}$.

Update

Having computed the prediction $p(\boldsymbol{x}_m^{(k)}|\mathbb{Z}_{\mathcal{M}}^{(1:k-1)})$, updated beliefs are proportional to the product of the likelihood of local measurements $\boldsymbol{z}_m^{(k)}$ and the predicted state

$$p\left(\boldsymbol{x}_m^{(k)} \middle| \mathbb{Z}_{\mathcal{M}}^{(1:k)}\right) \propto p\left(\boldsymbol{z}_m^{(k)} \middle| \boldsymbol{x}_m^{(k)}\right) p\left(\boldsymbol{x}_m^{(k)} \middle| \mathbb{Z}_{\mathcal{M}}^{(1:k-1)}\right), \tag{117}$$

which is again approximated with importance sampling, by updating the weights of the propagated particles $\boldsymbol{\chi}_{i,m}^{(k)}$ with

$$w_{i,m}^{(k)} \propto w_{i,m}^{(k|k-1)} \cdot p\left(\boldsymbol{z}_m^{(k)} \middle| \boldsymbol{\chi}_{i,m}^{(k)}\right). \tag{118}$$

Since measurements are independent, the likelihood can be expressed as the product of PDF associated to each single measurement

$$p\left(\boldsymbol{z}_m^{(k)}\middle|\boldsymbol{\chi}_{i,m}^{(k)}\right) = \prod_{n\in\mathcal{M}_m^{(k)}} p_{n\to m}\left(r_{n\to m}^{(k)} - \left\|\boldsymbol{p}_n^{(k)} - \boldsymbol{\pi}_{i,m}^{(k)}\right\|\right)\cdot \prod_{a\in\mathcal{A}_m^{(k)}} p_{a\to m}\left(r_{a\to m}^{(k)} - \left\|\boldsymbol{p}_a^{(k)} - \boldsymbol{\pi}_{i,m}^{(k)}\right\|\right)\cdot \prod_{s\in\mathcal{S}_m^{(k)}\setminus S_m} p_{s\to m}\left(d_{s\to m}^{(k)} - \left(\left\|\boldsymbol{p}_s^{(k)} - \boldsymbol{\pi}_{i,m}^{(k)}\right\| - \left\|\boldsymbol{p}_{S_m}^{(k)} - \boldsymbol{\pi}_{i,m}^{(k)}\right\|\right)\right), \quad (119)$$

where $\boldsymbol{\pi}_{i,m}^{(k)}$ is the position component of particle $\boldsymbol{\chi}_{i,m}^{(k)}$, in this case $\boldsymbol{\pi}_{i,m}^{(k)} \equiv \boldsymbol{\chi}_{i,m}^{(k)}$. $p_{a\to m}$ is the PDF of the terrestrial range error $\nu_{a\to m}$ while $p_{n\to m}$ is the PDF of the error $\nu_{n\to m}$. $p_{s\to m}$ is the PDF of the pseudorange difference error $\nu_{s\to m}$. Since all the errors are assumed to be Gaussian distributed $\mathcal{N}\left(\mu,\sigma^2\right)$, these contributions are calculated using the well known normal PDF:

$$f_{\mathcal{N}}(x;\,\mu,\sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}}\, e^{-\frac{(x-\mu)^2}{2\sigma^2}}. \quad (120)$$

Cooperation

Given the prediction (time message), pseudorange and range measurements (assuming to have received the cooperation messages from neighbor nodes), the goal now is to compute the cooperation messages to be passed to neighbors.

In order to properly use range measurements (coming from mobile neighbors) in the update phase, the algorithm should take into account not only the uncertainties related to the range measurements but also the ones corresponding to the estimated positions. Denoting by $p(\boldsymbol{p}_n^{(k)})$ the PDF of the position of peer n ($n \in \mathcal{M}_m$), the likelihood expression (119) should be calculated as follows:

$$p\left(\boldsymbol{z}_m^{(k)}\middle|\boldsymbol{\chi}_{i,m}^{(k)}\right) = \int p\left(\boldsymbol{z}_m^{(k)}\middle|\boldsymbol{\chi}_{i,m}^{(k)}, \boldsymbol{p}_{n_1}^{(k)}, \ldots, \boldsymbol{p}_{n_l}^{(k)}\right) p\left(\boldsymbol{p}_{n_1}^{(k)}, \ldots, \boldsymbol{p}_{n_l}^{(k)}\right) d\boldsymbol{p}_{n_1}^{(k)} \ldots d\boldsymbol{p}_{n_l}^{(k)}, \quad (121)$$

where $n_1, \ldots, n_l \in \mathcal{M}_m$ are all mobile neighbors of the node m. In principle, this likelihood can be calculated by approximation of particle filters.

Conclusion

This paragraph presented a novel hybrid and cooperative positioning paradigm, distributed particle filtering. Based on Bayesian inference theory and probabilistic graphical models, we proposed a novel hybrid cooperative particle filters, fusing information from satellites (pseudoranges) and terrestrial wireless systems.

3.2.4 Semantic Localization

Location information is one of the key drivers for many BUTLER use cases. The objective was to to enrich the geo-localization (x,y co-ordinates) contextual information generated by other partners with semantic annotations and enable rich spatial reasoning. The key features of the semantic reasoner are,

- Ability to describe the spatial characteristics of different locations in BUTLER environment, and represent sensors and their locations using W3C SSN ontology.
- Translate location information represented in coordinates into semantic locations (e.g. Home, Living_Room, etc.) and semantically link locations with activities (e.g., Watching_TV with Living_Room).

```
curl -X POST -H 'Content-type: application/json'
  --data '{ "rule": "select x,y,location(x,y) from LocationEvent.win:time(30 sec)" }'
  https://butler.cs.kuleuven.be:8443/samurai/rest/esper/statements/semlocation
```

Figure 52: ESPER operator that accepts geo-location and returns semantic location

- Provision for modelling the Quality-of-Context (QoC) parameters for location information and considering it for semantic reasoning.

Semantic enrichment of events is supported through custom operators in Esper that leverage background knowledge stored in a semantic database. For semantic and spatio-temporal reasoning, the system uses a GeoSPARQL enabled storage backend (such as Parliament or Strabon). The benefits of such a building block are manifold:

- Describe the spatial characteristics of different locations in your environment
- Use the W3C SSN ontology to describe the sensors and their position
- Translate positions in coordinates into semantic locations (e.g. Bedroom)
- Semantically link locations with activities (e.g. Sleeping in a Bedroom)

The following example illustrated in Figure 52 shows how to add a new operator called location(x,y) that accepts to coordinates, and returns the symbolic name of that location:

The command shown in Figure 52 would not only match the x and y coordinates, but also its corresponding symbolic or semantic name (e.g. kitchen). The mapping of coordinates to symbolic names (i.e. rooms) is stored in a semantic database and the location(x,y) operator calls this semantic database to infer and retrieve the corresponding name. The semantic database holds a catalogue of rooms, where each room is represented as a polygon. For each room, we also include a classification of relevant activities. This can then later be used to detect activities of interest or rule out the occurrence of other ones. The integration is done under the hood. Every custom Esper operator has to be translated into the corresponding SPARQL query, i.e. the language commonly understood by most semantic reasoning engines and databases. For spatio-temporal reasoning, we make use of the GeoSPARQL standard. Here is a small example of how this is achieved shown in Figure 53:

In summary, the mapping of coordinates to symbolic names (i.e. rooms) is stored in a semantic database (where each room is represented as a polygon) and an ESPER operator calls this semantic database to infer and retrieve the corresponding name. Every custom Esper operator is translated into the corresponding SPARQL query, the language commonly used by most semantic reasoning engines. Moreover, the semantic reasoner goes beyond the existing systems in modelling the uncertainties of the geo-location estimators. The traditional point representation (x,y co-ordinates) is replaced by (x + error, y + error) to represent the position of a user. In order to model and reason about semantic locations of realistic environments such as home, office, we have modelled the walls thicker (say 0.2 m) and replaced the scheme that tests for a point (individual) within a polygon (room) to one that tests for a polygon (individual with inaccuracies) within a polygon (room), rather than changing the polygons for all the rooms. The location of the individual is not represented as a point with x,y coordinates, but as a square or rectangle box with side lengths of 0.4 m, with the x,y coordinates as the centre. Rather than hard-coding the "inaccuracy" to pre-defined values such as 0.2m, we let the localization system to specify the maximum error along different axes to dynamically define the box of the individual (say a value of 0.2 would mean a square box of 40 cm by 40 cm). Thus, the reasoner provides options to return all possible rooms where the user can be present when the resulting polygon covers more than 1 room. The developed system (semantic enrichment block) is part of the BUTLER user behaviour engine described in Section 3.3.2.1 and is intended to be used in the some of the upcoming BUTLER field trials. For example,

```
PREFIX afn: <http://jena.hpl.hp.com/ARQ/function#>
PREFIX fn: <http://www.w3.org/2005/xpath-functions#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX gml: <http://www.opengis.net/ont/gml#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX par: <http://parliament.semwebcentral.org/parliament#>
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX sf: <http://www.opengis.net/ont/sf#>
PREFIX time: <http://www.w3.org/2006/time#>
PREFIX units: <http://www.opengis.net/def/uom/OGC/1.0/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX ex: <http://localhost/samurai/>

INSERT DATA INTO <http://localhost/samurai#Location>
{ex:Room              a                owl:Class;
                      rdfs:subClassOf  geo:Feature.

ex:LivingRoom         a                ex:Room;
                      rdfs:label       "Living Room";
                      geo:hasGeometry  ex:GeoLivingRoom.

ex:GeoLivingRoom      a                sf:Polygon;
                      geo:asWKT        "POLYGON ((290 600, 580 600, 580 700, 900 700,
                                        900 260, 290 260, 290 600))"^^sf:wktLiteral .

ex:MasterBedroom      a                ex:Room;
                      rdfs:label       "Master Bedroom";
                      geo:hasGeometry  ex:GeoMasterBedroom .

ex:GeoMasterBedroom   a                sf:Polygon ;
                      geo:asWKT        "POLYGON ((290 880, 580 880, 580 600, 290 600, 290 880))"^^sf:wktLiteral .

ex:activity           a                owl:DatatypeProperty;
                      rdfs:domain      ex:Room;
                      rdfs:range       xsd:string

ex:LivingRoom         ex:activity      "Watch TV"
ex:LivingRoom         ex:activity      "Play Game"
ex:MasterBedroom      ex:activity      "Sleep"
}
```

Figure 53: Spatio-temporal reasoning with GeoSPARQL standards

in the SmartHealth use case the semantic reasoner will be leveraged to reduce the number of false positives in estimating the user fall detection.

3.3 Behavior Modelling and Synthesis

As detailed in the deliverable D2.3 [136], the goal of this task is to recognize the user contexts both effectively and efficiently without depending extensively on input from the users. In accordance with the road map envisioned in the deliverable D2.3 [136], we have developed/enhanced algorithms and tools in the scope of the horizontal architecture of BUTLER. One of the primary contributions is in advancing the integration between deterministic and probabilistic modelling of human behaviours. On the one hand, the behavioural SmartServer (SAMURAI) seamlessly integrates various technologies such as semantic reasoning, Complex Event Processing (CEP), stream mining, etc., as building blocks in a scalable way. On the other hand, various algorithms that capitalize on correlations between user contexts and causal information enable versatile and robust recognition of user contexts. Multi-modal user behaviour modelling and recognition is enhanced by indirect inference from semantic locations and improved electrical appliance usage algorithms. As behaviour recognition systems are complex with lot of parameters, our methodology to learn deployment trade-offs facilitate the efficient deployment of software components in different BUTLER platforms. Other works in the domain of transfer learning aims at reducing user efforts in training as well as aids demographic analysis of user behaviours. Also, the work on contextual networking pays more attention to mobility support for masses by providing context-aware adaptation of various networking mechanisms. Moreover, from the user perspective context synthesis and management approach proposed on the basis of CEP would enable them to define their own contexts of their interest even without much technical skills. The sections below give a detailed overview of the tech-

nical achievements with extensive citations to relevant scientific publications and level of integration of those scientific contributions in BUTLER platform.

3.3.1 Algorithms and Techniques for Advanced User Context Recognition

3.3.1.1 Direct and Indirect Context Recognition Using HARD-BN

The modern ubiquitous computing environments is characterized by the requirement to sense, interpret and anticipate multiple user contexts (both simple and complex contexts) simultaneously often with the help of resource constrained devices. Moreover, given the mobile capabilities of the user and their devices, the operating conditions are continuously changing for the ubiquitous applications giving rise to other non-trivial challenges such as sensor ambiguities (e.g., sensor failures/availability and missing data).

Traditionally, context-aware applications have modelled user contexts such as activities as high-level contexts which are either inferred directly from low-level sensors or indirectly through other context informations such as location [136]. A major drawback of these systems is that they either support direct or indirect sensing of human activities (or other high-level contexts) but not both. This would severely limit the ability of a robust context-aware system to gather information from as many sources as possible. Another major drawback of the above mentioned works is their inability to explicitly handle missing data.

Hence, we propose a graphical Bayesian framework [137] which would help in the horizontalization of smart applications from multiple vertical domains the following four key features: ability to combine both simple and high-level contexts (with a heterarchical structure), robustness against partial observability and missing data during inference (using recursive inference algorithms), flexibility to add new contexts and discover its relationship with existing contexts (support for distributed topology), and reduction in training period by supporting autonomic and incremental learning.

The guiding principle behind the framework is generating informative priors based on the correlations between high-level contexts (e.g., user locations and user activities) which enable selective contextual fusion of auxiliary information as and when required. We loosely define the informative priors as the priors for user contexts which are estimated from the posteriors of other context nodes i.e., replacing uniform discrete distribution by a beta distribution with parameters estimated from other correlated contexts. For instance, physical activities of a user (a high-level user context) inferred from accelerometer data (low-level context) are modelled as a Bayesian network whose priors are generated from the user location (a correlated high-level context), i.e., when the user is indoors, the probability of running is close to zero. The correlated contexts used for prior generation are specified either manually or discovered automatically with correlation mining algorithms based on Kullback-Leibler (KL) divergence [138].

3.3.1.1.1 Incremental learning and Self-adaptation

Learning or training is an important step in any supervised machine learning algorithms and most classical algorithms assume that sufficient training data are available prior to classification. The characteristics of the modern environments does not guarantee this assumption for the context-aware applications. Owing to this, there is a need to design machine learning algorithms to learn from new training examples even after deployment of the models and to add/remove a new context or sensors at run-time. For instance, in HARD-BN the likelihood values (measurement model) of the context souces are incremented with each new instance and the updated values are used to predict the class value probabilistically for the new instance, i.e.:

$$P(Z_j|X_{ik}),\ \forall\, i = 1, ..., n;\ k = 1, ..., m\ and\ j = 1, ..., p$$

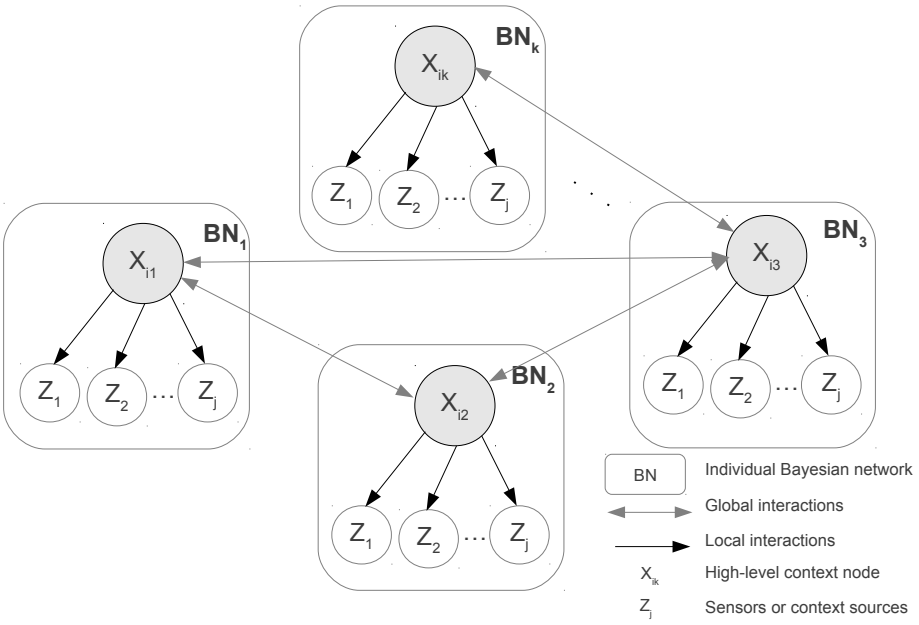


Figure 54: A conceptual overview of HARD-BN [137]

where Z_j is the observation of the j^{th} sensor source information conditioned on the i^{th} context value in k^{th} Bayesian network as shown in Figure. 54. In our running example, the likelihood values of energy in the accelerometer readings (Fast Fourier Transform - fft values) corresponding to each class value of the *User's Physical Activity* context node will be incremented according to the new example. Similarly the measurement model of the individual Bayesian networks that is incremented when observing the other high-level context nodes i.e.,

$$P(X_{ik}|X_{ab}),\ \forall\ i, a = 1, ..., n;\ k, b = 1, ..., m\ and\ ik \neq ab$$

where Z_{ik} is the observation of the i^{th} context value in k^{th} Bayesian network conditioned on the a^{th} context value in b^{th} Bayesian network with a constrain that k^{th} Bayesian network and b^{th} Bayesian network are not the same network. In our running example, a distribution over semantic locations corresponding to each context value of the *User's Physical Activity* node is learnt incrementally at this step. As the learning in HARD-BN involves estimating the likelihoods described above in the individual BNs, they can be easily parallelized as concurrent tasks to reduce the overall training time. As a result, HARD-BN can accommodate the context drifts in long-term such as learning new Device types for User Location2 context node and new WiFi ssid for the User Location1 node in Figure. 55.

In addition to incremental parameter learning, HARD-BN also supports self-adaptation of the network structure in response to addition or deletion of context sources by modifying the inference algorithm at run-time. If any particular context value is not available, then its likelihood value is omitted while calculating the posterior. For instance, while calculating the posterior of User's physical activity based on other high-level contexts in Figure. 55, if the User Location3 context node (localisation based on ambient sensors) is not available any more because user switched to another smart phone which does not support those sensors, then the likelihood of that node will be omitted while calculating the posterior. For addition of new context nodes, the corresponding likelihood value is included after evaluating the stability of the node's prediction on the last *n* data.

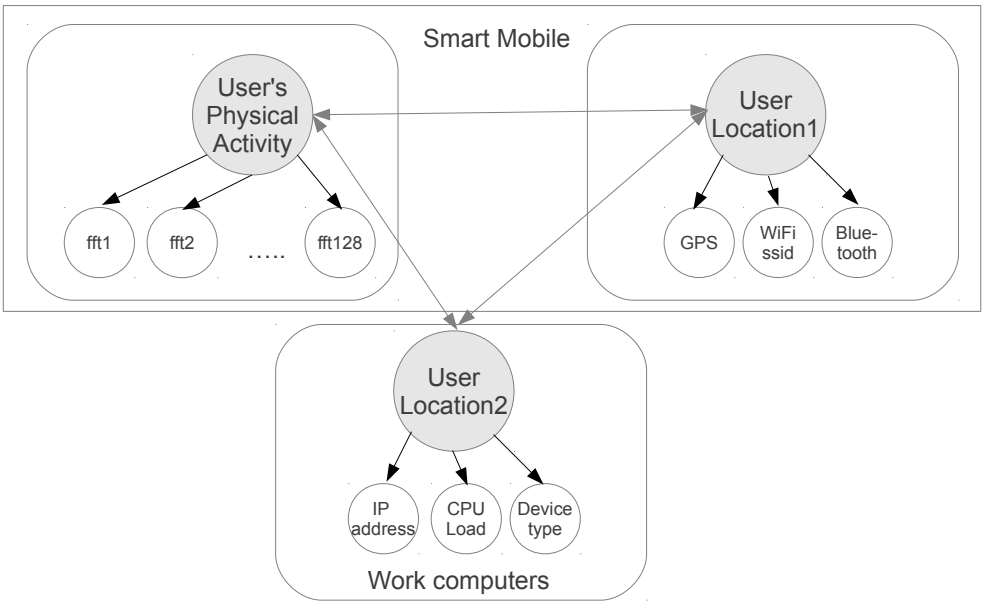


Figure 55: HARD-BN framework modelling the contexts for Personal Assistant use case [137]

HARD-BN leverage multi-view learning not only to improve the prediction performance but also to introduce robustness in ubiquitous application by creating redundant views of different context information without compensating much on performance.

Inference in HARD-BN is done in bootstrap mode by recursively executing two steps to combine the global and local views of the individual Bayesian networks. First, a global view on the estimate of each high-level context nodes is generated by combining objective prior probabilities (uniform distribution) and evidences from other high-level contexts. Later, in order to generate a local view from dedicated sensors, the posterior estimated from the global view is used as an informed prior to determine the most probable value for each of the high-level contexts.

3.3.1.1.2 Generating a global view on context nodes

The advantages of the informed prior is well known in the literature, but often paid less attention to because of the practical difficulties in acquiring them for highly dynamic systems. The objective of this inference step is to utilize the global influence of a context on other co-related context information to update the uniform prior distribution for improved prediction.

$$P'(X_{ik}) = P"(X_{ik}) \prod P(X_{ab}|X_{ik}),$$

$\forall\ i, a = 1, ..., n;\ k, b = 1, ..., m\ and\ ik \neq ab$

In our running example, an estimate of the current physical activities of the user is inferred from the location information. For example, if the location of the user is outdoors, then the probability of him/her being active is higher.

3.3.1.1.3 Generating a local view in individual Bayesian networks

This step acts as a correction step where the estimated context values from the previous step are adjusted according to the evidence from the local observations.

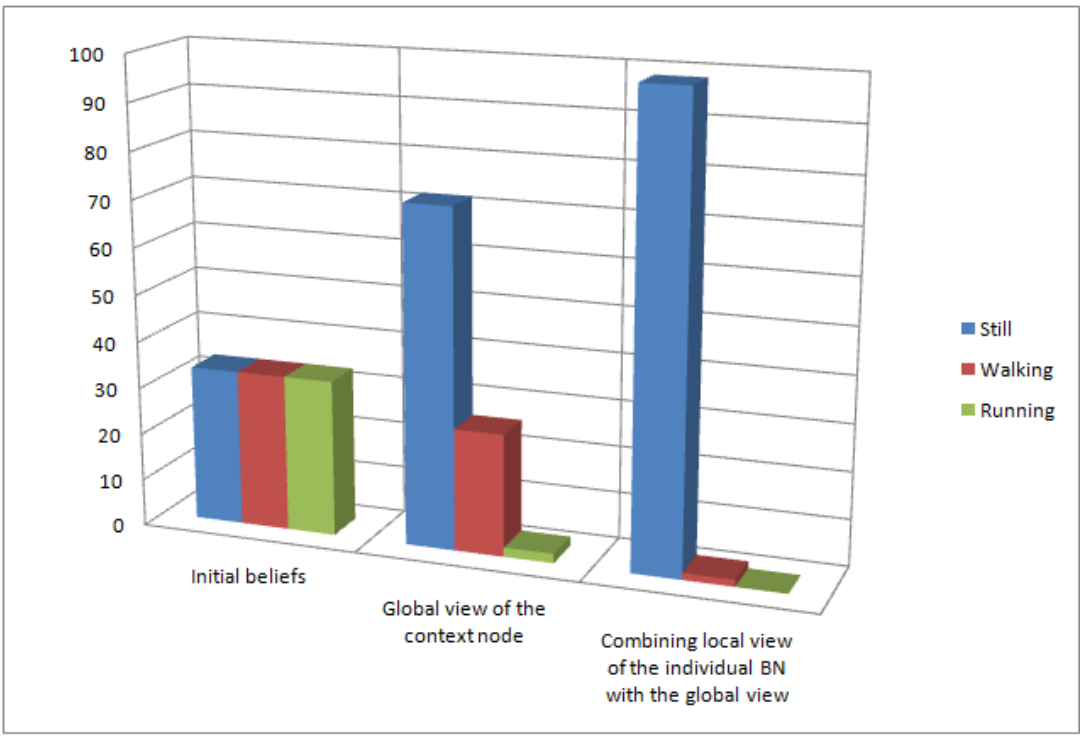


Figure 56: Inference in HARD-BN for *User's Physical Activity* node in the running example

$$P(X_{ik}) = \arg\max_{i} P'(X_{ik}) \prod P(Z_j|X_{ik}),$$
$$\forall\ i = 1, .., n;\ k = 1, .., m\ j = 1, .., p\ and\ ik \neq ab$$

In this final step of inference, the informed prior is combined with the local evidence from dedicated sensors (i.e, features from the accelerometer) to mitigate the influence of errors from the other high-level contexts (i.e., location node).

Figure. 56 illustrates the influence of different inference steps on the probability distribution of *User's Physical Activity* node of the running example. Initially, all the class values *staying still, walking, running* had same probabilities. At the end of the first step, contextual information from other high-level context nodes (i.e., semantic location of the user is office) is used by the framework to modify the distribution favouring *staying still and walking* over *running*. In the next step, the accelerometer readings are combined with the informative priors to arrive at the final estimate of the probability distribution for the *User's Physical Activity* node.

To summarize, the term $P"(X_{ik})$ is the objective prior for the context variables with equal probability distribution for the possible values of a context node and $P'(X_{ik})$ is the informed prior obtained from the global view of the framework. Note that the likelihood estimates used while generating the global view is the latest likelihood estimates of the other high-level context nodes available from previous iteration or time step. This is understood to provide a good approximation of temporal dependencies of the context nodes on its own value in the previous time step. Furthermore, in general, the local view can be related to causality whereas global view estimates the co-relations among various loosely coupled high-level contexts.

Another major objective of combining multiple views of context in HARD-BN is to create robustness for contexts to missing values. Most of the existing works handle the missing data issue at prediction by imputation of raw data through various statistical methods, imputation decision trees, k-means methods, etc. In this section, HARD-BN imputes the data at the classifier level where the global view

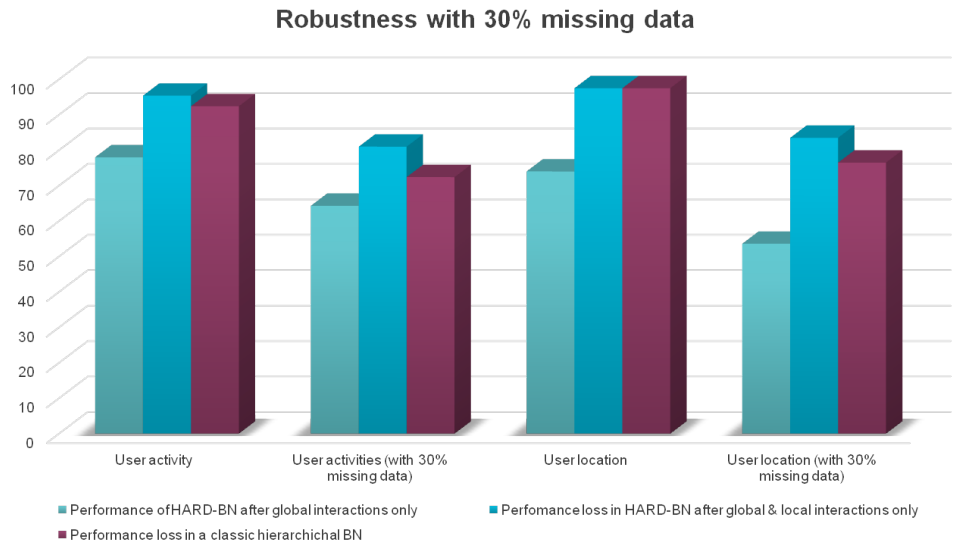


Figure 57: Robustness HARD-BN for missing values

generates an estimate of the possible context values of the BN (based on the other co-Bayesian networks) for which the low-level sensor data are missing. Figure. 57 illustrates the robustness resulting from multi-view learning in HARD-BN in the presence of 30% missing data.

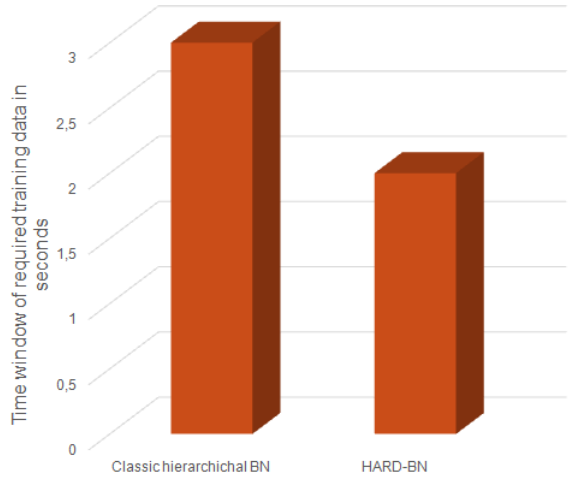
Also note that under the minimal performance requirement criteria in HARD-BN, combining global and local views show improvement in realizing non-functional requirements such as response time and network traffic. Figure. 58 shows the reduction in time for context sensing due to reduction in window size of accelerometer data for predicting user physical activity and amount of data required for predicting with the required minimum confidence compared to a typical hierarchical Bayesian network.

The proposed algorithms are integrated in SAMURAI - Behavior SmartServer.

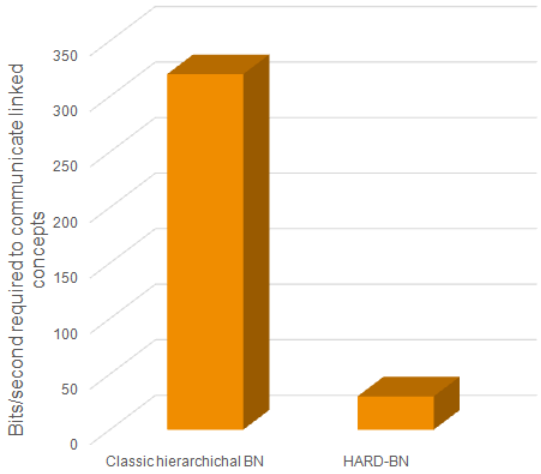
3.3.1.2 Learning Deployment and Re-configuration Tradeoff for Context-aware Applications

Smart homes and offices, smart health, assisted living, smart cities and transportation are only a few examples of possible application scenarios where IoT is playing a vital role where achieving self-* properties is one of the foremost challenges. For example, one challenge on self-optimization is how to change the behavior of a system to achieve a desired functionality, while maintaining a balance with Quality of Service (QoS) and resource usage. Self-optimization in the Internet of Things (compared to traditional software systems) shifts the focus from design and deployment of a single or a few elements operating autonomously to a large complex ecosystem of a network of autonomous elements.

The main objective of our work is to find optimal distributed deployments and configurations of application components. We use annotated component graphs to model application compositions and Pareto-curves to represent the optimization options for each (type of) platform, i.e. the Smart Object, Smart Mobile and Smart Server. The resource optimization objectives are chosen with respect to the QoS requirements and the tradeoffs on the computation vs. communication cost-



(a) Reduction in window-length of accelerometer data to predict the physical activity of the user



(b) Reduction in over all network traffic while prediction

Figure 58: Reduction in data required for prediction due to multi-view learning

benefits. For the runtime (re)configuration and (re)deployment, we use Markov Decision Processes and dynamic decision process to achieve the self-optimization objectives of the system.

In this work, we challenge the hypothesis that using the cloud (i.e, SmartServer) for all data storage and processing will always provide resource and performance benefits. We explore examples where the decision of deploying an application (or some of its subcomponents) on either the sensor (i.e., SmartObject), the mobile (i.e., SmartMobile) or in the cloud (i.e, SmartServer) is not clear-cut. In our previous work [139, 140] we identified that many resource and performance trade-offs exist, and we demonstrate that a modular application design philosophy helps to support optimal mobile cloud application deployments. As shown in Fig. 59, the overall aim is to achieve a distributed intelligence by finding optimal distributed deployments and configurations of application components in the following way:

1. We use annotated component graphs to model application compositions at design time.
2. Pareto-curves are used to represent the optimization options for each (type of) platform. The resource optimization objectives are chosen w.r.t. to the QoS requirements and trade-offs between computation and communication.
3. We use Dynamic Decision Networks for the runtime configuration and deployment to achieve the self-optimization capabilities of the system.

Our experiments show that with this combined approach, our framework is able to learn deployment trade-offs of smart applications for Intelligent Environments and capable of learning from earlier deployment or configuration mistakes to better adapt to the setting at hand.

As the primary objective of the proposed methodology is to assist the developers to better understand the deployment and re-configuration trade-offs, it is not integrated into SAMURAI SmartServer which exposes semantic reasoning and stream mining technologies to the end user applications.

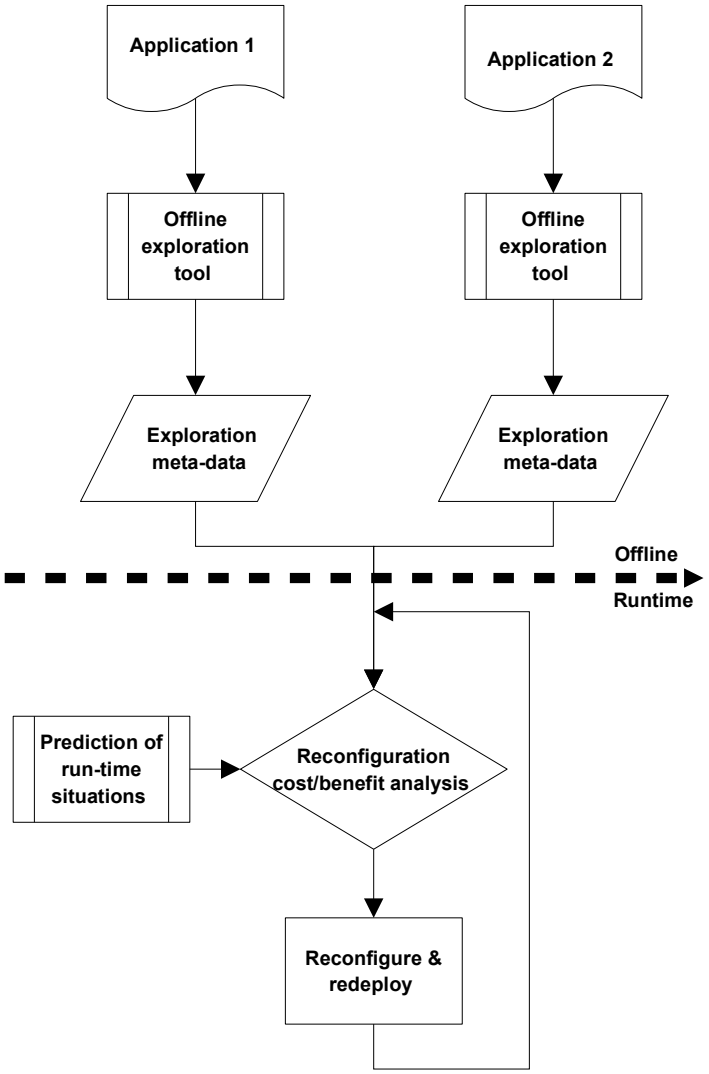


Figure 59: An overview of the approach illustrating the offline and runtime phases

3.3.1.3 Enabling Life-long Learning in Context-aware Applications

With mobile applications tapping into their environments, their operating conditions are continuously changing, giving rise to non-trivial challenges such as sensor ambiguities (e.g., sensor failures and missing data). Therefore, the IoT applications have to cope with heterogeneity, context drifts and continuous changes in operating conditions with minimal input from the end-users necessitating the need for life-long learning principles. We analyse the applicability of highly relevant yet understudied machine learning principles such as multi-view learning, incremental learning and meta-learning that enable life-long learning in context-aware mobile applications [141].

- Incremental learning to realize adaptable and flexible context models to tackle dynamic environments
- Multi-view learning to leverage the inherent heterogeneity of ubiquitous environments
- Meta-learning techniques to capitalize domain/application/user specific knowledge for optimizing the resource consumption of context inference tasks and to detect view disagreements between distributed context models.

The incremental learning enables continuous learning and adaptability in the HARD-BN to cope up with any continuous drifts in the user contexts, provided the newly seen data are annotated. More

advanced tasks such as identifying replacements for a temporarily unavailable primary contexts are still at large as they requires automatic correlation mining between user contexts. Owing to the heterogeneity and complexity in user contexts, we divide the task of correlation mining into two sub-tasks: First, the coarse grained correlations including any many-to-many, many-to-one and one-to-many mappings (i.e., hierarchies) between user contexts is discovered with an adapted apriori frequent set mining algorithm [142]. Later, the degree of similarity between the contexts that exhibit many-to-many relation is measured using a mutual information metric, KL divergence [143].

3.3.1.3.1 Frequent set mining for hierarchy discovery

The choice of using an adapted frequent set mining algorithm is motivated by the fact that these fast algorithms were designed to handle very large databases making them suitable for the modern IoT environments generating large amount of continuous data. In the scope of this work, a frequent item set is a set of context values that occur together between the user context nodes modeled by the HARD-BN framework.

Algorithm 1 explains the proposed apriori algorithm where the join and prune steps are modified (from the original algorithm [142]) to discover the frequent sets containing context values of the context node of interest with any other context node. This will make sure that the obtained frequent sets contain values from two context nodes only. Then, the ratio of the number of context values occurred in the frequent sets for a context node to the total number of values possible for that context node is calculated. Depending on the ratio, the relationship between contexts are categorized into either many-to-many, many-to-one or one-to-many mappings. Note that the high-level context values modeled in HARD-BN are multi-nominal.

Algorithm 1: Adapted apriori frequent set mining algorithm to discover context correlations
Join Step: C_k is generated by joining L_{k-1} with itself under the condition that C_k contains values from only two different context nodes *Prune Step:* Any (k-1) item set that is NOT frequent cannot be a subset of a frequent k-item set and at least one of the items in the set belongs to the new context node that is of interest
C_k: Candidate item set of size k; L_k: frequent item set of size k; L_1 = frequent items; **for**(k = 1; L_k != $\emptyset$; k++) C_{k+1} = candidates generated from L_k; **foreach** (transaction t in database) increment the count of all candidates in C_{k+1} that are contained in t; L_{k+1} = candidates in C_{k+1} with minimum support; **return** Union over index k L_K;

Fig. 60 illustrates the applicability of the proposed algorithm for the use case described in Section 3 with three user contexts - *location1, location3* and *physical activities*. Let *location3* be the context node of interest which is newly introduced in the HARD-BN framework. Now, in iteration 1, frequent sets with only one item are generated from each context node. In iteration 2, the candidate sets are generated by combining frequent sets from two context nodes as described in the Algorithm 1. Then, as home is the only context available in the frequent sets out of four possible context values for the context node *location1* and all the context values of the context node *location3* occurs in the frequent sets, we conclude that a one-to-many relation exist between *location1* and *location3*. Similarly, we can conclude that a many-to-many relation exist between *location3* and physical activities. Such information about hierarchical relations are used by the HARD-BN framework to intelligently

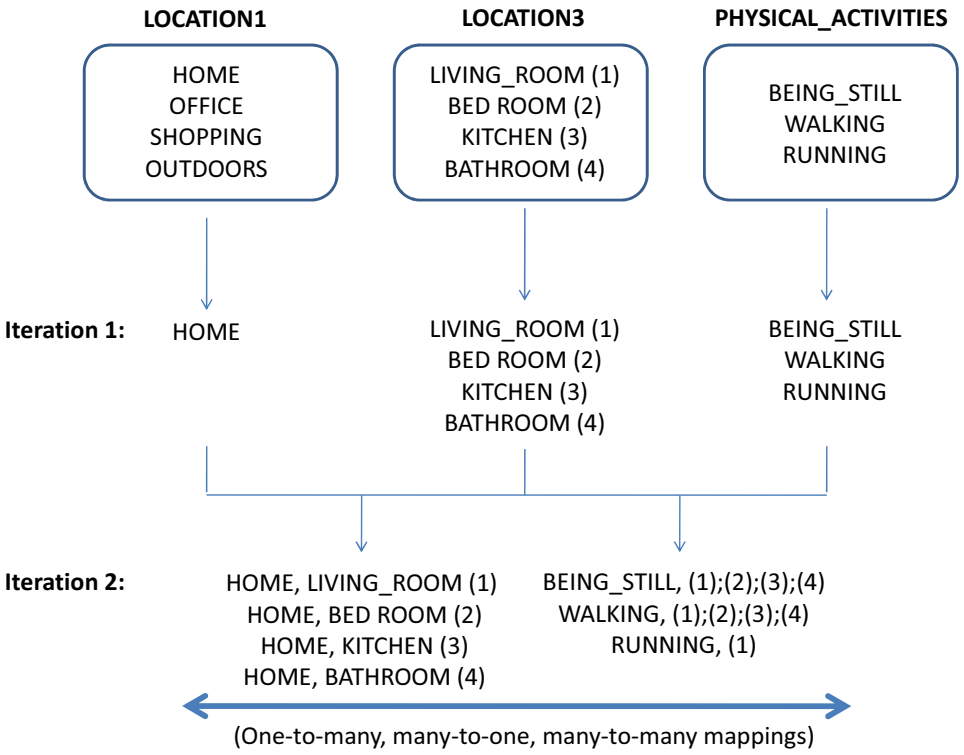


Figure 60: An example of hierarchical clustering of contextual nodes using *Algorithm 1*

switch-on/off high cost (or) resource consuming sensors. For instance, the above discovered hierarchical relation between *location1* and *location3* is utilized to intelligently switch-off the primary context network for *location1* (i.e., GPS as shown in Fig. 54) when the *location3* is active (i.e., the user is at home).

3.3.1.3.2 KL divergence based correlation mining

In the previous subsection, we demonstrated how to discover the possible hierarchical relations between the existing contexts and a newly introduced context. Out of the three types of relations considered, the one-to-many and many-to-one relations can be utilized directly by the smart applications as the correlations are explicit. Whereas many-to-many relations can be utilized only if we know the degree of correlation between those contexts. Consider the case when a new context source is introduced for a user context. Now, the HARD-BN framework can use it as a replacement for the primary context network (or source) for the user context provided its relative goodness is known with respect to the primary context network. For instance, consider that a new context network is available for *location1* where the context values *home/office/shopping/outdoors* are modeled from various ambient sensors such as light sensor, humidity sensor, etc. In this case, one will expect that the newly introduced context source will be less informative compared to high-resource consuming yet information rich sensors such as GPS. Accordingly, HARD-BN needs to know how informative the new source can be especially with respect to the primary context network (or source).

Hence, we propose the KL-divergence based algorithm which can measure the correlations between any two probability densities [143] in order to identify the strengths of these many-to-many mappings. In other words, such an algorithm can measure the quality of a new context source

compared to an existing one. Our preference to use the whole probability distribution of the context nodes and not just the maximum a posteriori estimations (which can be easily obtained from the support thresholds calculated in Algorithm 1) is justified by the fact that former can be helpful to achieve optimal Bayesian decisions. Also, KL-divergence is the asymptotic limit of the Maximum Likelihood (ML) criteria which is noted as the most appropriate similarity function for context nodes with equal priors for its values [143]. Here, we aim to deduce the difference between the probability distribution of an context predicted by its primary contextual network (i.e., $P(y = i|x)$) from that of the distribution predicted by the new context node (i.e, $Q(y = i|z)$). Assuming the probability of any of the values of context of Y is equal,

$$d(x) = \int P(y|x) log \frac{P(y|x)}{Q(y|x)} dy \tag{122}$$

$$d(x) = KL(P||Q) \tag{123}$$

where P is the probability distribution predicted by its primary contextual network and Q is that of the distribution predicted by the new context node. Then the primary contextual network-P can be said to be highly correlated with a secondary context network-Q_i, when the KL divergence between their probability density functions is minimal, i.e.,

$$d(x) = \underset{i}{\operatorname{argmin}}\, KL(P||Q_i) \tag{124}$$

Fig. 61 illustrates the probability distribution of the context *location1* under three conditions: (1) estimated from its primary context network modeled by HARD-BN (P), (2) estimated by the physical activities which has many-to-many relation according to Algorithm 1 (Q_1) and (3) estimated by a secondary network (modeled from ambient sensors) (Q_2). Now, the degree of similarity between them is given by their respective KL divergence:

$$KL(P||Q_1) = 0.6215; KL(P||Q_2) = 0.0745; KL(P||Q_3) = 0.8082 \tag{125}$$

where Q_3 is the uniform distribution for the context *location1*. Hence, we can conclude that despite their many-to-many relations the contexts *physical activities* and *location1* are not correlated enough, instead the distribution estimated from the secondary network (Q_2) is better correlated with respect to (P). Also, note that $KL(P||Q_1) \approx KL(P||Q_3)$, i.e., the distribution estimated by the physical activities is only as good as the uniform distribution (or uninformative distribution) for the context *location1*.

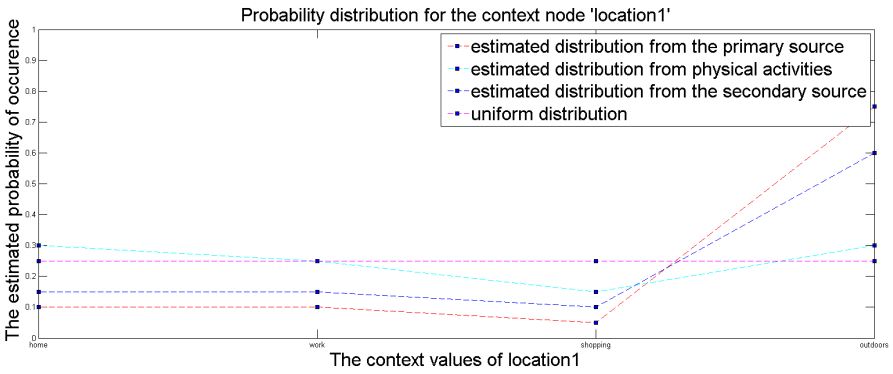


Figure 61: Estimated probability distribution for the context values of *location1* from different sources

Our experiments on the Personal Assistant application scenarios demonstrate the advantages of these techniques confirming their applicability for dynamic modern ubiquitous environments.

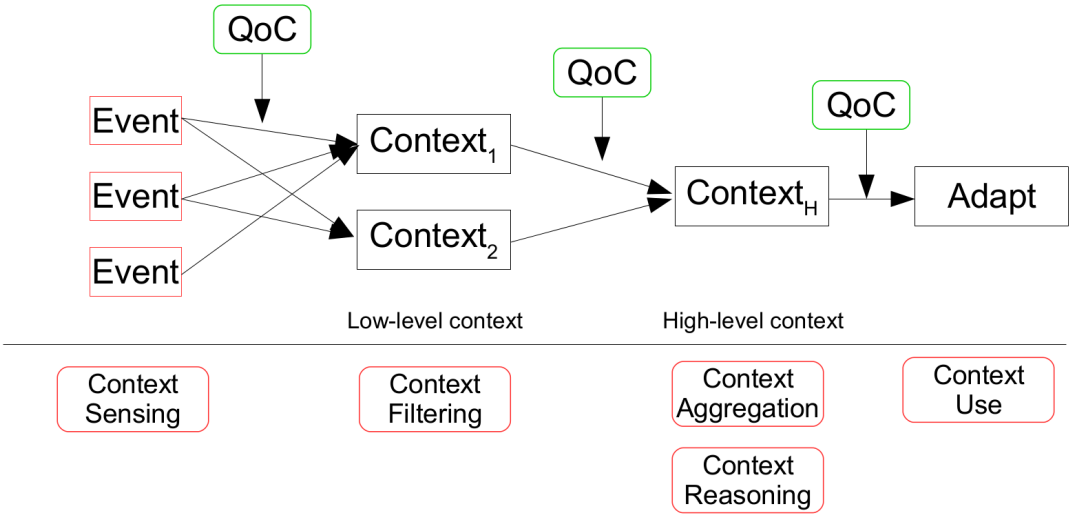


Figure 62: End-to-end quality management in context processing [144]

As these algorithms were developed in the final year of the project, the integration work is still ongoing and will be available in the future versions of SAMURAI SmartServer.

3.3.1.4 Assuring End-to-end QoC Support

Automating decision making in a continuously evolving dynamic context can be challenging. First of all, the right thing to do usually depends on the circumstances and context at hand. What might be a good decision today, could be a bad one tomorrow. Secondly, the system should be made aware of the impact of its decisions over time so that it can learn from its mistakes as humans do. In this study [144], we formulate a technique for decision support systems to mitigate runtime uncertainty in the observed context, and demonstrate our context-driven probabilistic framework for ubiquitous systems that addresses the above mentioned challenges. Our framework incorporates end-to-end Quality of Context (QoC) as a key ingredient to make well-informed decisions. It leveraged Dynamic Decision Networks (DDN) to deal with the presence of uncertainty and the partial observability of context information, as well as the temporal effects of the decisions. We have extended existing QoC frameworks and modeling paradigms by processing the quality attributes in an end-to-end fashion(as shown in Fig. 62), rather than considering QoC only at the source of the information (e.g. the sensors).

As it is in the early stage of research, integration is on-going. As a first step, DDNs will be integrated and exposed as APIs from SAMURAI SmartServer.

3.3.1.5 Causality Graphs

We will clarify how the probabilistic dependence is distinct from the causal dependence and motivate the use of this novel aspect of directivity into learning and prediction models.

Causality is the relation between an event or a set of events, i.e. the *cause* and another event or set of events, called *effect*. Granger [145] gave the following definition of causality in a bivariate time series context, which was later on used for testing the causation.

Given two time series, $X = X_i, i \geq 1$ and $Y = Y_i, i \geq 1$, the goal is to determine whether X causally influences Y.

Let Ω_t be the information which is available at time t. At time t, R_1 is the optimal forecasts of Y_{t+1} using Ω_t and R_2 is the optimal forecast of Y_{t+1} using all the information in Ω_t apart from the past

and present values of series of X, i.e., X_{t-j}. If R_1 is superior to R_2, then the series X_t contains information about Y_t not available elsewhere, and X_t is said to *cause* Y_t.

The motivation of using causal quantities instead of simply correlation coefficients comes from the fact that correlation cannot simply infer the causal relationship between variables [146–148]. Empirical correlation is proved to be inconsistent estimate of a causal relation.

If two events X and Y are correlated does not necessary implies that X causes Y or vice-versa. Their correlation gives certainly a hint of causation, yet one cannot simply infer also the direct causal relation. Such a reasoning is a fallacy.

The following cases of misuse of correlation to infer causation:

Reverse causation One case in which the causality is misinterpreted is the reverse causation case. The correlation between two variables can be actually reversed, when additional factors are considered. This case is illustrated in Figure 63. So, for example, the fact that it is rainy

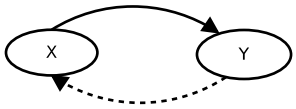


Figure 63: Reverse Causation Diagram

outside is highly correlated with the fact that people are having umbrellas on. Therefore, carrying an umbrella would increase the probability to rain. Here the correlation between the two events does not imply that wearing an umbrella would cause the rainy weather, but rather the other way round: it is more likely for people to carry umbrellas when it is rainy outside. Therefore the conclusion is again misleading.

String of causation In this case the correlation between two variables can be caused by a transitory factor. The string of causation structure is illustrated in Figure 64. For example, the

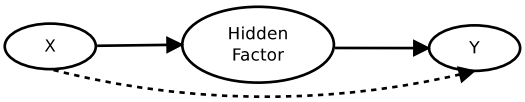


Figure 64: String of Causation Diagram

lack of religion can be correlated with high rates of depression, concluding that lack of religion leads to depression. While these two events are correlated the lack of religion is perceived differently and, in some cultures, may lead to discrimination, which can be the actual factor of depression. Therefore inferring that depression is caused directly by the lack of religion is yet another example of similar logical fallacy.

Common causal factor It can be the case that both X and Y have a common cause which makes the change observed in X to be reflect in Y in a linear manner. Figure 65 depicts the case.

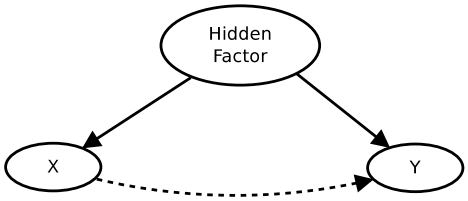


Figure 65: Common Cause Diagram

For instance, take X to be the event of sleeping in one's shoes and Y the fact of walking up with headache. While the two events are strongly related, it does not imply that walking up with headache is caused by sleeping with one's shoes. They are rather both caused by a third

factor, Z, that is going to bed drunk. Therefore this reasoning would be misleading in this situation.

The misinterpretation of such associations as if they were necessary is called by Yule a *fallacy* [149, 150]. His discussion indicates the same fallacious reasoning is reflected in the case of time series data as well.

For instance, he considers the decomposition of two events, such as the female cancer death rate and the quantity of imported apples consumed per head, into linear series, whose components are independent of time. Hence the two series, X and Y have the following expressions:

$$X_t = \alpha_1 t + \ldots + \alpha_n t^n + x_t,$$
$$Y_t = \beta_1 t + \ldots + \beta_n t^n + y_t,$$

where x_t and y_t are independent and identically distributed random variables but correlated with contemporaneous correlation r_{xy}. Since x and y have the same correlation, each time step, also the way X and Y progress induces a correlation which can give a false impression that there is also a relation between X and Y. Instead, the r_{XY} is the correlation which denotes if there is really a dependency between the two series. However, r_{XY} is still insufficient to infer if there is actually also a relation between them.

The causality of a system plays an important role and is yet difficult to handle. The lesson to be learned is that causal dependence is different from correlation and cannot be inferred easily. Essentially, the covariance can be misleading in several situations such as the ones described in this section. Hence it is clear that the use of causally relevant metrics, such as Directed Information (DI), has a better potential exploiting prediction problems.

We will give an example mentioned in [151] in which directed information is capable to detect also the information flow, while the mutual information fails to do so.

Consider two processes X^N and Y^N denoting each a T-tuple $[X_1, X_2, ...X_N]$ and $[Y_1, Y_2, ...Y_N]$ respectively, whose components are discrete Bernoulli i.i.d. and equiprobable random variables. To infer a causal relationship between X and Y we define the Y series based on the X process delayed with one:

$$Y_i = X_{i-1}, i \geqslant 1 \tag{126}$$

It can be seen that process X causally influences the process Y, while there is no causal influence the other way round.

We will compute the Mutual Information (MI) between X and Y and the DI in both directions.

The MI between X^N and Y^N is:

$$\begin{aligned} I(X^N; Y^N) &= \sum_{i=1}^{N} I(X^N; Y_i | Y^{i-1}) \\ &= I(X^N; Y_1) + \sum_{i=2}^{N} I(X^N; Y_i | Y^{i-1}) \end{aligned} \tag{127}$$

By replacing Y_i with X_{i-1} we get:

$$I(X^N; Y^N) = I(X^N; X_0) + \sum_{i=2}^{N} I(X^N; X_{i-1} | X^{i-2}) \tag{128}$$

Since $(X_i)_{i \geq 1}$ are i.i.d. then their mutual information is zero. Therefore,

$$I(X^N;Y^N) = \sum_{i=2}^{N} I(X^N; X_{i-1}|X^{i-2}) \tag{129}$$

From the definition of conditional mutual information we know that:

$$I(X;Y|Z) = H(Y|Z) - H(Y|X,Z) \tag{130}$$

So, by using (130) into (129) we get:

$$\begin{aligned} I(X^N;Y^N) &= \sum_{i=2}^{N} [H(X_{i-1}|X^{i-2}) - H(X_{i-1}|X^N, X^{i-2}))] \\ &= \sum_{i=2}^{N} [H(X_{i-1}|X^{i-2}) - H(X_{i-1}|X^N)] \end{aligned} \tag{131}$$

And since $(X_i)_{i\geq 1}$ are i.i.d. and they are Bernoulli distributed then:

$$I(X^N;Y^N) = \sum_{i=2}^{N} [H(X_{i-1}) - 0] = n-1 \tag{132}$$

We will compute now the DI in both directions, following the same rules applied in the MI case.

$$\begin{aligned} I(X^N \to Y^N) &= \sum_{i=1}^{N} I(X^i; Y_i|Y^{i-1}) \\ &= I(X_1;Y_1) + \sum_{i=2}^{N} I(X^i; Y_i|Y^{i-1}) \\ &= I(X_1;X_0) + \sum_{i=2}^{N} I(X^i; X_{i-1}|X^{i-2}) \\ &= \sum_{i=2}^{N} [H(X_{i-1}|X^{i-2}) - H(X_{i-1}|X^i, X^{i-2}))] \\ &= \sum_{i=2}^{N} [H(X_{i-1}) - H(X_{i-1}|X^i)] \\ &= \sum_{i=2}^{N} [1-0] = n-1 \end{aligned} \tag{133}$$

And the DI from Y^N to X^N is:

$$\begin{aligned} I(Y^N \to X^N) &= \sum_{i=1}^{N} I(Y^i; X_i|X^{i-1}) \\ &= I(Y_1; X_1) + \sum_{i=2}^{N} I(Y^i; X_i|X^{i-1}) \\ &= I(X_0; X_1) + \sum_{i=2}^{N} I(X^{i-1}; X_i|X^{i-1}) \\ &= \sum_{i=2}^{N} [H(X_i|X^{i-1}) - H(X_i|X^{i-1}))] \\ &= 0 \end{aligned} \tag{134}$$

As it can be seen from (133) and (134) DI has different value in the reverse direction, which provides us with the correct result, namely that X causes Y and not the other way round. This example shows clearly the advantage of the DI over the MI.

Various information-theoretical metrics that can adequately capture the directivity of information flow between variables has been proposed, including directed coherence [152,153], directed information [154–156], predictive information [157, 158] and directed coherence [153].

Let us examine now the current modeling frameworks dealing with the problem of prediction by means of causation on one hand, while on the other hand we motivate our attempt to fill the current gaps based on an improved information-based solution. We will introduce a conventional way to represent the causal relationship, i.e. trough directed graphs. Consider that this section will not contain details regarding the current causal models and structural learning but rather a short analysis of their gaps and the ways that can be improved. For a detailed overview refer to [147] and [159].

A directed acyclic graph $G =< V, E >$ represents a causally sufficient *causal structure* C for a population of units where the vertices of G denote the variables in C, and there is a direct edge from X to Y in G if and only if X is a direct cause of Y relative to V.

An example of such graph is shown in Figure 66.

Given such a causal structure, the aim is to infer the direct interactions between nodes, i.e. determine for each pair of nodes (X, Y) whether X is the direct cause of Y or Y is the cause of X in the sense defined by Granger [145].

Consider the following example that illustrates a case of study. We are interested to examine the relationships between shopping habits and factors such as discount prices, brand, etc and determine its direct and indirect causes. The variables and their values are shown in Table 10. Figure 66 depicts one causal model for our example.

The causal modeling task is to identify the deterministic functional relationships between the variables, that nature imposes, some of which are not observable. Unlike the classical statistical analysis driven by covariation and not causation, causal inference methods in machine learning are designed to identify the causal directions based on the available structures and build causal graphs. Our aim is therefore to identify the causal interactions by studying the information flow between variables of such dynamic processes. Our causal modeling framework can be summarized into the following points:

- Estimation of the causality between each two time series $\mathbf{X}$ and $\mathbf{Y}$ based on a finite number of samples of both processes, $i.e.$ $\mathbf{X}^N$ and $\mathbf{Y}^N$.
- Causal Graph Inference based on the information flow estimation in the previous point.

Acronym	Variable	States
AvgP	Average prices	[cheap, average, luxury]
DiscP	Discount prices	[yes, no]
StC	Store category	[clothes, drugs, food, super-market, jewelery]
Br	Brand	[non-famous, regular, famous]
Inc	Income	[poor, average, rich]
PerY	Period of year	[beginning-spring, end-summer, beginning-autumn, end-winter]
Shopping	Shopping	[at all, little, much]
Gender	Gender	[female, male]
Age	Age	[young, adult, old]

Table 10: The variables and their possible states in the shopping example

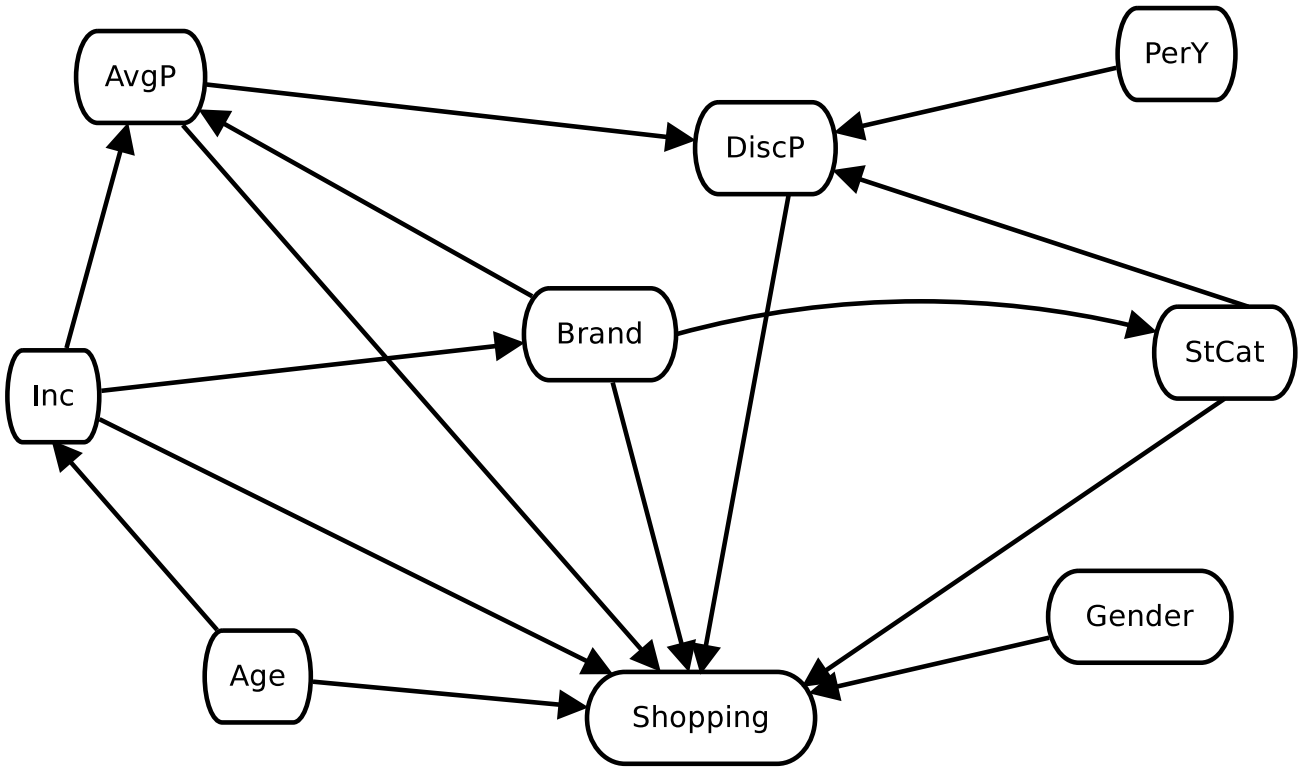


Figure 66: Example of Causal Relationships resented as a Causal Graph

- Improvement of the structure inferred in the previous point.
- Estimation of the root and child processes.
- Reproduction and evaluation of the data.

A detailed description of the causal modeling framework is given in Subsection 3.3.1.6.

3.3.1.6 Prediction of Multi-variate Stochastic Processes

Let us first clarify the class of processes we address. A stochastic process is a random variable X_t distinguished by its index t, which often represents time, but also distance. [160] A set of such indexed processes form a network of multi-variate stochastic processes. Our task is to infer the dependence relationships amongst the system's variables in general, and the causal relationships in particular, reproduce and predict the system evolution over time, based on snapshots of empirical

data. In order to identify the driving processes that dictate the evolution in one direction or another, we need to estimate the informational causality.

Estimating Information Causality

Information theoretic measures have a high potential in a variety of fields including neuroscience, telecommunications, mechanics and machine learning. We focus on exploiting such metrics in machine learning applications. In particular we want to examine and improve the potential of information theoretic techniques in order to solve prediction problems.

In the past researchers have used correlative measures in order to capture the interactions between elements of a system.

Both MI and Directed Information DI are two measures with a canonical role in a variety of statistic, signal processing and data analysis problems.

However, DI was more recently introduced and has been used to define the statistical causality between processes in such a matter that MI cannot. In his paper [151], Quinn summarizes the two quantities in short, the mutual information as being the *degree of correlation* (statistical interdependence) and directed information as being the *degree of causation*. The usability of directed information was outlined by Massey [155] based on Marko's definition for free information. Unlike the previous statistical studies dealing with prediction of random variables, we base our study on the information dynamics.

Consider m correlated processes: $\mathbf{X}_1, \mathbf{X}_2, \cdots \mathbf{X}_m$. Each $\mathbf{X}_k$ represents a stochastic process. The aim is to identify the information flow between each two time series $\mathbf{X}^N$ and $\mathbf{Y}^N$, based on finite number of samples.

Various information-theoretical metrics that can adequately capture the directivity of information flow between variables has been proposed. One such metric is DI.

In order to estimate DI out of samples, we apply a recursive partitioning scheme, by generalizing the algorithm proposed by Darbellay et. all in [161] to estimate multi-information in higher-dimension.

In his paper, he proposes a way to estimate the mutual information of 2 data vectors, X and Y using the *equiquantization principle*: the probability distributions are equalized so that at the end of the partitioning procedure the histogram of each vector is individually reshaped by using bins of variable widths but containing similar number of samples. The algorithm applies a recursive partition of the space $\mathcal{R}$ into rectangles of type $C = A \times B$ until the rate $\frac{P_{X,Y}(A\times B)}{P_X(A)P_X(A)}$ is the same in the next sub-cells, i.e. when the partition cells are *conditionally independent*. Once this condition is reached, it holds that:

$$\hat{I}(X;Y) = \sum_{A\times B\in\mathcal{R}} P_{X,Y}(A\times B) log(\frac{P_{X,Y}(A\times B)}{P_X(A)P_X(A)}) \tag{135}$$

Specifically he defines *restricted divergence* $\mathcal{D}^{\mathcal{R}}(X;Y)$ and *residual divergence* $\mathcal{D}^{\mathcal{R}}(X;Y)$ of two random vectors, and proves that $\mathcal{D}^{\mathcal{R}}(X;Y)$ converges to the true value of the MI $I(X;Y)$.

$$\begin{aligned}\mathcal{D}^{\mathcal{R}}(X;Y) &= \mathcal{D}(P^{\mathcal{R}}_{X,Y} \| (P_X \times P_Y)^{\mathcal{R}}) \\ &= \sum_{A\times B\in\mathcal{R}} P_{X,Y}(A\times B) log \frac{P_{X,Y}(A\times B)}{P_X(A)P_Y(B)}\end{aligned} \tag{136}$$

where $P^{\mathcal{R}}_{X,Y}$ and $(P_X \times P_Y)^{\mathcal{R}}$ are the *restrictions* of the corresponding distributions on the σ-algebra generated by $\mathcal{R}$, Σ.

$$\begin{aligned}\mathcal{D}_{\mathcal{R}}(X;Y) &= \sum_{A\times B\in\mathcal{R}} P_{X,Y}(A\times B) \\ &\cdot \mathcal{D}(P_{X,Y|A\times B}\|(P_X\times P_Y)_{|A\times B}) \\ &= \sum_{A\times B\in\mathcal{R}} P_{X,Y}(A\times B)\int f_{X,Y|A\times B} log\frac{f_{X,Y|A\times B}}{f_{X|A}f_{Y|B}}\end{aligned}$$

where $(P_X \times P_Y)_{|A\times B}$ denotes the conditional $P_X \times P_Y$ distribution on $\mathcal{X}\times\mathcal{Y}$ with density $f_{X|A}f_{Y|B}$. Darbellay develops his algorithm based on the following two propositions, whose proofs can be found in [161].

Proposition 3.1 *For every partition* $\mathcal{R}$

$$I(X;Y) = \mathcal{D}^{\mathcal{R}}(X;Y) + \mathcal{D}_{\mathcal{R}}(X;Y)$$

Proposition 3.2 *If a nested sequence of partitions* $\mathcal{R}^{(k)}$ *is asymptotically sufficient for* X, Y*, then*

$$\lim_{k\to\infty} D^{\mathcal{R}^{(k)}}(X;Y) = I(X;Y)$$

and

$$\lim_{k\to\infty} D_{\mathcal{R}^{(k)}}(X;Y) = 0$$

The main advantage of the improved procedure proposed by Darbellay is that, unlike the ordinal equiquantization, the space is divided into non-uniform cells which depends and adapts on the data distribution. In order to test if the cells are statistically similar, the chi-square goodness of fit test was used.

Estimation of the root and children processes

Once we know the information flow between each two processes we can build the tree structure of the causal relationships we estimate the root processes - using a non-linear estimator for ARX processes, and the child processes in relationship with the parents - using estimated linear coefficients.

The root processes are estimated using a non-linear estimator for Autoregressive Exogenous Model (ARX) processes, using the observed data.

$$X_t = f(X_{t-1}, X_{t-2}, \cdots, u_t, u_{t-1}, \cdots) \tag{137}$$

where f is a non linear function and X and u are the model regressors. The function f can include both linear and non-linear functions.

To estimate f we have used a wavelet estimator of the form [162]:

$$F(x) = L^T(x-r) + d + g(Q(x-r)) \tag{138}$$

where x is a vector of the regressors, $L^T(x)+d$ is the output of the linear function block and is affine when $d \neq 0$, d is a scalar offset, r is the mean of the regressors, $g(Q(x-r))$ is the nonlinear function in the shape of wavenet function [163] and Q is a projection matrix that makes the calculations well conditioned.

In order to estimate the children processes we assume they are linear AR and use a Least Squares Estimate for linear AR Process [164].

Consider we want to estimate the child node X_n having the parents $\{X_{r_1}, X_{r_2}, \cdots X_{r_p}\}$.

Then the process will be modeled by the following system of time series:

$$\begin{aligned} X_n(t) &= w + A_{11}X_n(t-1) + A_{12}X_{r_1}(t-1) + \dots \\ &+ A_{1,p+1}X_{r_p}(t-1) + noise(\sigma) \end{aligned} \tag{139}$$

We can write (139) in compact form:

$$X_t = B \cdot Z_{t-1} + \varepsilon_t, \quad \varepsilon = noise(\sigma) \tag{140}$$

where, T is the sample size and

$$\begin{aligned} X &:= (X_1, \cdots X_T) \\ B &:= (w \quad A) \\ Z_t &:= \begin{bmatrix} 1 \\ X_t \\ X_{r_1} \\ \vdots \\ X_{r_p} \end{bmatrix}; \quad Z = (Z_0, \dots Z_{T-1}) \end{aligned}$$

By A' we denote the transpose of matrix A.
By multiplying (140) with Z'_{t-1}:

$$\begin{aligned} X_t &= B \cdot Z_{t-1} + \varepsilon_t | \cdot Z'_{t-1} \\ X_t \cdot Z'_{t-1} &= B \cdot Z_{t-1} \cdot Z'_{t-1} + \varepsilon_t \cdot Z'_{t-1} \end{aligned} \tag{141}$$

We obtain then an estimate for B [165]:

$$\hat{B} = X \cdot Z' \cdot (Z \cdot Z')^{-1} \tag{142}$$

We can write B in terms of moment matrices [164]:

$$\hat{B} = W \cdot U^{-1} \tag{143}$$

where,

$$\begin{aligned} W &= \sum_{t=1}^{T} X_t \cdot Z'_t \\ U &= \sum_{t=1}^{T} Z_t \cdot Z'_t \end{aligned}$$

and the residual covariance matrix:

$$\begin{aligned} \hat{\sigma} &= \frac{1}{T - n_p} \sum_{t-1}^{T} \hat{\varepsilon}_t \cdot \hat{\varepsilon}'_t; \quad \hat{\varepsilon}_t = X_t - \hat{B} \cdot Z_t \\ &= \frac{1}{T - n_p}(V - W \cdot U^{-1} \cdot W') \end{aligned} \tag{144}$$

where $n_p = p + 1$ is the dimension of Z and

$$V = \sum_{t=1}^{T} X_t \cdot X'_t \tag{145}$$

The least squares estimates can be computed from a QR factorization of the data matrix K [164]:

$$K = \begin{bmatrix} Z_1' & X_1' \\ \vdots & \vdots \\ Z_T' & X_T' \end{bmatrix} \tag{146}$$

with the upper triangular matrix:

$$R = \begin{bmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{bmatrix} \tag{147}$$

According to [164], the QR factorization of K leads to a Cholesky factorization of the moment matrix:

$$\begin{aligned} \Gamma = \begin{bmatrix} U & W' \\ W & V \end{bmatrix} &= \sum_{t=1}^{T} \begin{bmatrix} Z_t \\ X_t \end{bmatrix} \cdot \begin{bmatrix} Z_t' & X_t' \end{bmatrix} \\ &= K' \cdot K = R' \cdot R \end{aligned} \tag{148}$$

from which we can derive the estimates for the parameter matrix B and noise covariance σ:

$$\hat{B} = (R_{11} \cdot R_{12})' \quad and \quad \hat{\sigma} = \frac{1}{T - n_p} R_{22}' \cdot R_{22} \tag{149}$$

3.3.1.7 Semantic Positioning via Structured Sparsity Models

Introduction

Semantic positioning is a new paradigm emerging with the IoT technology and its application to context-aware services in smart-spaces. Specifically, it refers to the problem of detecting user actions and locations based on prior characterization of the space and sensed data. Differently from classic positioning, input data are measurements of the interaction between human and sensors and location information is not a vector of coordinates but a a point in a topological map.

In this work, we tackle this challenge with a mere passive monitoring system in order to preserve user privacy, handle device heterogeneity, energy efficiency and utilizing low-complexity sensors that are able to capture events generated by human actions. We develop a structured sparsity model based on the notion of discrete Radon transforms on homogeneous space in order to construct mappings from events to actions and from actions to semantic locations. We propose algorithms for human activity detection and semantic positioning. Specifically, the Least Absolute Residual and Shrinkage Operator (LARSO) for human action detection, and a mixed-norm optimization to perform semantic positioning. Simulation results are shown to validate the proposed model and compare different algorithms.

The development of smart environments and IoT opened up an opportunity to reach the paradigm of an integrated context-aware technology [166]. While this pervasiveness of information offers the possibility to personalize context-aware services to user needs, if mishandled, it can represent a dangerous breach into the users' privacy. In this regard, a considerable effort has been invested into the development of novel security protocols for the IoT [167].

Differently, in this work we consider a mere passive monitoring problem of smart buildings, *e.g.* smart offices, in which a centralized system detects the actions occurring during a surveillance time [168] while preserving user privacy, $i.e.$, without user identification and communication between users and sensors. Generally, a sensor can register a type or multiple types of events generated by an object, $e.g.$ state-on or state-off. Moreover, based a given ontology, an action can be described by a set of events and different actions can share a sub-set of events. Thus, if we assume that sensors can transmit only the number and types of detected events, an action detection problem arises when multiple actions occur within the same monitoring time-interval (see Section 3.3.1.7.3

for more details). Furthermore, under the realistic assumption that sensors can be moved and participate to different actions in different locations, spatial-ambiguity is also introduced. Therefore, if we assume that a location ontology of the space is given, $i.e.$ a *semantic* characterization of the space depending on the actions that can take place, we consider the additional challenge of detecting the locations on a topological map of the environment sensed data only, $i.e.$, *semantic positioning*.

In light of the all the above, and with the additional needs of simple sensors and energy efficiency, we tackle the problems of action detection and semantic positioning under the assumptions of: unknown sensor location, un-identifiable users (passive monitoring), unknown number of users, no time-synchronization amongst sensors, unknown human behaviour patterns, simple state-based sensor output, known sensor-to-action mapping (action ontology) and known action-to-location mapping (location ontology). We utilize the notion of discrete Radon transforms on homogeneous space to construct a graph-based structured model, which yields sparse representations of events in the action space and of actions in the semantic location space. We propose a novel sparsity-inducing algorithm, referred to as Least Absolute Residual and Selection Operator (LARSO), to estimate human actions and a mixed-norm optimization to perform semantic positioning. Simulation results are shown to validate the model and compare different algorithms.

3.3.1.7.1 Problem Statements

Consider a smart-space consisting of a network of N_S sensors deployed to monitor human actions by means of *simple* measurements. Namely, assume that sensors are low-complexity devices capable to passively detect *events*, *e.g.* presence, or interaction between human and objects, *e.g.* touch. An event, for instance, can be the change in the state of an object ($e.g.$, firing, touching, etc.) or the movement of an object/person in the visibility field of a proximity sensor. Within a time-interval Δ_t, sensors can transmit to a central unit timestamped messages whenever an event is detected, $e.g.$, 1 when a sensor has detected an event, and 0 otherwise.

Let a_i denote a human action and $\mathcal{A} \triangleq \{a_1, \ldots, a_{N_A}\}$ be the set of actions of interest, where N_A is the total number of known actions. Although, actions can be modelled as a temporal sequence of events [168], in many application scenarios, such as those contemplated for the IoT [166], this model can be disrupted by practical problems such as the synchronization errors amongst sensors clocks, or unknown behaviour patterns. To circumvent these issues, we consider an action ontology where each action a_i is defined as a set of events, $i.e.$,

$$a_i \triangleq \{e_1, \cdots, e_{m_i}\}, \tag{150}$$

where e_j is the event captured by the j-th sensor, m_i is the number of events forming a_i and events can not be repeated within the same action.

Based on this model we construct an hypergraph $\mathcal{H}_a = (\mathcal{E}, \mathcal{A})$, where the action a_i corresponds to the i-th hyperedge (clique) of $\mathcal{H}_a$ [169]. Then, using the notion of discrete Radon transforms on homogeneous space [170], we derive a transformation matrix $\mathbf{A} \in \mathbb{R}^{N_S \times N_A}$ from events (sensors) to actions

$$A_{ij} \triangleq \begin{cases} 1, & \text{if } e_i \in a_j \\ 0, & \text{otherwise} \end{cases}, \tag{151}$$

where A_{ij} is the ij-th element of $\mathbf{A}$.

The matrix $\mathbf{A}$ can be considered a discrete representation of the action ontology, in which each column describes an action by the forming events. Notice, however, that this choice of transformation is motivated by the considered scenario in which nodes can only detect events in a completely independent manner and without any association to the users. Under less stringent conditions, $e.g.$ knowledge of the active user's ID, other choices of mappings can be considered in order to describe an action as a function of user-to-objects and objects-to-objects relationships.

Next, let $\mathcal{A}_t \subset \mathcal{A}$ be the set of actions executed during a time-interval Δ_t and let $\mathbf{y} \in \mathbb{R}^{N_S}$ be the observation vector whose i-th component is a signal *feature* extracted from the i-th sensor output[17], $e.g.$ state transitions [168]. By applying the Radon transform $\mathbf{A}$, the vector $\mathbf{y}$ can be written as

$$\mathbf{y} = \mathbf{A}\mathbf{x}, \tag{152}$$

where $\mathbf{x} \in \mathbb{R}^{N_A}$ is referred to as the *action-spectrum* of $\mathbf{y}$ and its i-th component relates to the frequency of occurrence of the action a_i occurring in Δ_t.

Unknown or incomplete actions that occur during the interval Δ_t are modelled by a noise vector $\mathbf{n} \in \mathbb{R}^{N_A}$, which is linearly combined with $\mathbf{y}$ yielding the noisy observation vector

$$\tilde{\mathbf{y}} = \mathbf{y} + \mathbf{n}. \tag{153}$$

Following, we categorize actions into semantic *classes*, where each class c_i contains actions sharing a certain semantic property, $e.g.$ "actions can occur in the same place". Thus, we define the class c_i as

$$c_i \triangleq \{a_1, \cdots, a_{k_i}\}, \tag{154}$$

and $\mathcal{C} \triangleq \{c_1, \ldots, c_{N_C}\}$ as the set of semantic classes.

In so doing, we construct a location ontology describing the space by classes. Thus, for each class c_i we associate a location s_i and, define the pair (c_i, s_i) as *semantic location* ℓ_i, $i.e.$ $\ell_i \triangleq (c_i, s_i)$. Finally, we consider the *semantic positioning problem* as the computational problem of detecting the subset of semantic locations ℓ_i's that best matches the observation $\tilde{\mathbf{y}}$.

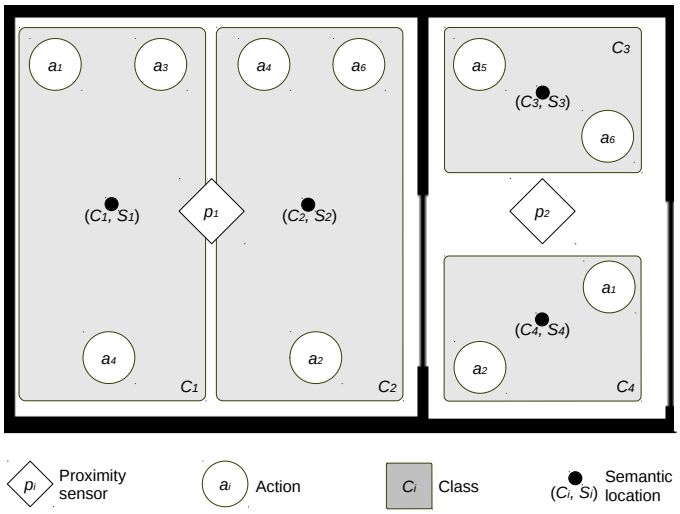


Figure 67: Illustration of a semantic map in which 4 semantic locations are used to identify 4 four significant spaces in two rooms.

In Figure 67 a small scenario for semantic positioning is illustrated. For the sake of clarity, sensors nodes are not drawn. In this scenario, there are 6 actions, 2 proximity sensors, 4 classes and, consequently, 4 semantic locations. Proximity sensors are active when any of the actions in their radio range are executed. For instance, if a_3 is computed, then p_1 is active.

Different classes can have common actions, $e.g.$ c_1 and c_4. Thus, ambiguous action classification can exist. However, to resolve this problem proximity sensors can be utilized. For instance, if p_1 is active and p_2 is not, then p_1 can resolve the ambiguity between the class c_1 and c_4. Hence, if p_1 and a_1 are active, then the semantic localization algorithm will return the semantic location $\ell_1 = (c_1, s_1)$. In a more general setting, this association is not trivial, and therefore, action and locations must be jointly estimated, $i.e.$ semantic positioning.

[17] An example is described in the simulation scenario in Section 3.3.1.7.3

3.3.1.7.2 Semantic Positioning Algorithm

In this section, we provide a two-step algorithm as a solution to the aforementioned semantic positioning problem. Specifically, we first perform action detection by estimating the action-spectrum $\mathbf{x}$ and, following, infer the semantic locations based on the estimated actions and semantic classes.

Action Detection

The action detection problem consists of estimating the set $\mathcal{A}_t$ from the noisy observation vector $\tilde{\mathbf{y}}$. This problem can be solved, for instance, by searching for an estimate of the action-spectrum $\mathbf{x}$, denoted by $\hat{\mathbf{x}}$, such that contains a minimum number of non-zero components [18] and $\mathbf{A}\hat{\mathbf{x}}$ best matches $\tilde{\mathbf{y}}$.

Therefore, for the the noiseless case, the action detection problem can be formulated with the sparse-recovery problem

$$\begin{aligned} \min_{\mathbf{x}\in\mathbb{R}^N} \quad & \|\mathbf{x}\|_0, \\ \text{subject to} \quad & \mathbf{A}\mathbf{x} = \mathbf{y} \end{aligned} \tag{155}$$

where $\|\cdot\|_0$ is the 0-norm.

The above minimization is a well-known non-convex optimization problem and, in [171], it is shown that the tightest convex approximation can be obtained by replacing the 0-norm with the 1-norm. When observations are affected by errors and statistical information on the noise is available, the Dantzig-Selector (DzS) and the Least Absolute Shrinkage and Selection Operator (LASSO) can be considered [172]. For the specific problem at hand, however, the noise is a vector representing unexpected or incomplete actions. Therefore, the vector $\mathbf{n}$ is generally sparse too. In order to induce sparsity also in the noise term, a novel sparse-recovery method, hereafter referred to as LARSO, is proposed

$$\hat{\mathbf{x}} = \arg \min_{\mathbf{x}\in\mathbb{R}^{N_A}} \|\mathbf{x}\|_1 + \frac{1}{2}\|\mathbf{A}\mathbf{x} - \tilde{\mathbf{y}}\|_1, \tag{156}$$

where the first and second terms of the objective function account for the sparsity in $\mathbf{x}$ and in $\mathbf{n}$, respectively.

The estimate of the action set $\mathcal{A}_t$ is then obtained by taking the action corresponding to the most significant values of $|\hat{\mathbf{x}}|$, for instance, by applying the "L-method" used in clustering [173]. Finally, note that the non-zero components of $\hat{\mathbf{x}}$ indicate how often the corresponding action is repeated.

3.3.1.7.3 Semantic Positioning from Actions

In this subsection we focus on the semantic positioning problem defined in Section 3.3.1.7.1. To this end, we derive a new Radon transform to construct a transformation matrix $\mathbf{C} \in \mathbb{R}^{N_A \times N_C}$ that maps actions into classes. Specifically, the ij-th element of $\mathbf{C}$, denoted by C_{ij}, is given by

$$C_{ij} \triangleq \begin{cases} 1, & \text{if } a_i \in c_j \\ 0, & \text{otherwise} \end{cases}. \tag{157}$$

Next, we assume that N_P proximity sensors are deployed in the monitored area and, for each class c_i, we consider $q_i < N_P$ simultaneous active sensors. Based on this assumption, we construct a matrix $\mathbf{P} \in \mathbb{R}^{N_P \times N_C}$ given by

$$P_{ij} \triangleq \begin{cases} 1, & \text{if the sensor } p_i \text{ can be active in } s_i \\ 0, & \text{otherwise} \end{cases}. \tag{158}$$

[18] The minimum number of non-zero elements is implied by $\mathcal{A}_t \subset \mathcal{A}$.

We define the observation vector $\mathbf{z} \triangleq [\mathbf{x}; \mathbf{y}_p] \in \mathbb{R}^{N_A+N_P}$, where $\mathbf{y}_p \in \mathbb{R}^{N_P}$ contains proximity sensor outputs and $[\mathbf{a}; \mathbf{b}]$ is the column-wise concatenation of two vectors $\mathbf{a}$ and $\mathbf{b}$. Based on equations (157) and (158), $\mathbf{z}$ can be written as

$$\mathbf{z} = \underbrace{\begin{bmatrix} \mathbf{C} & \mathbf{0}_{N_A \times N_C} \\ \mathbf{0}_{N_P \times N_C} & \mathbf{P} \end{bmatrix}}_{\mathbf{\Phi}} \mathbf{v}, \tag{159}$$

where $\mathbf{\Phi} \in \mathbb{R}^{(N_A+N_P)\times 2N_C}$ is the "double" Radon transform of $\mathbf{z}$ and the vector $\mathbf{v} \in \mathbb{R}^{2N_C}$ contains in the first N_C elements the spectrum of actions into classes, and in the remaining ones the spectrum of proximity sensors into location s_i's.

Noticing that the semantic location ℓ_i is defined by the pair (c_i, s_i), the components of $\mathbf{v}$ are tied by a block-structure explicated by pairing the v_i-th with the v_{i+N_C} elements with $1 \leq i \leq N_C$.

In practice, this structure can be induced by reshaping $\mathbf{v}$ into a $N_C \times 2$ matrix $\mathbf{V}$, and considering each i-th row-vector of $\mathbf{V}$, denoted by $\bar{\mathbf{v}}_i \in \mathbb{R}^2$, as a variable [174]. Thus, the semantic positioning problem reduces to the estimation of the matrix $\mathbf{V}$ with a row-structure sparsity[19], $i.e.$ estimating $\mathbf{V}$ with the least number of non-zero row-vectors.

Based on this model, we propose a mixed-norm optimization problem inducing row-structure sparsity that is given by

$$\hat{\mathbf{V}} = \arg \min_{\mathbf{V} \in \mathbb{R}^{N_C \times 2}} \lambda \|\mathbf{V}\|_{1,p} + \tfrac{1}{2} \|\mathbf{\Phi} \mathsf{vec}(\mathbf{V}) - \mathbf{z}\|_2^2, \tag{160}$$

where $\mathsf{vec}(\cdot)$ is the vectorizing function of a matrix, λ is a scalar controlling the structure regularization[20] and $\|\cdot\|_{1,p}$ is the $(1,p)$-mixed-norm defined as

$$\|\mathbf{V}\|_{1,p} \triangleq \sum_{i=1}^{N_A} \|\bar{\mathbf{v}}_i\|_p, \tag{161}$$

and $\|\cdot\|_p$ is the p-norm.

Notice that while the second term of the objective function in equation (160) is the residual error, the first one accounts for the aforementioned row-structure of $\mathbf{V}$. Moreover, based on the choice of p, different criteria and results can be achieved. For instance, utilizing the $(1,2)$ mixed-norm, $i.e.$ the *group* LASSO technique [176], the solution favourites row-vectors with the minimum squared-sum values. In contrast, the $(1,\infty)$ mixed-norm yields a solution with many components of equal magnitude [174]. As shown in the results, the latter provides the best trade-off between the probability of correctly detecting the number and the index of the unknown semantic locations.

Simulation Results

Before validating the framework proposed in Section 3.3.1.7.2, some considerations about the type of system and sensing mechanisms are necessary.

Firstly, we consider the sensor model. Sensors can simply react upon an event by emitting a binary signal, $e.g.$ $(0,1)$ [168]. Rather than counting the time-duration for each state change, a sensor processes only state-transitions by simple logics as those illustrated in Table 11. This eases the requirements on internal clock precision and global network synchronization. Additionally, sensors are only required to send the mere counts of events (feature) that they experience during the frame duration Δ_t, namely,

$$y_i = \sum_{m=1} f_{im}, \tag{162}$$

[19] The sparsity structure rises from the assumption that not $\mathcal{A}_t \subset \mathcal{A}$, and consequently, not all semantic locations are "active".

[20] In this framework, λ is computed as function of the maximum value of the projection $\mathbf{V}^{\mathsf{T}}\mathbf{z}$ [175], where $^{\mathsf{T}}$ is the matrix transpose operator.

where f_{im} is the value of feature measured by the i-th sensor at the sampling time t_m. In so doing, the communication cost measured as energy-per-bit between the sensor and the central system is minimal since the transmitted data is a finite and small number. Finally, the lack of communication amongst sensors and between sensors and users allows privacy issues and low-complexity network management. Following, we define two performance metrics, namely, the probability of set-detection and the set-cardinality ratio, which are used to validate both action detection and semantic positioning algorithms. Specifically, the probability of set-detection P_D and the set-cardinality ratio Υ are given by

$$P_D \triangleq 1 - \mathbb{E}\left\{\frac{|\hat{\mathcal{B}}\setminus(\hat{\mathcal{B}}\cap\mathcal{B})|}{|\mathcal{B}|}\right\}, \tag{163}$$

$$\Upsilon \triangleq \mathbb{E}\left\{\frac{|\hat{\mathcal{B}}|}{|\mathcal{B}|}\right\}, \tag{164}$$

where $\hat{\mathcal{B}}$ is the estimate of the set $\mathcal{B}$, $\mathbb{E}\{\cdot\}$ is the expected value, $|\cdot|$ is the cardinality of the set $\mathcal{A}_t$, $\mathcal{G}\setminus\mathcal{B}$ and $\mathcal{G}\cap\mathcal{B}$ are the set-difference and the set-intersection between two sets $\mathcal{G}$ and $\mathcal{B}$, respectively.

Hereafter, when referring to the performance achieved in the action detection, the probability P_D and the ratio Υ are renamed with $P_{\mathsf{D_a}}$ and Υ_{a}, respectively. Whereas for the performance obtained with the semantic positioning algorithms, they are indicated with P_{D_ℓ} and Υ_ℓ.

For the validation of the action detection algorithms, we consider a scenario with $N_S = 20$ sensors and $N_A = 30$ actions. If not specifically indicated, the sensor processing is based on Logic-1. We construct the set of actions $\mathcal{A}$ such that each action is formed by 4 events selected randomly and one event is uniquely associated to a sensor. The events of an action are randomly selected with uniform distribution such that two events can not be repeated within an action. Notice, that the events forming action can also be selected based on a given ontology, thus reflecting more realistic situations.

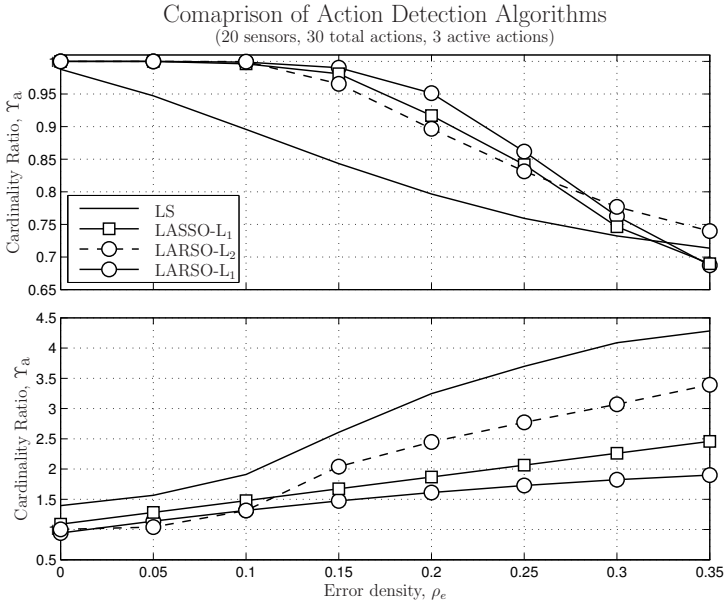
From the set $\mathcal{A}$ we build the sub-set $\mathcal{A}_t$ by selecting actions with uniform distribution. We assume that the cardinality of $\mathcal{A}_t$ is unknown. Using equation (152), we compute the vector $\mathbf{y}$ and simulate the observation $\tilde{\mathbf{y}}$ by adding a noise term $\mathbf{n}$. The non-zero elements of $\mathbf{n}$ are selected with uniform distribution and, with probability $1/3$ and $2/3$, they can assume values 2 and -1, respectively. We repeat this experiment 1000 times in order to compute the evaluate the performance of the proposed action detection algorithm.

The two subplots in Figure 68 show the probability of set-detection and the set-cardinality ratio as a function of the error density and action density given by $\rho_\epsilon \triangleq \|\mathbf{n}\|_0/N_S$ and $|\mathcal{A}_t|/|\mathcal{A}|$, respectively. The proposed LARSO technique is compared with two alternative optimization methods, namely, the Least Square (LS) and the LASSO. Moreover, the performance of the LARSO algorithm are evaluated with Logic-1 and Logic-2. It is shown that the proposed LARSO algorithm with Logic-1 is able to detect the unknown set of actions with higher probability than the alternative techniques while maintaining the set-cardinality ratio as close as possible to one.

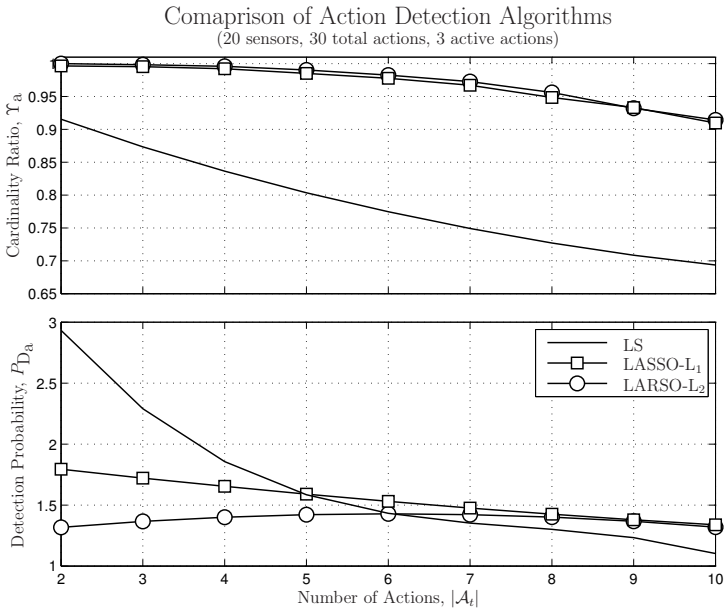
Table 11: Sensor's Logic

Signal		Logic-1[a]	Logic-2
Features	⎍	2	1
	⎿	−1	0
	⏌	−1	0
	0	0	0

[a] Considered in the simulations.



(a) Performance comparison as a function of the noise density



(b) Performance comparison as a function of the action density

Figure 68: Comparison of action detection algorithms. Solid and dashed lines refer to the results obtained with Logic-1 and Logic-2, respectively.

Additionally, Logic-1 provides better performance than Logic-2 since it enhances the difference between complete event and noise. In other words, Logic-1 yields higher SNR.

For the validation of the semantic positioning algorithm, we consider a scenario with $N_S = 20$ sensors, $N_A = 30$ actions, $N_P = 9$ proximity sensors and $N_C = 30$ semantic locations. Proximity sensors are equispaced in a 2-dimensional space covering a square of unitary edge-length and forming a grid of 9 points. Each proximity sensor has a detection range with $R = 0.35$. Each class c_i is formed with 4 actions randomly selected from $\mathcal{A}$ and, within the same class, actions can not be repeated.

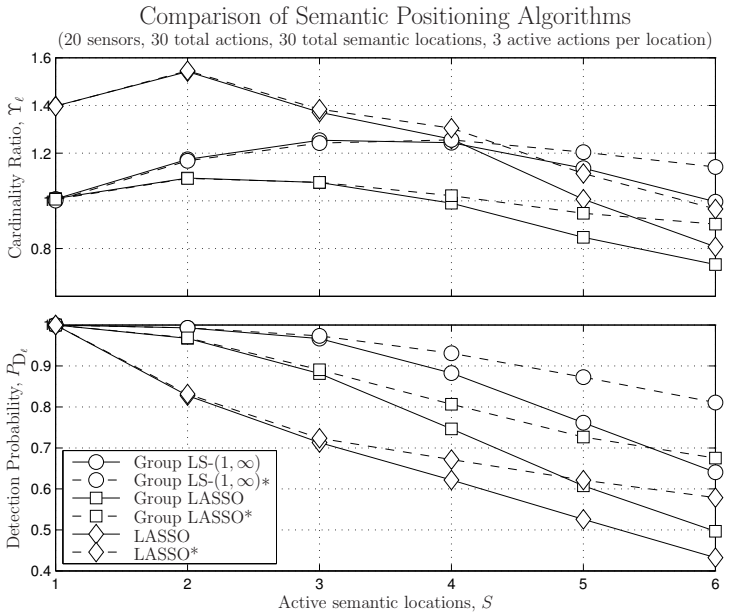


Figure 69: Comparison of semantic positioning algorithm performance as a function of the number of active semantic locations. Solid and dashed lines refer to the results achieved with estimate and exact information on the executed actions. The results are obtained with $\lambda = 0.6 \max(\mathbf{V}^{\mathsf{T}}\mathbf{z})$.

The locations s_i are uniformly distributed within the union of the proximity sensor coverage areas.

We construct the location ontology – relationship between classes and locations – by randomly assigning to each class a set of neighbouring proximity sensors. In so doing, we can correlate the execution of an action with the activation of the set of proximity sensors. In other words, we create a methodology to select a semantic location ℓ_i and, consequently, generate those events associated to actions and proximity sensors related to ℓ_i. As before, this synthetic data and ontology can be replaced by a more realistic logic based on the application scenarios.

The results obtained with this simulation scenario are shown in Figures 69 and they illustrate the probability P_{D_ℓ} and the ratio Υ_ℓ as a function of the number of active semantic locations, denoted by S. In this simulation we assume that for each active ℓ_i there are 3 active actions and, actions can repeat in multiple locations s_i's.

Three algorithms are compared, named, the proposed group LS with $(1,\infty)$-mixed-norm regularization, the group LASSO and the LASSO. Furthermore, we consider that actions and their frequency are either estimated with the LARSO method or that are known a priori. The performance related to the latter are shown with the dashed-line.

As a general assessment of the results, the proposed $(1,\infty)$-mixed-norm regularization provides higher probability of detection than all the other alternatives. However, this comes with a slight increase of the cardinality ratio metric with respect to the LASSO method. The impact of the action estimation for each algorithm can be noticed by the gap between dashed and the corresponding solid lines. It is interesting, however, to remark that for few active semantic locations this gap is negligible. The reason is that action detection algorithm is estimating a moderate size of action set.

Conclusions

We addressed the problem of semantic positioning with basis on IoT technologies. Namely, we tackled the problem of detecting the users' locations on a topological map of the environment based on prior characterization of the space and sensed data. To this end, we considered a mere passive monitoring approach posing stringent assumptions such as low-complexity sensors, simple measurements, unknown number of users, unknown sensor locations and unknown user identification in order to preserve user privacy, minimize the system requirements and energy consumption.

The main contribution of this framework is the original formulation of the problem via structured sparsity models. Specifically, we utilized the notion of discrete Radon transform to create mapping from events to actions and from actions to classes and, used proximity sensor information to enable semantic positioning. The LARSO algorithm was proposed to efficiently perform action detection from a noisy vector observation comprising the data transmitted by sensors. Finally, we proposed a block sparsity method based on mixed-norm objective function to perform semantic positioning.

Sufficient probability of action and semantic location detection were achieved, however, further investigation is necessary to quantify the theoretical bounds, improve the performance and, ideally, develop a single-step method. These are the research objectives for the future work.

3.3.1.8 Techniques to Improve Robustness of Context-aware Applications for Unexpected Events

In this work [177], we aim for a first step towards a more systematic method to analyse inconsistencies in context-aware rule-based behaviour at runtime to be more robust against unforeseen human interventions, exceptional circumstances and unexpected events. Faced with these observations, we try to provide an answer to the following questions:

1. How can we detect inconsistencies in context-aware decisions and actions that drive the dynamic behaviour of smart systems?
2. How can learning of patterns in event streams be used to anticipate and prevent failures?
3. How can context-aware event-based interaction patterns be used to identify significant deviations from common or expected human interactions?
4. How can we embed a software safeguard in the design of the application to reduce the risk of failures during unexpected contextual circumstances?

The objective is to decide which action to take given a contextual state of the environment, and to ensure safe or avoid unwanted behaviour. The software-based safeguard that we propose is based on the fact that we can translate rule-based adaptation into a classification problem. As input we use the contextual state information about the IoT environments and the actions taken by the actuators, and where the classifier learns whether the reached outcome of rule reached its objectives (i.e. matching with the desired effects). A key concern is that commands executed by different actuators (or humans for that matter) might not be independent. We therefore need a conflict resolution strategy that can identity hidden dependencies for conflicting actions. Our approach consolidates the different Event-Condition-Action (ECA) rules for each actuator into a separate decision tree, where each of the nodes in the tree can be annotated with an action and corresponding effect.

Due to its restriction to one sub-set of context modeling techniques, i.e., ECA rules, it is not fully integrated with SAMURAI SmartServer.

3.3.1.9 Mining Behavioral Patterns with CEP

As described in section 2.5 in deliverable D2.4 [124], one of the experimentation lines of BUTLER has been the possibility of enabling end-users to define virtual entities and, associated to said entities, a specific context. The BUTLER platform would then provide automatic synthesis of said context contents. The context abstraction has been defined within the BUTLER Information Model in section 4.3.3 in deliverable D3.2 [1]. An excerpt of the BUTLER Information Model section related to the context can be seen in the Figure 70:

The approach was novel not from the technological point but mainly by enabling end-users to define what context actually meant for them without tying them to a predefined (and limited) set of entities

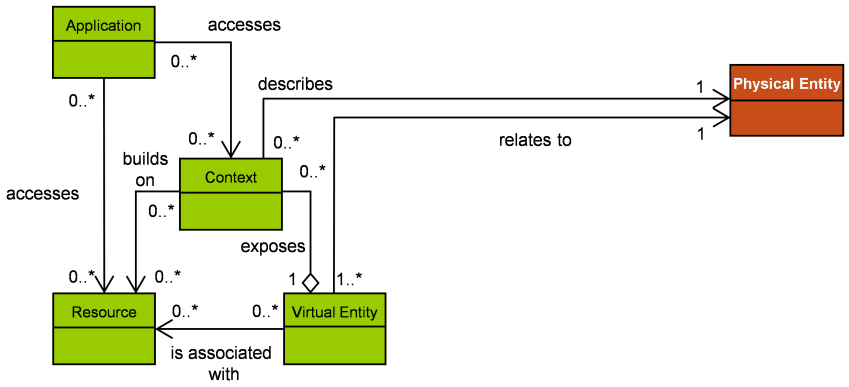


Figure 70: Context abstraction in the BUTLER Information Model.

and associated contexts. Another key point of the proposal is the automatic synthesis of the context contents by means of CEP technologies.

Our CEP-based approach for the management and automatic synthesis of context contents has not been devised to cope with behavior modelling although it could be considered in some specific (when the entity being defined by end-users can be mapped to a physical entity that shows a “behavior” and when the behavior can be defined in terms of simple mathematical and statistical operations applied over data streams).

The initial election for the CEP) infrastructure was Apache S4, “a general-purpose, distributed, scalable, partially fault-tolerant, pluggable platform that allows programmers to easily develop applications for processing continuous unbounded streams of data” [178] originally developed by Yahoo Labs and later transferred to the Apache Foundation for its incubation. There were several reasons for this election:

- S4 is an open-source framework (with an Apache license), which enables its use in research and experimentation projects.
- S4 is pure Java-based. It eases its integration with other Java-based frameworks especially those for creating web front-ends and associated business logic.
- S4 can be used with jBoss Drools [179], a rule engine, to define in real-time the rules that an S4-based infrastructure executes. It is possible to deploy new rules in real-time and therefore, accept new incoming data streams or generate new outbound data streams from the same set of incoming data streams.
- Although being in an alpha stage, when compared with other similar frameworks, such as Twitter’s Storm [180], it seemed to be the most “mature” at the moment of the framework election.
- It was created by Yahoo Labs! and had been adopted afterwards by the Apache Foundation for its incubation (*a priori* it meant that a wide community of developers would take care of the project).
- There are open-source efforts to extend its scope by enabling its integration with SAMOA [181] and therefore adding online analytics features to the automatic synthesis of context contents.

We have verified that it is possible for end-users with a minimum set of computer skills to manage the user interface and create context contents from the combination of data streams and/or by applying simple mathematical and statistical operations). See the Figure 71 that shows the context definition interface:

However, some limitations have been found in the aforementioned approach:

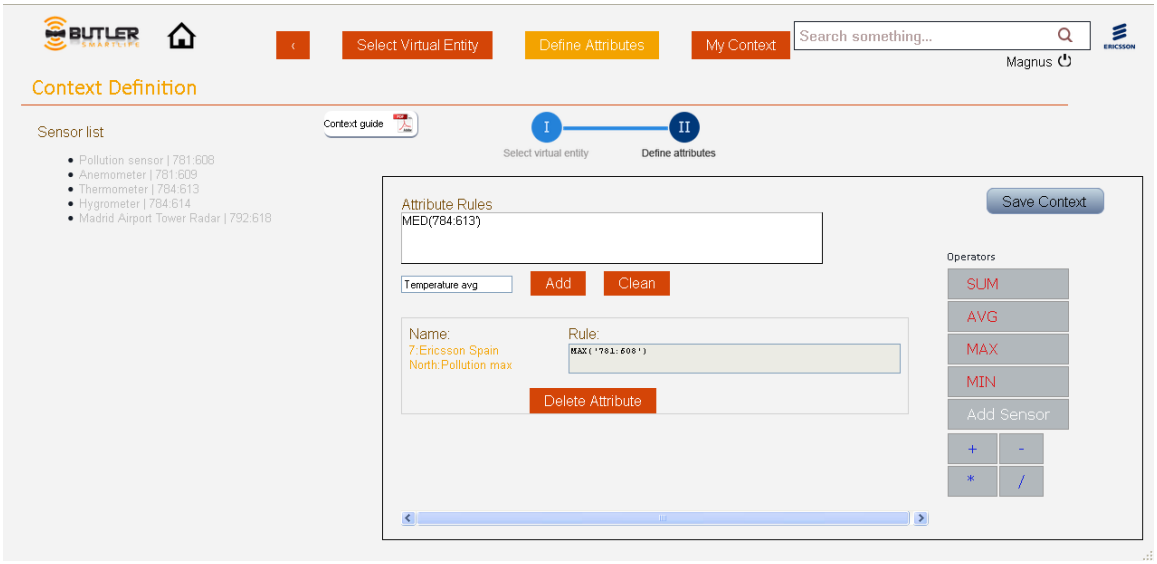


Figure 71: Context definition interface.

- Lack of a true web-based interface that eases the use of a rule engine. Here we face the usual trade-off between the power of the rule engine and the fact that only a small set of the end-users is really able to use a full-fledged rule engine language. However, we acknowledge that simplicity of use have to be prioritized if mass adoption is wished.
- Almost free definition of entities and associated context contents by end-users, which could lead to inconsistent definitions or resource-consuming computations.
- Lack of maturity of the CEP tools used in the BUTLER experimentation line. The initial election was Apache S4, based on the criteria described above. However, some of them became not as decisive as time passed by. The main issue with Apache S4 was the fact that the last release was June 2013 [182], and a general lack of progress in the next months. In the meantime, Storm was also transferred by Twitter to the Apache Foundation for its incubation [183] and could show some of the positive criteria already described with regard to Apache S4 (Apache license, wider community support, integration with SAMOA...).

3.3.1.10 Improving Performance of NIALM Algorithms for Usage Recognition of Home Electric Devices

The demand for electrical energy is constantly growing although the devices coming onto the market are getting more and more energy efficient. One of the main reasons for this effect is the overall number of devices which is increasing constantly outweighing the newly gained efficiency. Another influence comes from the rising human population. With different strategies and regulations the governments try to cut down the total energy consumption. To name a few of them [184]:

1. Device replacement by more efficient ones
2. Automatically switch off unused devices
3. Visualize energy consumption to give the users a better understanding of their energy consumption

But, on what grounds can a user decide which device is worth to be replaced? Or how does a system know which devices really are used at this moment and which can be switched off without harm or cut downs in user comfort? Electrical devices simply need to be monitored. Plug level energy monitoring solutions are commercially available, but they get very cost intensive if they are

managed and visualized from a central station. So, each monitored device requires additional hardware for the measurement process and additional installation effort for the communication to the central station [185].

Nonintrusive Appliance Load Monitoring (NIALM) algorithms represent a good way to monitor devices from a central location (e.g. the metering point). They even can be implemented on an available Smart Meter. This minimizes the hardware and installation costs. Hart [186] was the first person proposed a NIALM system. His method examined the steady-state behavior of loads.

A NIALM system is capable of reducing energy in all of the 3 previously named energy cut down strategies. If it recognizes an inefficient light bulb it may give the user a hint to replace it. On user absence it even can switch it off on its own.

Although NIALM research is going on for more than 20 years now and there are several existing products on the market, we have seen no major breakthrough in this field. Most existing algorithms are still too imprecise to come up with a constant acceptable device recognition rate. According to Zeifman [187] a minimal overall recogntion performance of 80 to 90% has to be reached in order for a NIALM system to be accepted by its users.

One of the aims of NIALM is to persuade the user to reduce the energy consumption of electrical devices by showing him his personal energy usage profile and provide him specific saving hints. There is a tradeoff between identifying a high amount of different devices with low accuracy but lots of visualization options and recognizing just a few device categories with a high accuracy but fewer user hint possibilities instead. Armel [185] shows that a simple improvement of the monthly energy bill can result in 3.8% energy savings. Real-time energy visualization without any disaggregation of the devices from the total load saves 9.2% whereas real-time consumption with disaggregation saves up to 12%. Even more energy can be saved with systems automatically controlling device states.

3.3.1.10.1 Related work and NIALM overview

An often studied topic in NIALM is the sampling rate applied in gathering the raw data. Basically, it is divided into macro and micro level measurements. For macro level measurements one is able to take use of a Smart Meter whereas micro level requires an external sub meter. In both research directions, many authors describe how to split up the devices according to their device states. Usually, the electrical loads are split up into rough groups of device types such as simple on/off devices, multi state devices, variable state devices and permanent on devices. Liang ([188], [189]) gives an overview over the whole subject of NIALM and the challenges in disaggregating the devices. A more recent publication that compares other works and their results was published by Zoha [190] and Reinhard [191]. Zoha also includes some products with state of the art technology and looks forward to possible new ways on how to tread the disaggregation problem.

Most existing products work in the macro level field. PlotWatt [192] and Bidgely [193] are two samples of companies providing such NIALM services. These companies don't publish how they disaggregate the device categories before they visualize the results for the user. Berges shows in [194] a prototype taking into account all aspects from the measurement system to the very end of device recognition.

Many NIALM researchers work with machine learning (ML) algorithms for the process of recognizing the devices. Figure 72 shows our proposed NIALM framework including the different processing activities needed in order to identify devices through general machine learning algorithms. In the following paragraphs a closer look into the different processing activities is provided.

A. Data acquisition

In the process of data acquisition input signals are measured by different sensors. In our application we used sensors for the voltage and the current but it is also possible to extend this set of sensors

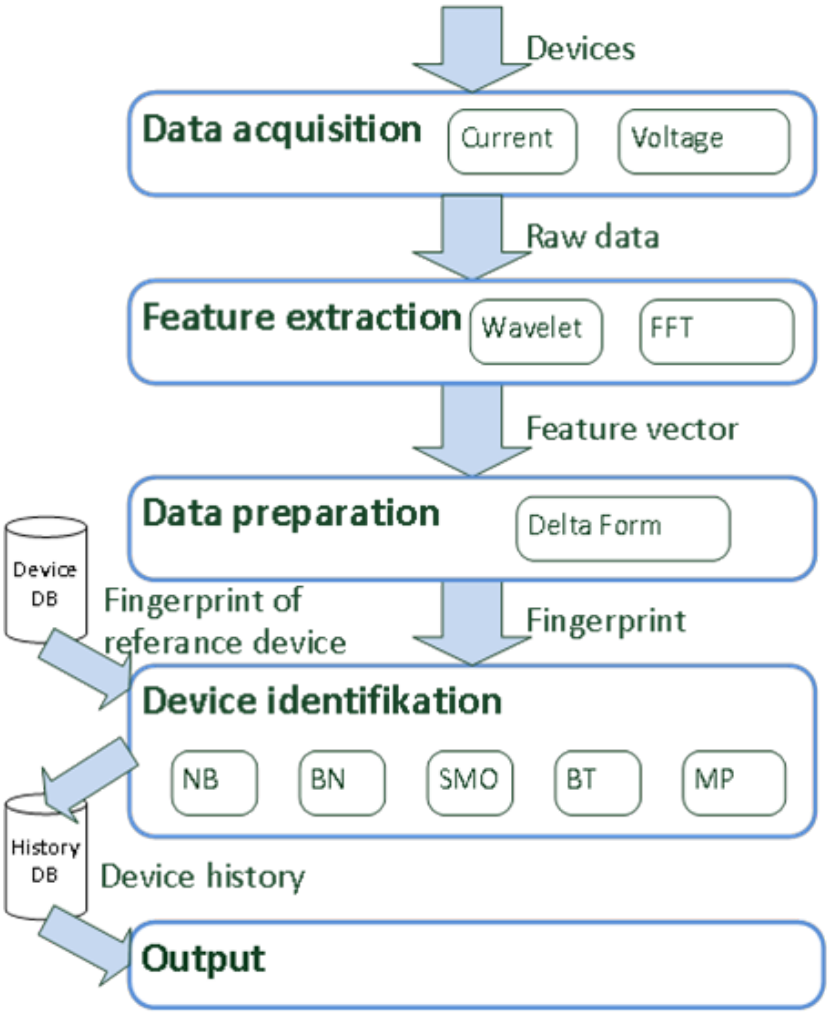


Figure 72: NIALM Framework of the accuracy comparison

e.g. by environmental sensors. With these 2 sensors measuring over a fix time duration we stay in the 3 dimensional space, that means we have a function with 3 unknown variables.

In our work we decided us for a sampling rate of the raw date of 5 kHz. On one hand a standardized low power microcontroller is capable of sampling at this frequency and on the other hand 5 kHz provides enough recognition information for the further processes. According to Armel [185] a recognition performance of 16 to 32 devices or device categories can be achieved with 5 kHz.

B. Feature extraction

A transformation of the raw data into a couple of parameters is called feature extraction. On the one hand features are used to get a robust transformation of the signal. Noise should be ignored as much as possible. On the other hand feature extraction tries to reduce the data amount of the measured raw data. The smaller the amount of data, the easier it is to send, store and also to handle it with machine learning algorithms.

We did several tests with different features. In the context of this work we only look to our most promising features regarding our task of recognizing devices. One is the Fourier Transformation of the waveform representing the electrical current flowing through the device. We only respect the low order odd harmonics 1, 3, 5, 7 and 9. Those are the harmonics that are produced by DC

converters. High harmonics are not considered in the calculation as they reduce their amplitude by the factor of $1/k$. Formula (165) shows the AC current of an ideal DC converter.

$$I_{AC} = \frac{4 \cdot I_{DC}}{\pi} \cdot \sum_{k=1}^{\infty} \frac{sin\left((2k-1)\,\omega t\right)}{2k-1} \tag{165}$$

Further we only looked at the absolute values in the Fourier Transformation because their information content is much higher than that of angle values. To still satisfy equation (165) the absolute values cannot be separated from the angle value before the data preparation process.

The second feature we utilized is the Wavelet Transformation of the current. We looked for an orthogonal and symmetric mother function that is often used in machine learning tasks. We found out that Wavelets of higher orders don't produce better results, even though they are more resource demanding. A Wavelet called Symlet fits best our demands. For a stronger parameter compression we followed a different strategy. We used the fact that Wavelet Transformation preserves the time localization information. Each transformation reduces data amount because of data redundancy. This is a similar effect as it can be observed with Fourier Transformation, where higher frequencies are mirrored. So we transformed the input signal in multiple iterations until it reaches the maximal parameter length. In our application we reduced the data amount to 4 parameters in 6 iterations. A disadvantage of wavelets is the more complex and resource demanding algorithm. Even the calculation time of a wavelet transformation with just one iteration depth is longer than a complete Fourier Transformation of the same signal.

The total number of parameters of a feature must be kept small because we are moving in the three dimensional space. The more parameter we have the higher gets the risk of finally over fitting a system.

C. Data preparation

The most efficient way to train machine learning algorithms is to provide a set of features representing just the interested device isolated from all others. This device fingerprint needs to be robust against noise and unique for each individual device. The same characteristics apply to the recognition process itself. That is the reason way almost all recent researches use the delta curve to isolate each device state change. We didn't work with delta curves until now. Instead we are directly using isolated measurements of single devices in a lab environment.

The feature extraction is calculated from the raw data, whereas a fingerprint represents the feature extraction of a single device or device state change or in other word the output of the delta curve calculation.

The fingerprint of a device can change if it is combined with other devices working simultaneously and if the used feature vector is not satisfying formula (166). In formula (166) $\vec{\Omega}$ is the feature vector of the cumulative curve and $\vec{g}_k$ the feature vector of the device k of a total of K devices. $\vec{g}_{\Delta t}$ represents the device state change during the time period Δt.

$$\vec{\Omega}\left(t + \Delta t\right) = \vec{\Omega}\left(t\right) + \vec{g}_{\Delta t} = \sum_{k=1}^{K} \vec{g}_k \tag{166}$$

D. Device identification

Device identification is the process where machine learning algorithms are used with their artificial intelligence. That means this learning algorithms learn their behavior from experience. We have looked at five different types of learning algorithms, all with promising capabilities. For this we used the implementations out of the tool Weka [195].

Bayes Net (BN) algorithms take a provided training set as input and produce a probability density function usually in form of a normal distribution. So the devices will be assigned to the category

with the highest probability. Full BN algorithms assign a conditional probability to each parameter whereas simpler forms of BN algorithms assume that all parameter are completely independent of each other. An example of such an algorithm is the Naïve Bayes (NB) algorithm.

The Sequential Minimal Optimization (SMO) algorithm was taken as a representation of the category of support vector machines. This binary classification algorithm works with the one versus one method. BrTree (BT) has been chosen from the category of decision trees. Like SMO it is also a binary classificator and works with the Best-first method.

Another category of machine learning algorithms that have been looked at is Multilayer Perceptron (MP) of the category of feed forward artificial neural networks. This algorithm has a very resource intensive training method, so it was limited to a maximum of 500 training iterations.

The last device identification algorithm can barely be treated as such and was taken in comparison to the other algorithms stated above. It simply classifies all devices to the category with the highest probability. Because our measurements are unbalanced, the accuracy of ZeroR (ZR) is always equal or above the inverse multiplication of the number of categories. ZR is independent of the input features and its kappa statistic is always equal to 0. See formula (3) for the definition of the kappa statistic.

3.3.1.10.2 Labeling and evaluation methods

A. Labeling

As already stated, the minimal recognition performance of a NIALM system must be between 80 to 90% in order to achieve the intended user acceptance [187]. This means that the performance of the device identification process must be significantly higher because the preceding data preparation process is already error-prone by itself. We assume that a minimal performance of 95% is needed to finally bring NIALM technology into applications with mass appeal. Most researches try to identify all devices in a household, but the rising number of devices in a household makes it more and more difficult to recognize all of them. Assigning the devices to categories results a constant number of recognition classes, even with a variable number of devices. So the recognition performance is independent of the device number. Statistically the performance can also be improved in lowering the amount of recognition classes. So there is a tradeoff between the amount of classes and the information that can be gained from each class. The more classes exist, the more specific tips can be visualized for the user.

A few existing NIALM products follow this trend. Most of them disaggregate the load in just a few classes [196]. One such a product is PlotWatt [192]. It collects the information from a Smart Meter and splits the devices into the categories Heating & AC, water and septic, dryer, always on, refrigeration and others. So it is capable of recognizing 4 categories (always on and others don't need to be recognized). In spite of that, it still is able to provide useful tips about how to minimize the energy consumption.

In our work, four different ways of labeling the data with different category sizes have been analyzed and compared with each other. TABLE 12 shows the used labeling list.

Table 12: List of recognition categories

Label	Label Name	Nr. of categories
DL	Device Labeling	31
TL	Technology Labeling	13
NL	Normed current curve Labeling	9
UL	User Labeling	6

1. Device Labeling

   The most used categorization in ongoing research and recent publications separates each device into an individual category or even to its device states. In most households the amount of devices is bigger than the 16 to 32 devices that can be recognized with 5 kHz. Another drawback is that some algorithms like support vector machines increase their complexity potentially with a raising number of categories. We labeled each device with a unique label that needs to be recognized.

2. Technology Labeling

   In order to decrease the number of categories a grouping of devices using the same technology looks most promising. So lighting devices are split up into light bulbs, fluorescent and LED lights, electronic devices to desktop computers, printers, notebooks and mobile phones and screens are split into ones running on LCD- and LED-technology. We assume that each device technology has its own power consumption pattern that can be discovered by machine learning algorithms.

3. Normed current curve Labeling

   A more intuitive labeling method is grouping the devices with similar normed current curves. The norming condition was defined as the mean current over the measurement time interval equals one. The curves have been normalized to look to the wave form and not to the power. The aim is to get an error rate in misclassification as small as possible with the highest amount of categories. We split the devices into the following categories:

   Ohmic devices have no harmonics in their power waveform and their current zero-crossing is at the same time as the one of the voltage. Typically they occur in devices producing light or warmth. Lighting devices usually require less than 100 W whereas warmth devices like a convenient hotplate have more than 500 W. So, ohmic devices can be split into 2 categories according to their power consumption. Miss recognition can be caused by dimmed lights, because they also have harmonics.

   If the zero-crossing of voltage and current are not at the same time, the load is inductive or capacitive. In our set of devices, there are only inductive loads. Typical inductive devices are motors or inductive hotplates.

   A more efficient way to produce light is using fluorescent or LED lamps. Both technologies produce sharp peaks in their current wave form. A dimmer moves this peaks along the time axis. Cooling devices have been recognized by their very unique wave form. The same was true for coffee machines.

   The DC voltage of all electronic devices is stabilized with a smoothing capacitor. Characteristically it is charged in a pulsed way which results in high amplitudes and lots of harmonics. Since 2001 Europe laws claim for certain devices to correct their power factor to reduce physical load in the grid. This correction can be measured in their current curve. So, electronic devices can be grouped into devices with and without corrected power factors.

   Another technical category that was identified is the stand-by consumption of electronic devices. Low power with high reactive part and lots of noise are the recognition features.

4. User Labeling

   The user category groups devices according to the interests of the users. It's reasonable to recognize devices from each category with high accuracy instead of each single device with a high error rate. The user categories are constructed to give the user as much information as possible to identify his personal energy saving possibilities. The categories are lighting, screens, computers, kitchen devices, coffee machines and cooling devices.

B. Performance evaluation

A number of publications evaluate a performance value in respect to the work done. The way they calculate this value differs for each appliance. How to rate a NIALM performance is discussed

in [197]. On one hand systems are using different input values for their validation. On the other hand different statistic methods are used in the calculation of their results. This makes it hard to objectively compare the different studies.

1. Data - Set
   The performance evaluation is made with a total number of 3104 measurements and 31 different devices. The scope of these measurements was to capture all possible device states in which devices remain for a longer time period, let's say 10 seconds. Very small energy consumption like standby states has been measured as well as running computer with highly variable power consumption. Special focus has been put on variable loads. For those means, different current curves of dimmed lamps and charging computers have been captured. Compared to simple static loads, variable loads demand on a much higher amount of measurements to capture most states. This resulted in a highly unbalanced dataset in which each device has a different amount of measurements. The calculation of the performance of an algorithm with unbalanced data can be misleading because there is no correlation between the number of real device events that must be recognized and the amount of different states that have been measured. Statistically it's easier to categorize unbalanced data as shown in TABLE 4. In statistics different ways are used to correct the unintended side effects created by unbalanced data sets. One way is in making all categories equal sized by randomly deleting data. Another way is in using a cost-matrix. A cost-matrix weights each category independently so they can be equally weighted. In our approach we take use of the cost-matrix because it still allows using all measurements available.
2. Validation
   To measure the performance of a machine learning algorithm usually the labeled input data is split into training data and validation data. The amount of data is mostly limited, but usually, the more training data the algorithm is fed the better it gets. More data also lowers the possibility of over fitting the algorithm. In our application we used a 10 fold cross-validation method. So, each measurement belongs during 10 validation processes exactly one time to the group of independent validation data. The results get more precise with a significant higher amount of validation data. The cross validation method guarantees that the test data is always independent from the training data. This is important to detect over fitted systems.
3. Number of categories
   The more categories the labeled data has, the more difficult it gets to choose the right answer in a random fashion.
   The kappa statistic measures the agreement of prediction with the true class, so the probability of choosing the right answer by chance is subtracted. In (167) $P(A)$ is the proportion of times the k raters agree, and $P(E)$ is the proportion of times the k raters are expected to agree by chance alone. $P(E)$ is represented by the ZR algorithm.

$$K = \frac{P(A) - P(E)}{1 - P(E)} \tag{167}$$

   We used the kappa statistic in our application to allow a more meaningful comparison when categorizing with a different number of categories while the absolute performance value decreases. However it still doesn't weight the categories equally.

3.3.1.10.3 NIALM algorithm performances and interpretations

We looked at the behavior of NIALM algorithms when labeling the measurement data differently. Not all NIALM algorithms behave the same on differently labeled data and also an influence of the different features selected is expected. In our comparison we tried to identify such a behavior.

TABLE 13 shows the results of unbalanced data whereas in TABLE 14 the results of balanced data are shown. Both tables show performances over 95% in bold.

Table 13: Cross-validation results with kappa statistics

Cat	Features	BN	NB	SMO	BT	MP
DL	Wlet	87.1%	74.2%	9.9%	90.0%	36.3%
	FFT	84.8%	52.5%	30.4%	93.8%	74.7%
TL	Wlet	82.3%	56.9%	12.4%	91.9%	43.9%
	FFT	83.2%	38.2%	25.6%	**96.1**%	71.0%
NL	Wlet	89.0%	40.9%	16.6%	**97.2**%	67.5%
	FFT	76.4%	29.9%	14.2%	**96.6**%	79.2%
UL	Wlet	87.4%	27.9%	16.9%	**96.3**%	56.0%
	FFT	78.5%	15.5%	15.3%	**96.7**%	70.5%

Table 14: Cross-validation results with equally weighted category cost-matrix and kappa statistics

Cat	Features	BN	NB	SMO	BT	MP
DL	Wlet	87.7%	77.5%	9.7%	88.4%	30.6%
	FFT	89.7%	64.5%	21.8%	**95.1**%	68.0%
TL	Wlet	90.6%	76.9%	15.9%	93.4%	48.8%
	FFT	92.3%	60.4%	21.8%	**97.3**%	72.2%
NL	Wlet	**97.1**%	83.1%	25.0%	**98.9**%	78.7%
	FFT	92.9%	77.1%	21.9%	**98.5**%	88.8%
UL	Wlet	**95.9**%	76.3%	18.6%	**98.6**%	47.1%
	FFT	93.4%	69.1%	15.0%	**98.1**%	62.1%

As expected the mean performance over all algorithms turned out better with unbalanced data but surprisingly the highest performance could be reached with balanced data. In all comparisons the rather simple algorithm BT performed the best and in the most label comparisons, the label NL reached the best results.

A. Feature

The difference between the compared features is minimal. The mean performance of the Fourier Transformation is slightly better although the highest performance is reached by a Wavelet Transformation. However, one has to take into account that the calculation time of a Fourier Transformation is by factors shorter. For our data, in which the Wavelet Transformation was calculated in 6 iterations, the result of the Fourier Transformation could be displayed 18.2 times earlier. Even a single iteration of the Wavelet Transformation takes about 3 times longer in the simulation environment Matlab [198].

B. Algorithm

Bayes algorithms clearly perform better with balanced data because they are based on a balanced model. A higher amount of training data samples improves only the precision of the model. A model for unbalanced data can be found in neural networks. Each training sample increases the possibility of the sample category and its precision.

A different behavior can be found in support vector machines. Their performance depends on the number of categories. Small numbers of categories perform much better than large ones. The used algorithm implements a one versus one method. That means with a total of 6 categories 15 comparisons have to be done whereas 31 categories require 465 comparisons.

The decision tree performed well on balanced as well as unbalanced data although its model is based on balanced data. It also performed great on a small and big number of categories. Surprisingly, it shows the best performance for each comparison done in the context of our work. Further, it belongs to the fastest classification algorithms as it builds a simple binary decision tree in its training phase with a logarithmic effort. This makes the algorithm applicable on most low power micro controllers.

C. Balanced / Unbalanced data

The significant influence of unbalanced data can be seen in TABLE 15. It shows the performance of the algorithms ZeroR on our data set. The category with the most devices with variable device states has the most measurements data. For example a category of NL labeling includes more than 50% of all data. A small correction is Kappa Statistics as it is already used in our results. Whether the data is balanced or unbalanced the performance of ZeroR with corrected Kappa Statistics is always zero.

Table 15: Classification performance of the ZeroR algorithm

ZeroR	DL	TL	NL	UL
Unbalanced	16.2%	33.2%	50.1%	51.6%
Balanced	3.2%	7.6%	11.1%	16.6%
Kappa statistics	0%	0%	0%	0%

D. Labeling

In most label comparisons the label NL reached the best results. No surprise was that device labeling has the worst performance. With a rising number of devices an even decreasing performance is expected. The performance of other categories will remain about the same as new devices are categorized to an existing category.

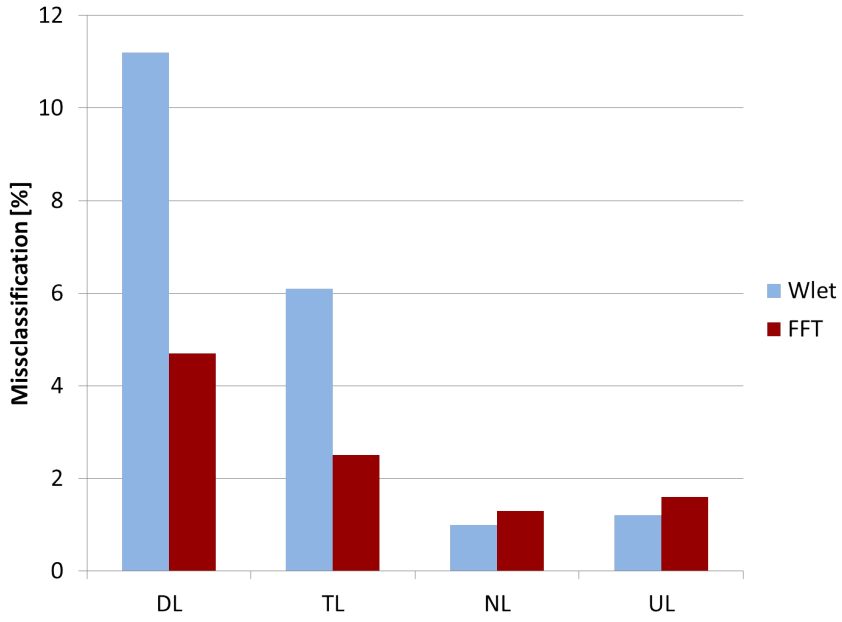


Figure 73: Misclassification of the algorithm BT on balanced data

The performance of the algorithm BT on balanced data is shown in Figure 73. Labeling with normed current curve (NL) has about 7 times less misclassification samples compared to device labeling (DL). For the calculation the mean performance of the FFT and Wavelet was taken.

As it is hard to compare different classifications with different amount of categories TABLE 5 provides a normalized misclassification rate divided by the number of categories. Cat Nr. of categories Unbalanced BT Balanced BT

Table 16: Misclafication per category of the BT algorithm

Cat	Nr. of categories	Unbalanced BT	Balanced BT
DL	31	0.22%	0.26%
TL	13	0.31%	0.33%
NL	9	0.17%	0.13%
UL	6	0.28%	0.23%

A significant performance improvement step in using NL can be identified on balanced and unbalanced data. As the results of balanced data are more of interest, a closer look to these results was taken. As shown in TABLE 16, a normalized error of about 0.13% per category could be achieved with NL. Compared to the other categories, with an average error rate of about 0.27%, the observed performance of NL was about double.

None of the investigated algorithms was able to find the expected individual pattern in labeling the devices according to their technology (TL).

In numerous NIALM approaches, the researchers try to identify even more categories (e.g. [194]) as most devices have multiple states. So each state change is a category by itself. An identification of device states even increases the number of categories beyond the number of different physical devices. One challenge resulting from this approach is how to identify variable state changes. Another problem is that a rising number of categories usually results with a significant increase of misclassification which also could be shown in this work. In our opinion it is clearly more valuable to use fewer categories with a higher accuracy.

3.3.1.10.4 Achieved improvements of NIALM and future activities

In the context of ongoing NIALM research, an evaluation of different recognition approaches was done. As one of the main barriers, preventing NIALM technology from being widely adopted, the dissatisfactory recognition performance of existing solutions has been identified. Different algorithms for device identification where examined with the prerequisite that a minimal performance of 95% must be reached in order for them to be acceptable for real users. A rising number of devices in a household results in a rising number of devices and device states to be recognized. As it is extremely difficult to identify each single device in a household with a high accuracy, grouping devices together limits the number of categories to be recognized and enables a high recognition performance. Statistically, as the number of categories decreases, the accuracy of recognizing them increases. This could be shown with a classification of 4 different labeled data sets, representing a total of 31 loads including several variable state loads. The comparison was done with the Fourier and Wavelet Transformation in the process feature extraction and different kinds of machine learning algorithms in the process device identification were compared. Statistical tools have been included to enable a better comparison of the results. The best recognition performance of 98.9% was reached by the BrTree algorithm of the family of decision trees with the Wavelet method. For this performance all measurement data has been split into 9 categories according to their normed current curve. Whereas the other three device labeling categories achieve an averaged misclassification of about 0.27% per category, we could reach a misclassification of 0.13%. The performance of our categorization remains constant with a rising number of devices. Although Wavelet performed

slightly with the best results, the mean performance of Fourier Transformation was better and its calculation time is much faster.

We are aware that recognizing less categories results in less possibilities of displaying information to the user. However, in our view the additional value of reliable real time information - clearly distinguishing between several device groups - is still a considerable additional benefit for the user compared to the energy bill he gets today.

Future improvements:

The normed current curve labeling (NL) was defined based on human experience and performed better than other categories with an even lower amount of categories. We expect a significant improvement of this approach in using cluster algorithms to split the devices or device states into NL categories. Unsupervised algorithms as used in cluster algorithms search for similarities in the features and not in the current curve. If supervised algorithms, as used so far in our work, are trained with labeled data coming from cluster algorithms, a significant improvement of our labeling is expected even with a higher amount of categories. Another try is by using a public data set like it was announced in [199]. This would allow comparing these algorithms directly with other works.

3.3.2 Frameworks and Tools to Support Behavioral and Situational Awareness

3.3.2.1 SAMURAI: Enhanced Stream Mining with Support for Data Persistency and Multi-tenancy

In the Internet of Things, heterogeneous and distributed streams of sensor events is a driver for contextaware behavior in intelligent environments. However, processing the event data usually cross-cuts the business logic of IoT applications and offering such reusable functionality as a service towards a variety of customers with different needs is often faced with scalability concerns. We developed SAMURAI, a multi-tenant streaming context architecture that integrates and exposes well-known components for CEP, machine learning, knowledge representation, NoSQL persistence and in-memory data grids. SAMURAI pursues a twofold approach to achieve scalability: (1) distributed deployment with horizontal scalability, (2) shared resources through multitenancy. For the scenario used in the experimental evaluation of our architecture, the results show little overhead to support multi-tenancy, with near-linear scalability and flexible elasticity for deployment schemes with data partitioning per tenant [200].

Our event-based streaming architecture uses the Spring Framework[21] and Jersey[22] to expose well-known software libraries for CEP, machine learning and knowledge representation as RESTful services. These are illustrated in Figure 74.

Events can be *simple events* that carry slivers of meaning in themselves, and *complex events* which summarize, represent, or denote a set of single events which combined denotes a 'pattern of events'. An event is represented as a set of typed key-value pairs that can be easily serialized into the JSON format. The example below illustrates the type and an instance of an accelerometer event that we use for activity recognition. In this example, the *x*, *y* and *z* values hold the acceleration values along these axes. The *timestamp* field represents the number of milliseconds passed since January 1, 1970 UTC:

The architecture offers publish/subscribe capabilities to have clients (applications or subsystems) notified when particular (patterns of) events occur with *push* notifications implemented as REST callbacks (see following section). The architecture has three basic components to hold events:

- **In-memory Data Grid**: This is a distributed in-memory container for events based on Hazelcast[23].

[21] http://projects.spring.io/spring-framework/
[22] https://jersey.java.net/
[23] http://www.hazelcast.com/

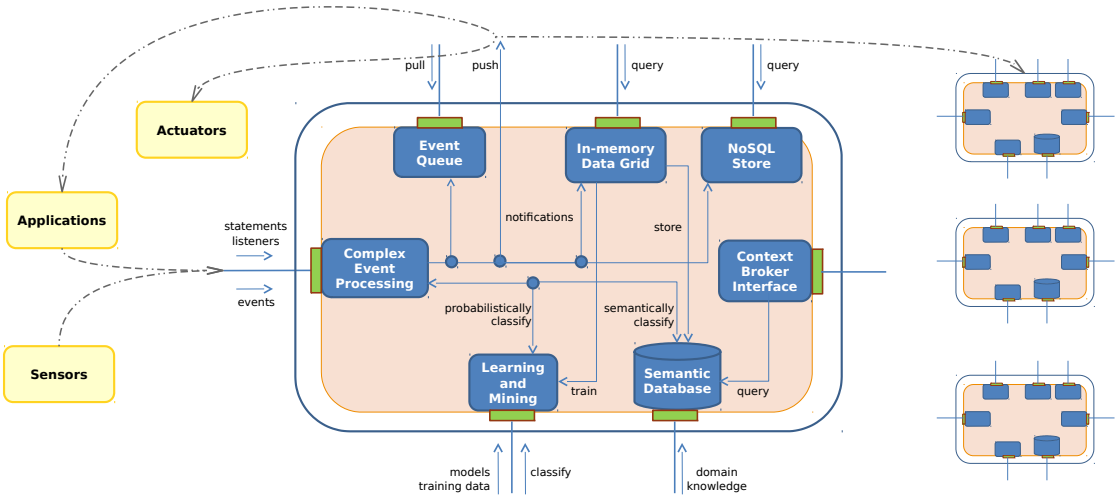


Figure 74: Conceptual overview of streaming architecture

```
{                                   {
  "timestamp": "long",                "timestamp":1340099550210,
  "x":"double",                       "x":-8.308,
  "y":"double",                       "y":-1.9477,
  "z":"double"                        "z":4.099
}                                   }
```

Figure 75: Example of an event type and an event instance

- **Event Queue**: Clients that do not support push notifications through REST callbacks can register a queue to hold events and poll that instead.
- **NoSQL Store**: The events can be optionally stored in a persistent way using CouchDB[24] as a RESTful database.

Their RESTful APIs follow the CRUD mapping on HTTP methods to create (POST), read (GET), update (PUT) or delete (DELETE) events.

3.3.2.1.1 Complex Event Processing (CEP)

For step counting and fall detection, we process events from the tri-axial accelerometer found in most smartphones. Below is a short overview of the domain specific event stream processing algorithms used in our use case:

- **Accelerometer**: It produces a continuous stream of X,Y,Z acceleration data at a certain rate (e.g. 100Hz).
- **Low-pass filter**: We use the 'moving average' to remove high-frequency noise to track steps as acceleration peaks at a frequency of max 5 steps per second.
- **Magnitude filter**: We carry out the signal analysis on the magnitude of the acceleration signal as the sensor orientation may change while moving around.
- **Peak filter**: This component extracts maxima and minima in the time domain. A single step is characterized by a particular pattern of these features.
- **Step detector**: It identifies the correct maxima/minima to correctly count the number of steps and to differentiate between standing still, walking and running.

[24] http://couchdb.apache.org/

```
1  curl -X POST --data '{ "timestamp": "long", "x": "double", "y": "double", "z": "double" }'
2      http://localhost/samurai/rest/esper/eventtypes/AccelerometerEvent
3
4  curl -X POST --data '{ "type": "AccelerometerEvent", "timestamp": 1234, "x": 5.0, "y": 1.3, "z": 2.1 }'
5      http://localhost/samurai/rest/esper/event
6
7  curl -X POST --data '{ "rule":"insert into MagnitudeEvent(timestamp, magnitude) select timestamp, Math.sqrt(x*x + y*y + z*z) as
8      magnitude from AccelerometerEvent" }' http://localhost/samurai/rest/esper/statements/magnitude
9
10 curl -X POST --data '{ "rule":"insert into MovingAverageEvent(timestamp, movingaverage) select timestamp, avg(magnitude) as
11     movingaverage from MagnitudeEvent.win:length(10)" }' http://localhost/samurai/rest/esper/statements/movingaverage
12
13 curl -X POST --data '{"url":"http://otherhost/myapp/steps/offer"}' http://localhost/samurai/rest/esper/statements/steps/listener
```

Figure 76: Examples for registering or sending event types and listeners

```
1  ex:Room            a                 owl:Class;
2                     rdfs:subClassOf   geo:Feature .
3
4  ex:LivingRoom      a                 ex:Room;
5                     rdfs:label        "Living Room";
6                     geo:hasGeometry   ex:GeoLivingRoom .
7
8  ex:GeoLivingRoom   a                 sf:Polygon;
9                     geo:asWKT         "POLYGON ((0.00 9.44,3.80 9.44,
10 3.80 8.13,8.00 8.13,8.00 13.90,0.00 13.90,0.00 9.44))"^^sf:wktLiteral .
11
12 ex:activity        a                 owl:DatatypeProperty;
13                    rdfs:domain       ex:Room;
14                    rdfs:range        xsd:string .
15
16 ex:LivingRoom      ex:activity       "Watch TV" .
17 ex:LivingRoom      ex:activity       "Listen to music" .
18 ex:LivingRoom      ex:activity       "Play game" .
19 ex:LivingRoom      ex:activity       "Read newspaper" .
```

Figure 77: Semantic representation of rooms and activities in an apartment

- **High-pass filter**: This component implements a FIR filter to detect sudden and high-frequency changes of the acceleration signal for fall detection.
- **Fall detector**: This component analyzes the signal magnitude area (SMA) of the high-frequency part of the acceleration signal, and identifies a fall if this feature passes a certain threshold.

A more detailed discussion of the algorithms can be found in our Intelligent Environments 2013 award winning work [139]. SAMURAI uses Esper[25] for on-the-fly processing of complex event streams. Esper enables:

- Feature extraction from low-level events (e.g. from accelerometer to steps)
- Publish/subscribe interaction with applications or SAMURAI subsystems.

Esper usually relies on Java POJOs to represent events at compile time. However, in our IoT ecosystem new event types can be created anytime. We therefore expose a RESTful API to dynamically register new event types at runtime. Figure 76 illustrates how to do this with *curl*, a command-line utility commonly found on Linux systems to transfer data from or to a server. Registering the other event types and sending events is done in a similar way as shown in the same figure.

Our architecture offers RESTful APIs to register *statements* and *listeners*. A statement is a continuous query registered with an Esper engine instance that provides results to listeners as new events arrive. In order for applications or subsystems to be notified about the *step* events, we add a listener to this statement as shown in line 13. The example adds a REST callback to *http://otherhost/myapp/steps/offer*, which gets called upon using a HTTP GET request for every step event. The event attributes are appended to the REST callback as url parameters. This way, the *myapp* subscriber is notified about all the event details.

[25] http://esper.codehaus.org

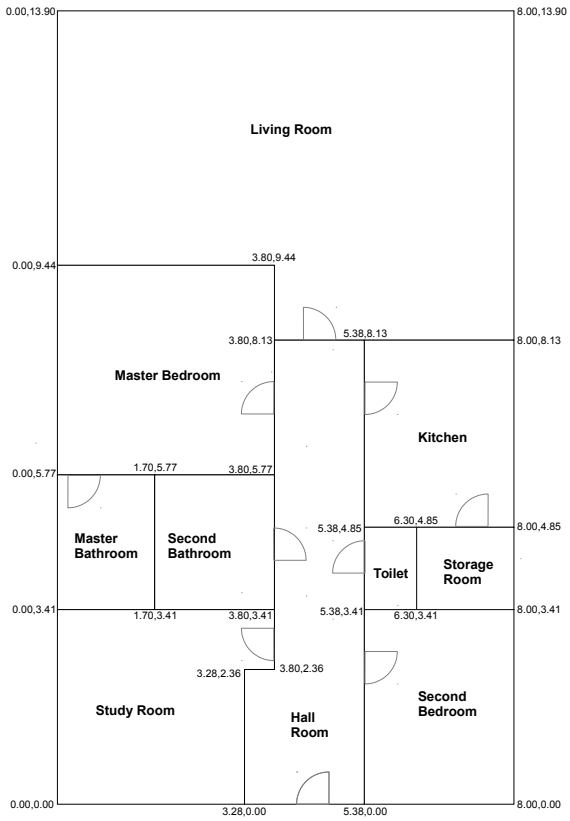


Figure 78: Visualization of the apartment

```
curl -X POST --data '{ "classifier": "weka.classifiers.bayes.NaiveBayesUpdateable" }' http://localhost/samurai/rest/weka/models/m01

curl -X POST --data-binary @m01.arff http://localhost/samurai/rest/weka/models/m01/arff

{ "rule" : "select time, x, y, classify('m01', hour(time), location(x,y), '?') from ..." }
```

Figure 79: Examples for registering, training and defining a custome Esper operator for a classifier

3.3.2.1.2 Semantic database with spatio-temporal reasoning

Beyond matching patterns of events and feature extraction, SAMURAI can also leverage background knowledge stored in a semantic database to increase the meaningfulness of an event. For semantic and spatio-temporal reasoning, SAMURAI uses a GeoSPARQL enabled storage backend (e.g. Parliament[26]). The benefits are manifold:

- Describe the spatial characteristics of different locations in your environment (see Figures 77 and 78).
- Use the W3C SSN ontology to describe the sensors and their position
- Translate positions in coordinates into semantic locations (e.g. [6.0, 10.0] being in the *Living Room*)
- Semantically link locations with activities (e.g. *Watch TV* in a *Living Room*)

The following (simplified) statement demonstrates the integration with Esper (see Figure 80):

This statement translates the *x* and *y* coordinates (e.g. obtained after signal strength triangulation) of incoming events of type *LocationEvent* with the custom *location()* Esper operator offered by SAMURAI. The operator is mapped onto a GeoSPARQL query which retrieves the semantic

[26] http://parliament.semwebcentral.org/

```
1  { "rule" : "select x,y,location(x,y) from LocationEvent" }
```

Figure 80: Custom geo-semantically enhanced event operator *location()*

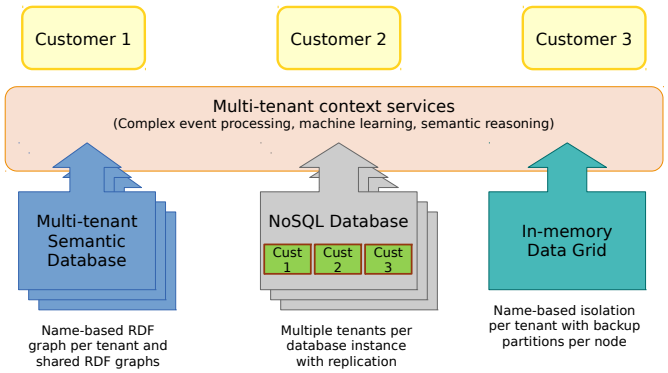


Figure 81: The multi-tenant data architecture of SAMURAI

location (e.g. *location(6,10)* $\rightarrow$ *'Living Room'*). Such higher level concepts are more suitable for classification.

3.3.2.1.3 Learning and mining with classification and clustering

When the relationship between co-occurrent events cannot be established in advance, we need classification and clustering mechanisms to probabilistically infer these dependencies. SAMURAI embeds the Weka machine learning library for this purpose and exposes its key features through RESTful APIs. SAMURAI allows every application to register one or more *models*, with each model having a particular attribute set and classifier. See line 1 in Figure 79. This example registers a model called *m01* using Naive Bayes as an incremental classifier. The attributes used for classification are described in the Attribute-Relation File Format (ARFF) and registered with the following REST API (see line 3). By specifying an appropriate statement and corresponding listener, Esper feeds events as training or test instances into the Weka model. The example in line 5 illustrates how to probabilistically classify activities from the current time (in hours) and location (e.g. *8,'Kitchen'* $\rightarrow$ *'HavingBreakfast'*). This example demonstrates the use of Weka to learn spatio-temporal correlations. The integration with Esper is again with custom Esper operations mapping the core classification and clustering features of Weka. Many technical details of the RESTful APIs and examples could not be elaborated upon in depth in the previous sections due to lack of space. These will be offered on the website *https://butler.cs.kuleuven.be/samurai/* that hosts a running instance of SAMURAI.

3.3.2.1.4 Multi-tenancy and data isolation per tenant

The Software as a Service (SaaS) / Platform as a Service (PaaS) and multi-tenancy paradigms of cloud computing are often positioned as practical approaches to offer the above functionality to a variety of customers with different needs. However, there are several concerns from a data management standpoint that make this endeavour not straightforward for context management.

- **Data isolation**: Separate each customer's data and context to reduce the risk of exposing the wrong data
- **Performance**: Manage the customer's data that allows for collocation or isolation based on service level agreements or performance

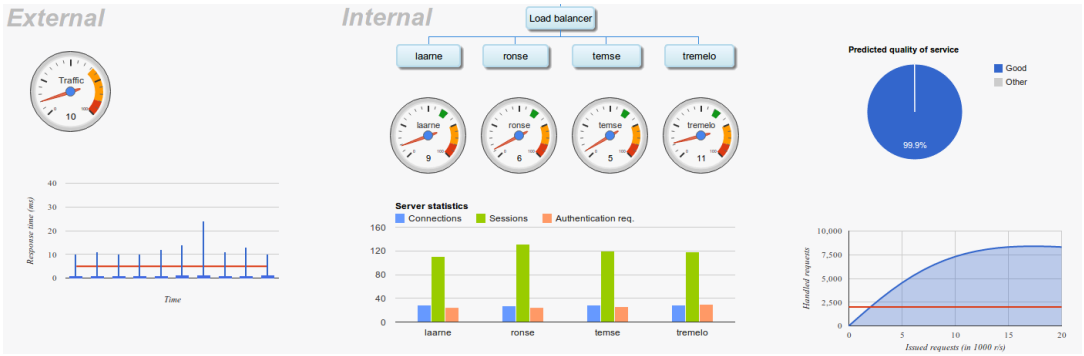


Figure 82: Monitoring dashboard

- **Management**: Add new customers in a flexible way and allow for cross-customer context support

In the SAMURAI architecture, data is provisioned in the *In-memory Data Grid*, the *Semantic Database*, the *NoSQL Store* and the *Event Queues* hosted on top of the in-memory data grid. Using dedicated servers per tenant offers good isolation from a security perspective, but wastes resources when the context management services are not used. Enabling multiple tenants to share a database and isolating tenant data by using separate tables for each tenant helps to reduce per tenant costs.

SAMURAI achieves multi-tenancy by first having each customer authenticate. Access based on their identity to the persistent and in-memory data is then achieved for the 3 aforementioned data subsystems as illustrated in Figure 81. To simplify isolation per tenant in the semantic database, we instrument SPARQL queries so that the tenant only accesses its own SPARQL graph. The NoSQL store is set up in replication mode to ensure high availability. The in-memory data grid acts as a distributed hash map and is the most frequently used and highest performant data access component.

We have evaluated several architectural tactics to ensure the scalability of the SAMURAI framework in a distributed multi-tenant deployment. We carried out experiments to analyze the performance impact of pure REST request handling and the impact of multi-tenancy for data isolation.

For the scenario used in the experimental evaluation of our architecture, the results show little overhead to support multi-tenancy, with near-linear scalability and flexible elasticity for deployment schemes with in-memory data partitioning per tenant. A limitation of our current experiments was that the load generation for simulation (external) and the multi-tenant distributed deployment of SAMURAI (internal) were all linked to the same local network. As our data tier consists of multiple components (e.g. a semantic database, an in-memory data grid and a NoSQL store) each with their own replication and partitioning techniques, we believe better results can be achieved to isolate further the internal and external parts of our setup, not only for the computational part but also for the network part.

SAMURAI is fully integrated with BUTLER platform through other SmartServers such as Localisation SmartServer, Context Manager and Trust Manager.

3.3.2.2 Situational Studio Tool for Design Time Modeling

As noted in [136], most existing works focus on a limited set of user context recognition and validate the accuracy of their approach with the implicit assumption that the activity of interest is taking place.

We explicitly model all possible user activities and consider situations where context models could lead to false positives. For example, the fall detection with the barometric pressure might detect a

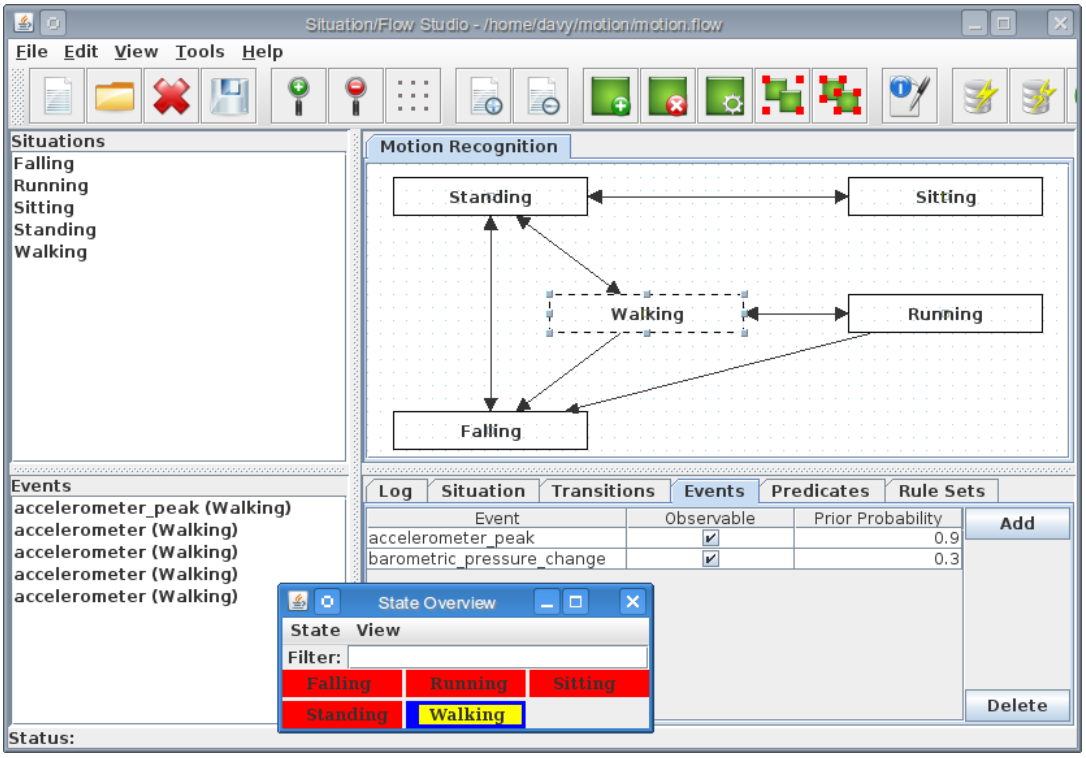


Figure 83: Situation Studio - a tool to model the activity transitions at design time [202]

false positive when going down the stairs, because with each step the accelerometer triggers the pressure sensor and the latter detects a lower altitude. However, one can also fall down the stairs. All of these interrelationships between different kinds of contexts and activities and corresponding recognition techniques are modelled with our Situation Studio [201]. This tool (see Fig. 83) borrows concepts from work on modelling languages, and represents situations that evolve from one to the next through constrained sequential and parallel transitions. For each of them, we identity the contextual boundaries, the likelihood of activities of interest, the relevant contextual events, and the recognition schemes available.

3.3.3 Contextual Networking

3.3.3.1 Transfer Learning Techniques as a Key Enabler for Contextual Networking of Macro-scale Behavior Recognition

With the availability of a wide variety of embedded sensors and powerful processors, physical activity recognition with off-the-shelf smart phones has become main stream. Most popular activity recognition applications are in the health-care domain where the tri-axial accelerometer embedded in a mobile phone is used to recognize the activities of a user (such as standing, walking) and monitor their physical activity levels and calorie expenditures. Although recent works in learning and classification have considerably improved the recognition accuracy for activities of daily living, most applications heavily rely on user or device specific inputs.

In our work [203], we studied the influence of various device specific parameters (e.g., sensitivity and sampling rate of the accelerometer) and the user specific variabilities in human activities (and behavior), on the performance of activity recognition algorithms. Also, investigated the transfer learning techniques as a key enabler for contextual networking of macro-scale behaviour recognition with entities linked in a contextual network characterized by similar features.

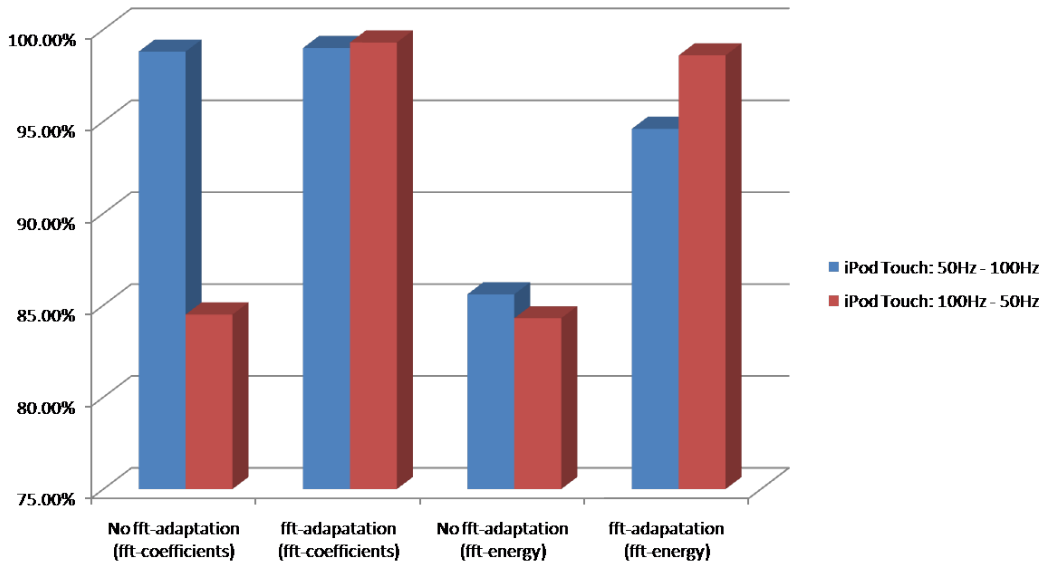


Figure 84: Recognition rate for physical activities with and without transfer learning techniques when the context models are trained and tested under different sampling rates

A middleware abstraction to handle the discrepancies in the varying characteristics of the devices, parameters of feature extraction algorithms and users is implemented using Weka library for Android in order to mitigate the influence of device specific parameter variations on the accuracy of user activities which has resulted in reduced training periods. Fig. 84 show how the recognition rate can be adversely affected when the sampling rates are different at inference compared to training period. Also, it showcase how the transfer learning techniques for FFT-adaptation can significantly improve the performance.

In the near future, we will investigate how a trained model for one individual can be applied for another individual, to reduce the overall training effort. For clusters of similar models, we will develop techniques to elicit the corresponding similarities in the user details and associated context. These similarity requirements will be used to ensure that transfer learning can be applied without adversely affecting the activity classification accuracy for any individual user within the cluster.

It is integrated at the client application for monitoring physical activities of the user.

3.3.3.2 Context-based Optimized Communications to Gateways in Wireless Networks of Smart Objects

3.3.3.2.1 CLUBCROM Architecture

In the specific context of Smart Cities, preliminary studies have been carried out regarding the definition of a new architecture suitable to a context-aware adaptation of MAC/networking mechanisms in order to establish adaptive and optimized wireless communications between mobile sensing nodes or end-devices (i.e. Smart Objects) and Gateways or WSN sinks over large distances. Existing networks have too many devices spread over a huge geographical area and use one single coordinator or a mesh topology. The use of only one coordinator could lead into a single point of failure as well as traffic congestion or bottleneck effect on the vicinity of the coordinator. In order to allocate communication resources to all the devices the coordinator must increase the size of the superframe, leading to increased delays for new nodes while joining the network and also increased end-to-end delays to transmit data packets. A more distributed approach with more than one coordinator and further clustering could provide improvements specifically into smart cities

scenarios. The IEEE 802.15.4 standard allows the interoperation of different Personal Area Networks (PANs) but under the supervision of a primary coordinator, which was covered by Zigbee. An extension to Zigbee is proposed in [204] with a multi-channel scheme and bridge functionalities for devices interconnecting surrounding coordinators to the primary coordinator. Alternatively, our solution will follow a locally centralized and globally distributed approach. In other words, inside each cluster the communication is centralized on the coordinator, reducing the number of collisions and easing reconfigurability, while between clusters the communication is distributed, allowing a more natural spatial reuse of the temporal resources. In other words our approach chooses to locally centralize the communications with the establishment and the management of clusters and to globally distribute the management of the large scale network with an optimized data routing to Gateways through dynamic bridges. Accordingly, a new protocol called CLUBCROM (CLUster-Based CROss-layer Multi-channel protocol) has been designed. CLUBCROM is suitable for large, dense and heterogeneous networks. The protocol is based on the IEEE 802.15.4 standard with some extensions to diminish its limitations in the context of Smart Cities: Cluster-based since it organizes itself under a cluster tree scheme, Cross-layer due to the fact it merges both MAC and routing layers to enhance its capabilities and Multi-channel in a sense it uses different channels to enable communications in or between two or even more numerous distinct clusters. Figure 85 illustrates different clusters using distinct channels with the bold lines representing the communication bridges between clusters.

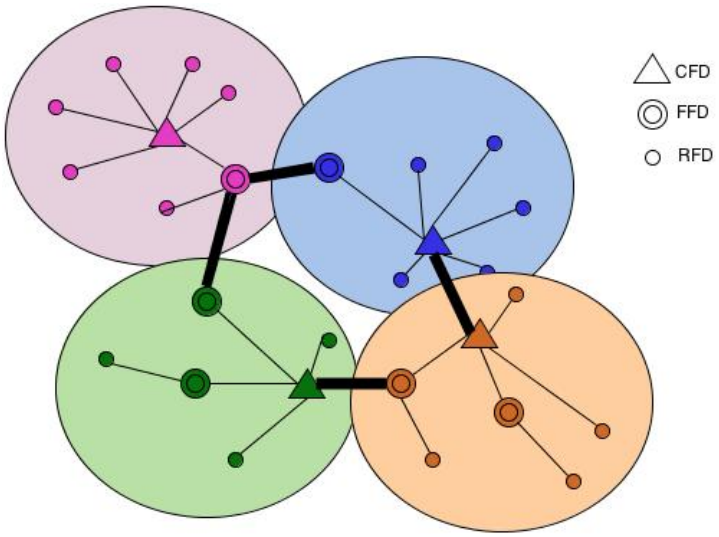


Figure 85: CLUBCROM Network Topology.

In the proposed architecture, the presence of different clusters under different coordinators is possible. It provides the interoperation and coexistence of several clusters thanks to the use of the different IEEE 802.15.4 channels. Hence interferences among clusters do not significantly increase while the protocol remains scalable. Each cluster could choose its protocols depending on its application requirements in order to provide an energy efficient solution. Our proposed CLUBCROM architecture provides interoperability among several clusters, regardless of their MAC and Routing schemes. This architecture is also compatible with inter-gateway connections through LTE. An extension to this study could be to compare the efficiency of horizontal (i.e. multi-hop schemes) and vertical (i.e. using LTE) communications between clusters. The architecture specifies three different device types, and four different possible roles. The devices can be either Coordinator-Function Device (CFD), Full-Function Device (FFD) or Reduced-Function Device (RFD) according to its power, computational and transmission capabilities. The roles they can play are as follows:

- Coordinator: Only one device (i.e. CFD) can fulfill this role per cluster. It has a similar role as that of a IEEE 802.15.4 coordinator. It is responsible for association management, beacon

scheduling and periods scheduling.

- Router: Nodes with this role have the ability to relay beacons and association messages to the CFD. During the beaconing period, it can relay messages in order to increase the cluster coverage. Normal devices join this node in a similar manner as when associating to the coordinator.
- Normal: The node with this role saves its energy and is responsible only for sending and receiving its data according to the selected MAC protocol.
- Bridge: Bridges enable the clusters interoperation and can communicate with other clusters through dedicated channels. FFDs can request bridge role after receiving an advertisement (ADV) from a device located in another cluster during the optimized scan procedure of the advertisement channel.

Different devices can have distinct roles in the protocol. For instance, FFDs can act as a router, a bridge or a normal sensor. CFDs can be coordinator and also act as a bridge. RFDs can only have one role, normal sensor. Moreover, CLUBCROM exploits the different channels available on the IEEE 802.15.4 standard, to increase coverage and to decrease interference and power-consumption while remaining scalable. This cross-layer MAC and routing protocol is characterized by a divide and conquer approach and uses a cluster-tree topology and 3 types of communications:

- Advertisement channel (ADVCH): Similarly to Bluetooth [205] and [206], the Advertisement channel is used to detect the presence of different clusters around the device. Only devices with higher energy capabilities (CFDs and FFDs) are allowed to communicate on this channel.
- Inter-cluster channel (INTERCH): The inter-cluster channel is used by bridges to transmit data from one cluster to another. Only nodes assigned with bridge role can use this channel. These devices must previously detect, through the Advertisement channel, other clusters' bridges in range. The bold line on Figure 2 illustrates the communication link on the Inter-cluster channel.
- Intra-cluster channel: All the remaining channels can be used for intra-cluster communications. In the Intra-cluster channel, only nodes belonging to the cluster can exchange messages and they shall not interfere with neighboring clusters communications.

Three different MAC approaches have been defined respectively for the Intra-cluster, Advertisement and Inter-cluster communication schemes. As CLUBCROM uses a cross-layer approach, the routing scheme must be specially designed according to the proposed MAC layer. The network layer acts differently depending on the destination and the channel it will route information. A centralized intra-cluster routing managing local changes and a distributed high level inter-cluster routing can be defined respectively on the device addresses and the cluster addresses. The intra-cluster routing is based on on-demand and scheduling tree schemes. The inter-cluster routing is defined to deliver a packet when the final destination is in another cluster. The network layer can be seen as a combination of two protocols: a local and a global one. The previous intra-cluster routing protocol defines the network layer in a local basis. This routing is used to link a coordinator to its bridge or to link two bridges inside the same cluster. On top of each local protocol, a high level inter-cluster routing protocol, based on AODV [207], is used to perform the routing between clusters. This protocol does not use the node address but the cluster address of the destination. Thus, each cluster can be considered as an entity in a higher level network and the routing extends multi-hop communications to multi-cluster communications.

3.3.3.2.2 Mobility Management

A particular attention has been paid more recently to mobility, which traditionally leads to a deterioration of the link quality, to frequent route changes and to increased delays for establishing a new route. Therefore, the support of mobility gains considerable importance while trying to extend the

protocol functionalities designed for smart objects into non-static scenarios. In this context, the previously defined cluster based architecture eases the management of the mobility by design. In fact, CLUBCROM (CLUster-Based CROss-layer Multi-channel protocol) can manage mobility at two distinct levels: locally by the coordinator inside the cluster without diffusing any information outside and globally between clusters through handover mechanisms.

Proposed scheme

The simplified scheme proposed for mobility management is depicted on Figure 4. After acquiring measurements for each input parameter, the mobility module initiates a filtering and prediction procedure. This procedure can rely on e.g., a Kalman Filter [208] for smoothing the input and eliminating possible variations on the input measurements caused by fast fading or shadowing. Within such filtering procedures, one can exploit the prediction step to determine future parameter values in the reasonably short term. The filtered or predicted value is subsequently fed into the Decision block of the proposed scheme. In this part, each node decides whether it stays connected with the same router, or if it performs a handover to some different cluster and/or router. The last part, after the decision is taken, corresponds to the execution it-self (i.e. the execution of what have been previously decided). The Cross-Layer Module is responsible for this Action part.

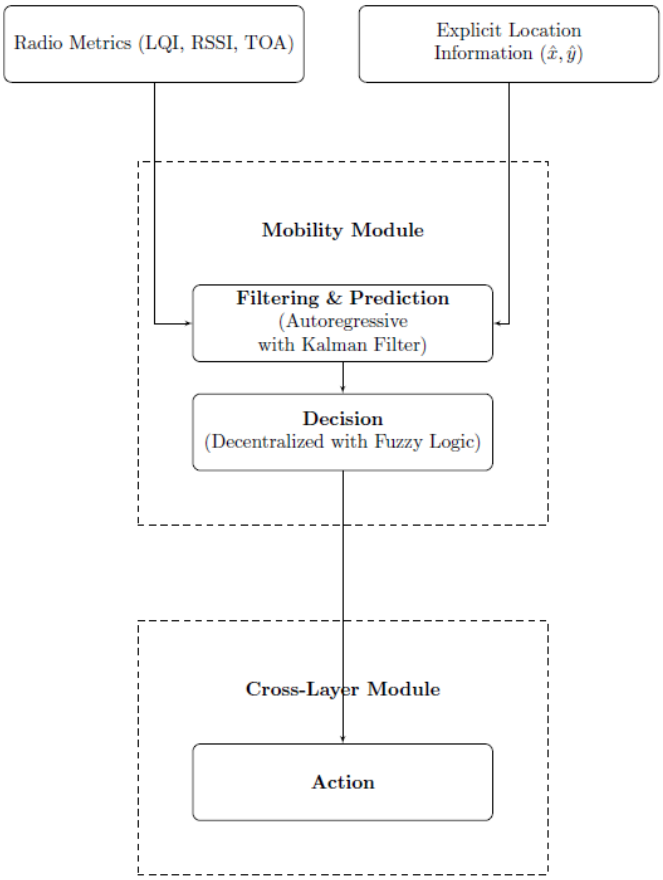


Figure 86: Proposed Mobility Management Scheme

Filtering and Prediction

The use of Kalman Filters for the filtering and prediction of radio-based metrics is a well-known area [209]. We are more particularly interested here in mitigating RSSI abrupt fluctuations and in predicting its future average values. RSSI fluctuations due to signal noise significantly affect both stability and efficiency [210]. Indeed RSSI over/under-estimation may trigger unnecessary predictions and RSSI-based handovers, thus lowering efficiency. Similarly to [210], our filter works by

minimizing process noise through a two phase algorithm: first, a predictor performs next RSSI estimation and then, a corrector improves RSSI estimation by exploiting current RSSI measurements. In practice, the process noise covariance and measurement noise covariance matrices might adaptively change at each time step or measurement. However here we assume they are constant as a first step of investigation for simplification.

Handover decision

In order to avoid overhead in data exchanges, the handover mechanism is implemented in each sensor node, i.e., each node detects locally the necessity of performing or not a handover to either a router inside the same cluster or to some other cluster. Four different algorithms for handover decision have been implemented. The different approaches to the mobility management have been created. The first, a classical one, is better known as hysteresis curve. It uses the RSSI value and tries to fit it under a hysteresis curve for each base station or access point (AP). It is an elementary approach, used for benchmark. The others use fuzzy logic for assessing the operating conditions and environment and then, from the measurements acquired and calculated, infer their handover decision based on a fuzzy logic system. All implemented solutions are briefly explained hereafter.

Hysteresis

For the classical solution explained on [210], the authors propose two variants of the Proactive handover policy, namely, Hard Proactive (HP) and Soft Proactive (SP). On the one hand, HP strategies trigger a handover any time the RSSI of a visible AP is larger than the RSSI of the currently associated AP plus a Hysteresis Handover Threshold (HHT). HHT is introduced mainly to prevent heavy bouncing effects. On the other hand, SP strategies are "less proactive" in the sense that they trigger handover only if i) the HP condition applies (there is an AP with RSSI greater than current AP RSSI plus HHT), and ii) the current AP RSSI is lower than a Fixed Handover Threshold (FHT).

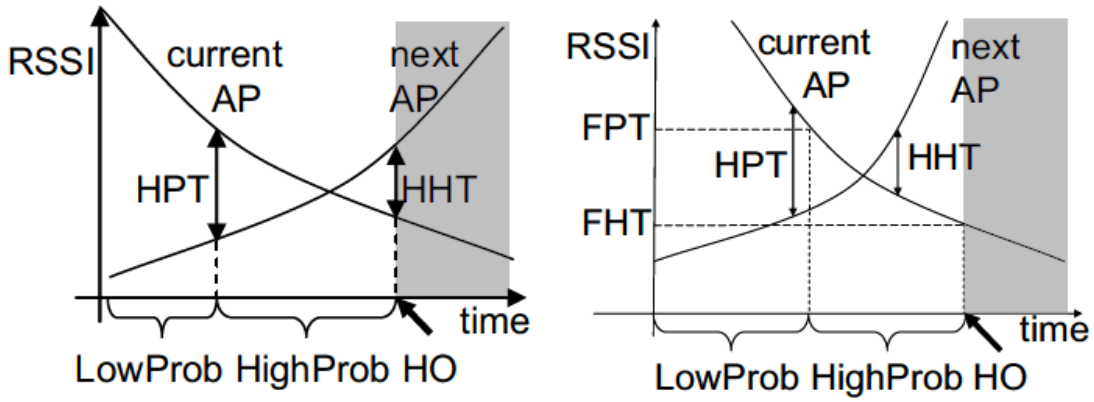


Figure 87: The different RSSI-based hysteresis handover variants considered for benchmark: HP (left) and SP (right). [210]

The HP-variant is in the state LowProb if the filtered value for the current AP RSSI is larger than the filtered RSSI values for any visible AP plus a Hysteresis Prediction Threshold (HPT) and in the state HighProb, otherwise. The SP-variant of the Prob module can occupy the following states: LowProb, if the filtered RSSI value for the current AP is larger than either a Fixed Prediction Threshold (FPT) or the filtered RSSI value for any visible AP plus HPT and HighProb, otherwise. Figure 87 represents the filtered RSSI values for current and next APs, in proximity of a HP (left) and SP (right) handover. A wireless client, moving from the origin AP locality to the destination AP one, is first associated with the origin AP (white background), then with the destination AP (grey background). In our simulations, the SP handover procedure described above, has been used for comparison purpose only.

Fuzzy Quantitative Decision Algorithm (FQDA)

In [211], the Fuzzy Quantitative Decision Algorithm (FQDA) is implemented to quantitatively evaluate the input parameters of candidate networks. The FQDA approach as well as its advantages (i.e. in comparison with a traditional fuzzy handoff algorithm) are summarized as follows:

- FQDA is able to get the QDV of a certain candidate network. QDV tells the probability that the certain network becomes the target one to handoff. There is no need to establish a database to store the enumerative handoff rule bases, which may occupy a large memory. For example, if 5 fuzzy sets are established for each of the 3 input parameters, the number of cases in rule bases will be $5^3 = 125$.
- The final handoff decision is solely based on the comparison of QDVs of the candidate networks. There is no need to search within the rule bases, which may take significantly longer time.

Moreover, FQDA can also provide more scalability in the case of increasing/decreasing the number of metrics to be taken into account. It can easily integrate new parameters without the need of rebuilding a big portion of the algorithm, which does not occur with traditional fuzzy-rules.

The membership function of each parameter is defined a priori. After this definition, the real time measurement of the metric in a candidate network is fed into these membership functions and then they are classified into one (RSS=Q in Figure 88) or two (RSS=P in Figure 88) of the fuzzy sets resulting in corresponding membership degrees. For example, when the input value is RSS=P, the membership degree of P is [0, 0, 0, 0.8, 0.3].

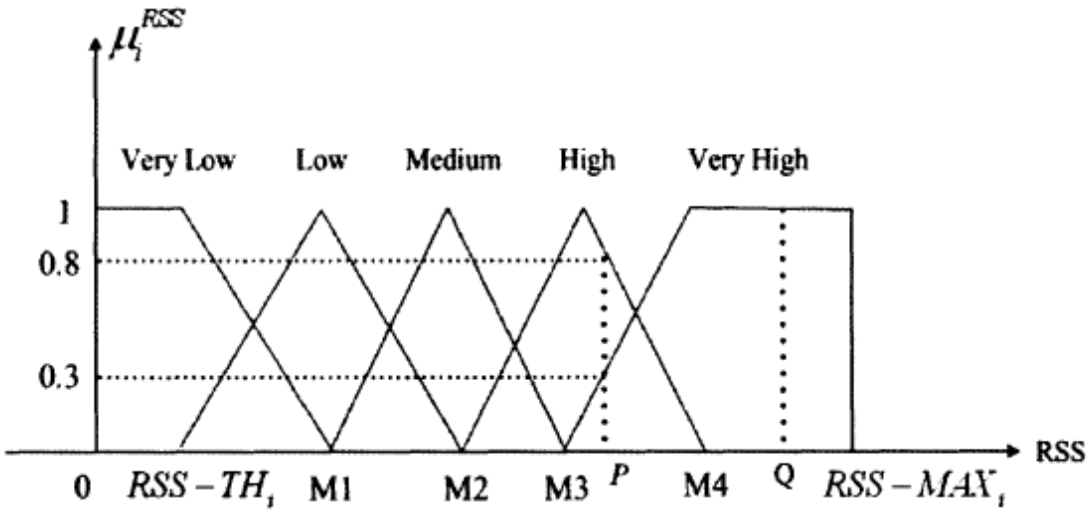


Figure 88: Membership function of RSS [211]

Moreover, in order to quantitatively evaluate the input factors, specific quantitative evaluation values have to be assigned to each fuzzy set [211]. These values can be seen as the desirability of each fuzzy set. In the case of RSS for instance, the higher the value, the more desirable it is for the node. Therefore, we can assign $[Q_{VL}, Q_L, Q_M, Q_H, Q_{VH}] = [0, 0.25, 0.5, 0.75, 1]$.

Based on the membership degrees and quantitative evaluation of each fuzzy set, the quantitative evaluation (QEV) of each input parameters k for a network i is defined as:

$$QEV_i^k = [Q_{VL}^k, Q_L^k, Q_M^k, Q_H^k, Q_{VH}^k].[\mu_{VL}^k, \mu_L^k, \mu_M^k, \mu_H^k, \mu_{VH}^k]^T$$

The final QDV_i of each network is obtained by integrating all the QEVs. In order to optimize the handoff decision, each metric weight for a candidate network should be adjusted to reflect this metric priority relatively to the other attributes in different candidate networks. Thus, weights should

dynamically reflect the importance and relationships of the continuously changing QEV under unpredictable wireless environments, and should magnify the dominant-difference among candidate networks. The modification of the weight of each parameter k, as proposed in [211], is defined as:

$$\phi^k = e^{\overline{-QEV^k}+\sigma^k}$$

$$w^k = \frac{\phi^k}{\sum\limits_k \phi^k}$$

Finally, the QDV of each candidate i is calculated as:

$$QDV_i = \sum_k w^k QEV_i^k$$

The candidate network with the largest QDV can be selected and will be the final target network to handoff [211]. The main drawback of such an approach is the equal comparison of different aspects and features of the network, giving the same importance to heterogeneous parameters of different nature, e.g., RSSI and bandwidth.

Weighted FQDA

In order to circumvent the main drawback of FQDA, a different approach is presented, introducing the use of a weighted system to differentiate the relevance of considered parameters. This weighting system can be defined a priori and once for all (e.g. according to the application needs, if relatively time-invariant) or automatically/dynamically.

Then, each parameter would be weighted according to the preference and/or importance given by the application. In this sense, the FQDA remains the same up to the point where QDV is calculated. In the Weighted FQDA, the QDV of each candidate i can be obtained as:

$$QDV_i = \sum_k \alpha^k w^k QEV_i^k$$

where α^k represents the weight assigned to the parameter k.

Group-Weighted FQDA

Another proposal briefly detailed hereafter is to cluster the parameters in homogeneous groups so that they could be comparable. After, we apply FQDA in these groups and, using weight for each of the different groups, the system generates a QDV for each AP. Finally, the one with the highest QDV is selected. As an example, we can select 4 different groups:

- **Group Link Quality** - Related to the RSSI value and to the gradient of the RSSI (i.e. RSSI dynamics/trends over time), thus reflecting somehow the link quality;
- **Group Position** - Related to the relative distance and also the relative speed (gradient of distance) of the mobile node in relation to the AP, thus reflecting explicit mobility aspects;
- **Group Energy** - Related to the energy that would be spent to remain connected or, eventually disconnect from actual AP and connect with other, thus anticipating on the impact/cost of handover in terms of system autonomy.
- **Group Network Distribution** - Related to the distribution of the network among different AP, a more distribute network requires nodes to be connected not with only one AP but with several ones.

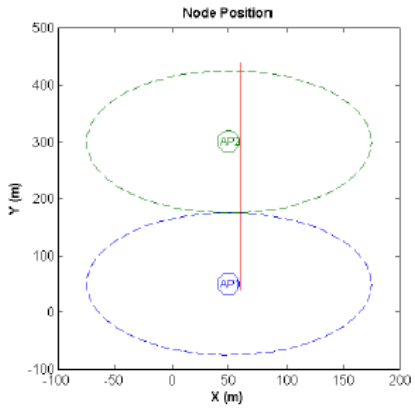
Inside each group, the importance (or weight) of each paramenter can be defined in comparison to other metric members of the group and then, each group would be weighted according to the preference and/or importance given by the application. This approach, allows the creation of profiles (i.e. set of group weights) specificly attributited according to the application needs. The QDV of each candidate AP i, group of metrics g and metric k_g belonging to group g is then defined as follows:

$$QDV_i = \sum_g \beta^g \sum_{k_g} \alpha^{k_g} w^{k_g} QEV_i^{k_g}$$

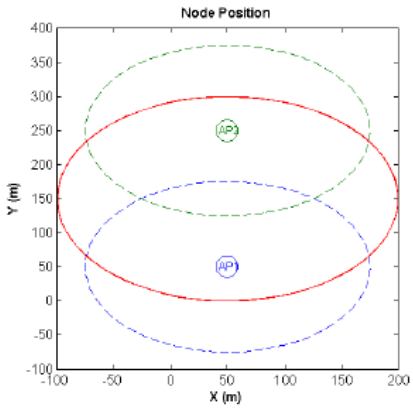
where α^{k_g} represents the weight assigned to the metric k of group g and β^g represents the weight assigned to the group g. The AP with the highest QDV is then selected.

Simulation

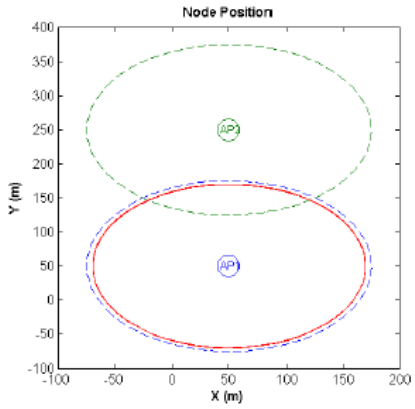
Initially, we simulated in Matlab the mobility of a node under three different activity patterns. The first is a simple walk from one AP to the other. The second is represented by a circle around one AP or even in the middle area between different APs. The last is the well-known Random Waypoint mobility pattern [212].



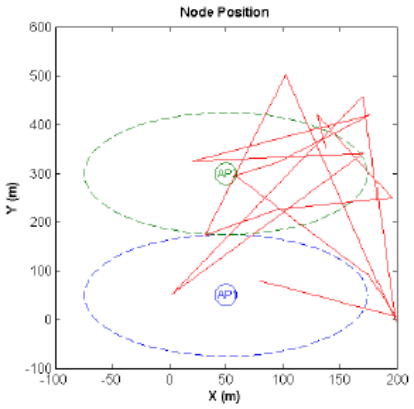
(a) Walk from AP1 to AP2



(b) Circular between APs



(c) Circular around one AP

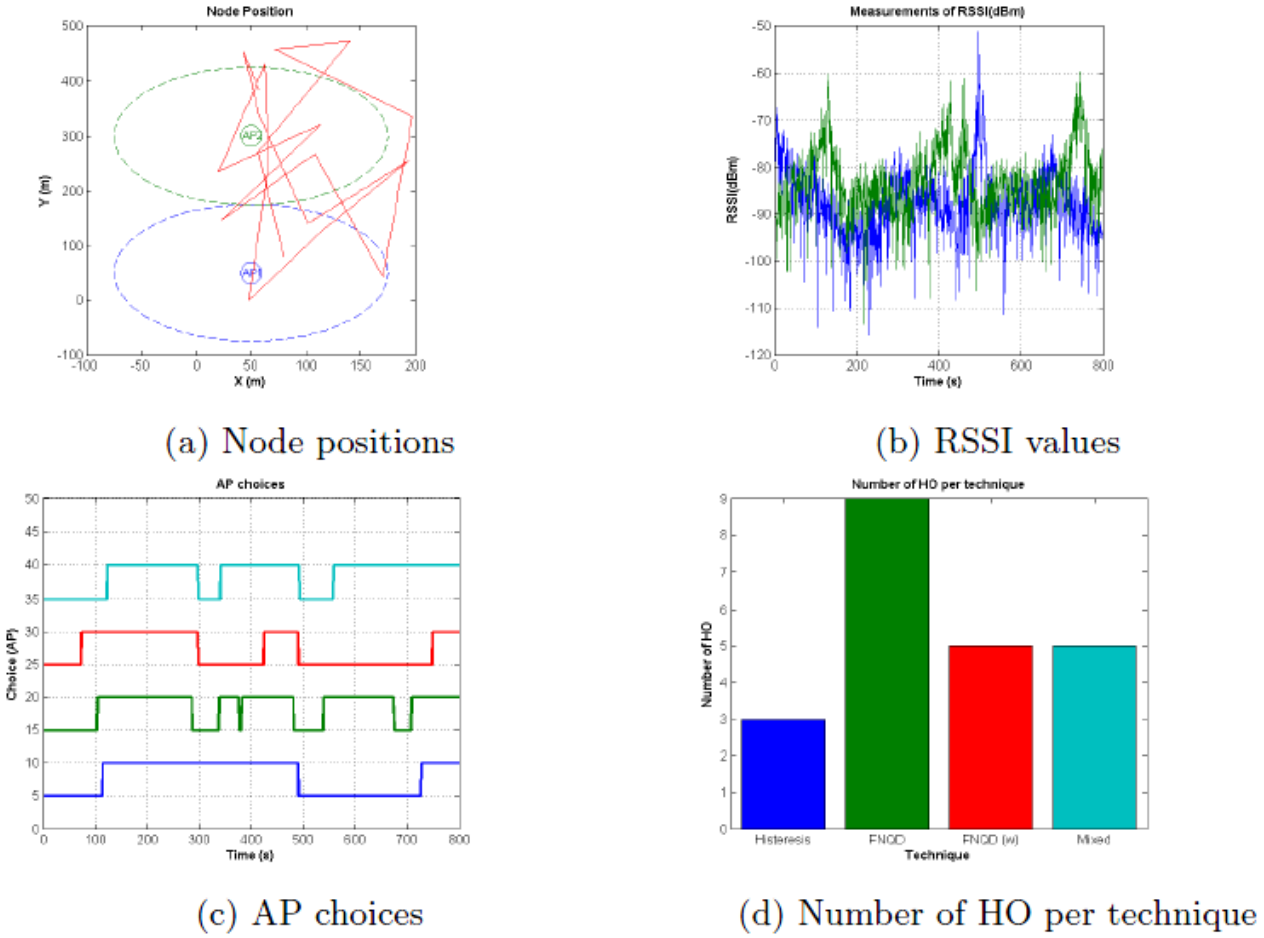


(d) Random Waypoint

Figure 89: Mobility Patterns considered for the performance assessment of handover decision rules

In this first scenario, in order to assess the performance uniquely of the decision stage, regardless of the quality of input values, we used only real instantaneous measurements to feed the decision module. Thus, the RSSI values obtained are subject to large fluctuations due to fading and

shadowing. Simulations were carried out for the four decision modules under the distinct mobility scenarios introduced above. As already described, the first classical scheme, better known as hysteresis curve, uses the RSSI value and tries to fit it under a hysteresis curve for each AP and, after selecting the best candidate, it performs (or not) the handover. The others use fuzzy logic to better capture the contextual operating conditions and, thus, based on the measurements acquired and calculated, infer handover decisions based on a fuzzy logic system. The Fuzzy Quantitative Decision Algorithm (FQDA) is implemented to quantitatively evaluate the input parameters of candidate networks. As said before, each candidate network is endowed with a QDV value, which tells the probability that the certain network becomes the target one to handoff, and the AP with highest QDV is the selected one. Finally, we consider using a weighted system to differentiate the relevance between input parameters. Thus, each parameter is weighted while calculating the QDV, according to the preference and/or importance given by the application. The obtained preliminary results are depicted on Figures 90, 91 and 92..

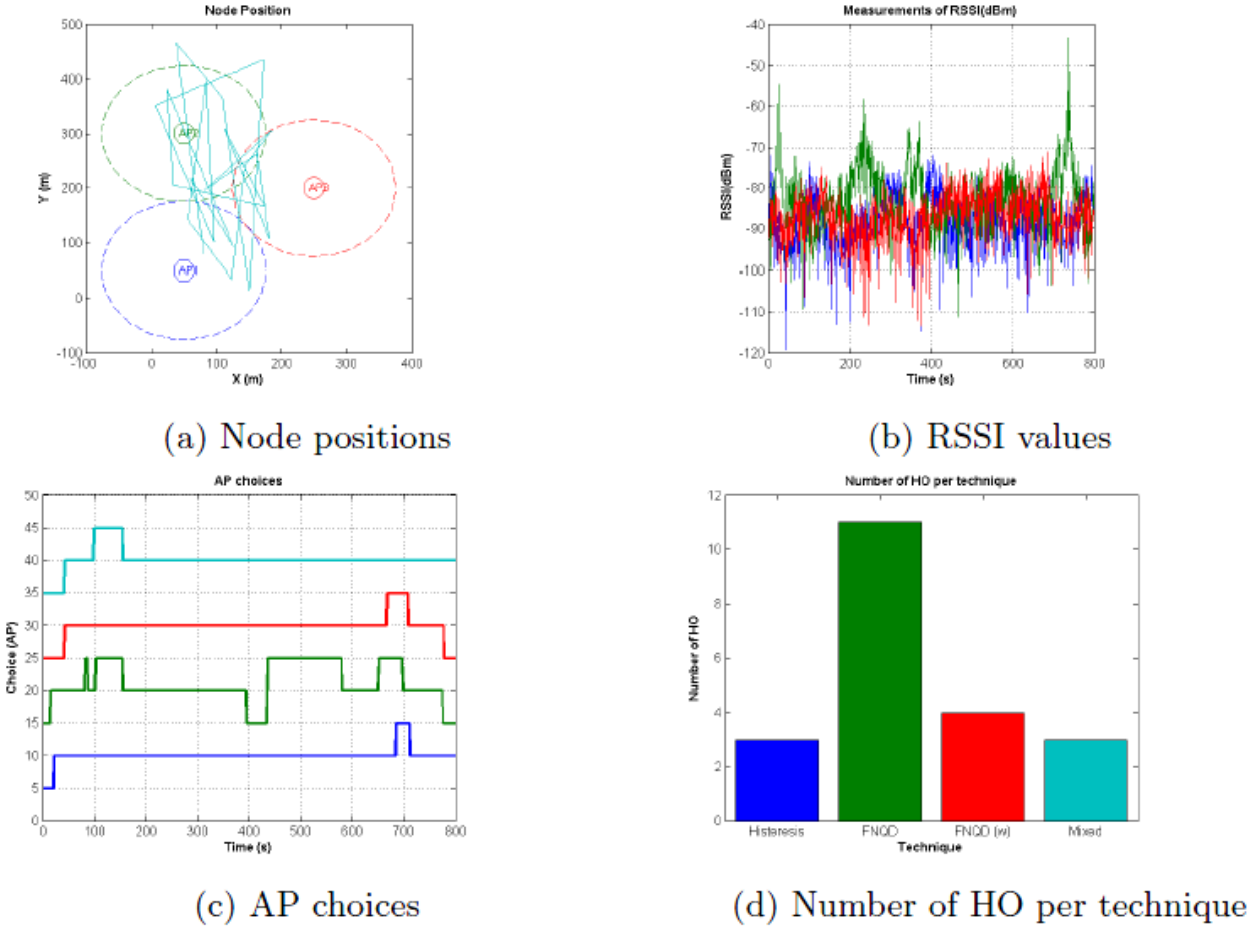


(a) Node positions (b) RSSI values

(c) AP choices (d) Number of HO per technique

Figure 90: Simulation handover results with walk between AP1 and AP2

These preliminary results tend to show that the new proposed fuzzy systems are globally able to reduce the number of handovers performed in several mobility situations in comparison with the other conventional techniques. However, It can still happen that the node remains more time connected to an AP when he should proactively and faster change to the other one to improve its QoS. Accordingly, we have implemented in an event-driven and packet-oriented simulation tool, named WSNet, this mobility management module, allowing to adapt in real-time the intra-cluster communications depending on the detected local and temporal context patterns (i.e. in terms of nodes activity, mobility, energy autonomy, etc.) and to design handover mechanism ensuring the



(a) Node positions (b) RSSI values

(c) AP choices (d) Number of HO per technique

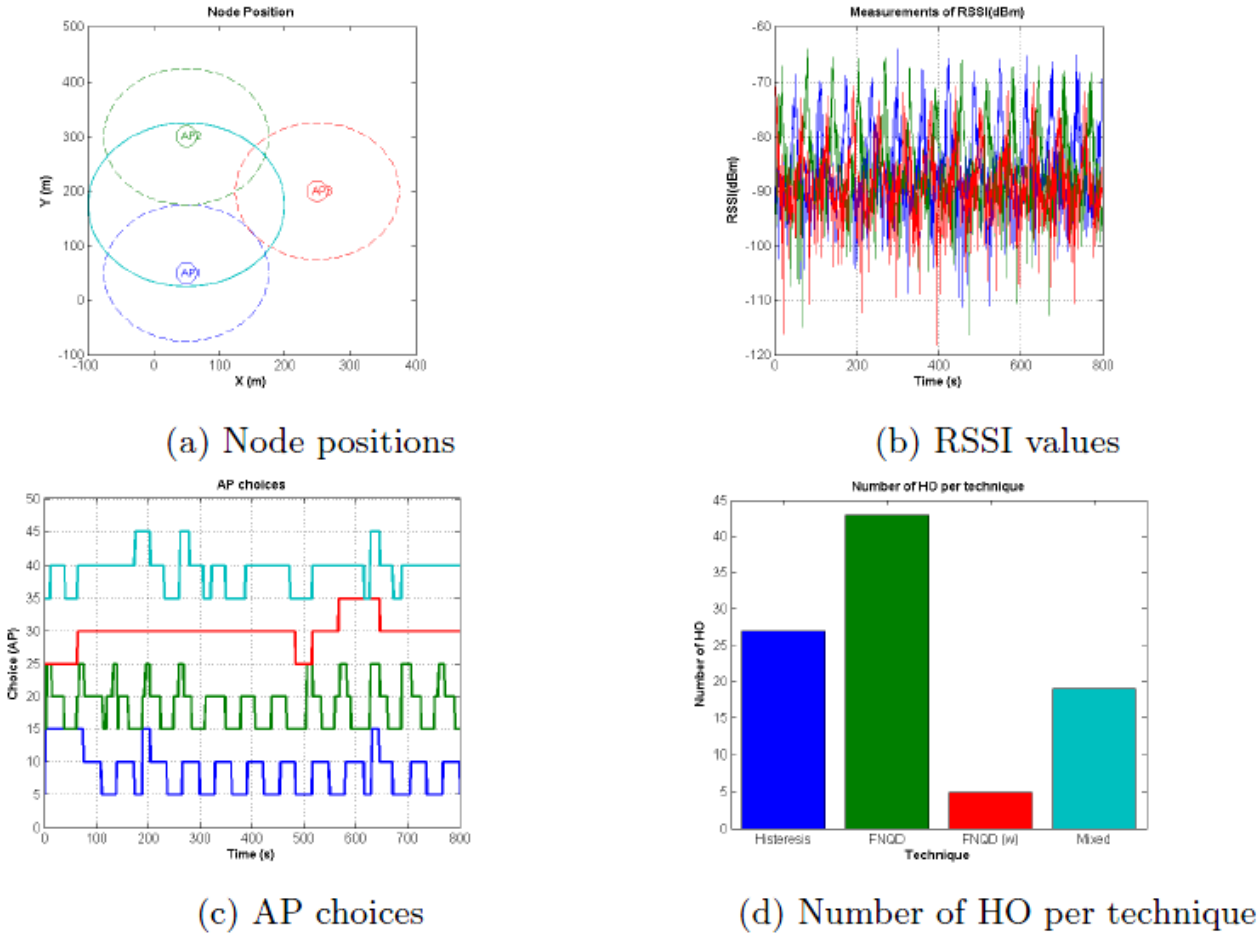
Figure 91: Simulation results with Random Waypoint mobility

stability of the whole network.

The scenario for this simulation, as depicted on Figure 93, is an area of 120x100m, divided in rooms of 20x20m. The propagation model inside the same room follow a LoS and NLoS for adjacent rooms, with all remaining rooms following a more degrading propagation model of Non Line of Sight Square (NLoS2). Each of these propagation models are breifly explained hereafter.

- **LoS** - Path loss exponent equal to 2, shadowing standard deviation equals to 0.5dB and Rician k factor equal to 9;
- **NLoS** - Path loss exponent equal to 3, shadowing standard deviation equals to 3dB and Rician k factor equal to 5;
- **NLoS2** - Path loss exponent equal to 3.3, shadowing standard deviation equals to 6dB and Rician k factor equal to 1;

Simulations were carried out for three different configurations: one with no decision module and two with decision modules implementing both hysteresis and group-weighted FQDA. Regarding the later, it is valid to stress that, during this simulation, only groups of metrics regarding link quality and network distribution have been used with weights of 0.8 and 0.2 respectively. The mobility profile used by mobile nodes is the random waypoint mobility pattern [212]. CFDs and FFDs acts as fixed AP and are depicted in Figure 93 as circles. Differently than previously, we used a sliding-window to smooth input RSSI values and prevent large fluctuations due to fading and shadowing. Results of simulations are presented on Figure 94.



(a) Node positions (b) RSSI values

(c) AP choices (d) Number of HO per technique

Figure 92: Simulation results with circular mobility between APs

Figure 94 shows that the mobility management decision provides a significant improvement on the packet delivery ratio while maintaining the average end-to-end delay relatively low in comparison with the standard (no decision module). Theses results occur once nodes remain connected to the network for longer periods. It is also valid to observe the relatively proximity between results of both hysteresis and group-weighted FQDA. These results are mainly linked to two main reasons: the simulation scenario and the weight selection. The simulation scenario includes only one cluster, therefore, inter-cluster handover is not present. This leads hysteresis performance close to the group-weighted FQDA as the RSSI metric can sufficiently provide a capable handover decision and the others metric does not significantly improve performance. Results on scenarios with inter-cluster are expected to present an inferior performance in comparison with group-weighted FQDA. The second reason is due to the fact that weights on each of the metrics of the group-weighted FQDA have been choosen manually and are probably not the optimal one.

In conclusion, we have defined a context-aware protocol adapting automatically and dynamically to its environment thanks to the combination of a multitude of metrics. A mobility and multi-context management module has been implemented in an event-driven and packet-oriented simulation tool named WSNet. This module adapts in real-time the intra-cluster communications depending on the detected local and temporal context patterns (i.e. in terms of nodes activity, mobility, energy autonomy, traffic, etc.) and provides a handover mechanism ensuring the stability of the whole network. Results provided by the present work show the importance of managing mobility on WSN. It can reduce the number of disconnections and thus improve the performance of the network. More research regarding the inter-cluster handover and the optimal selection of weights are expected in future works.

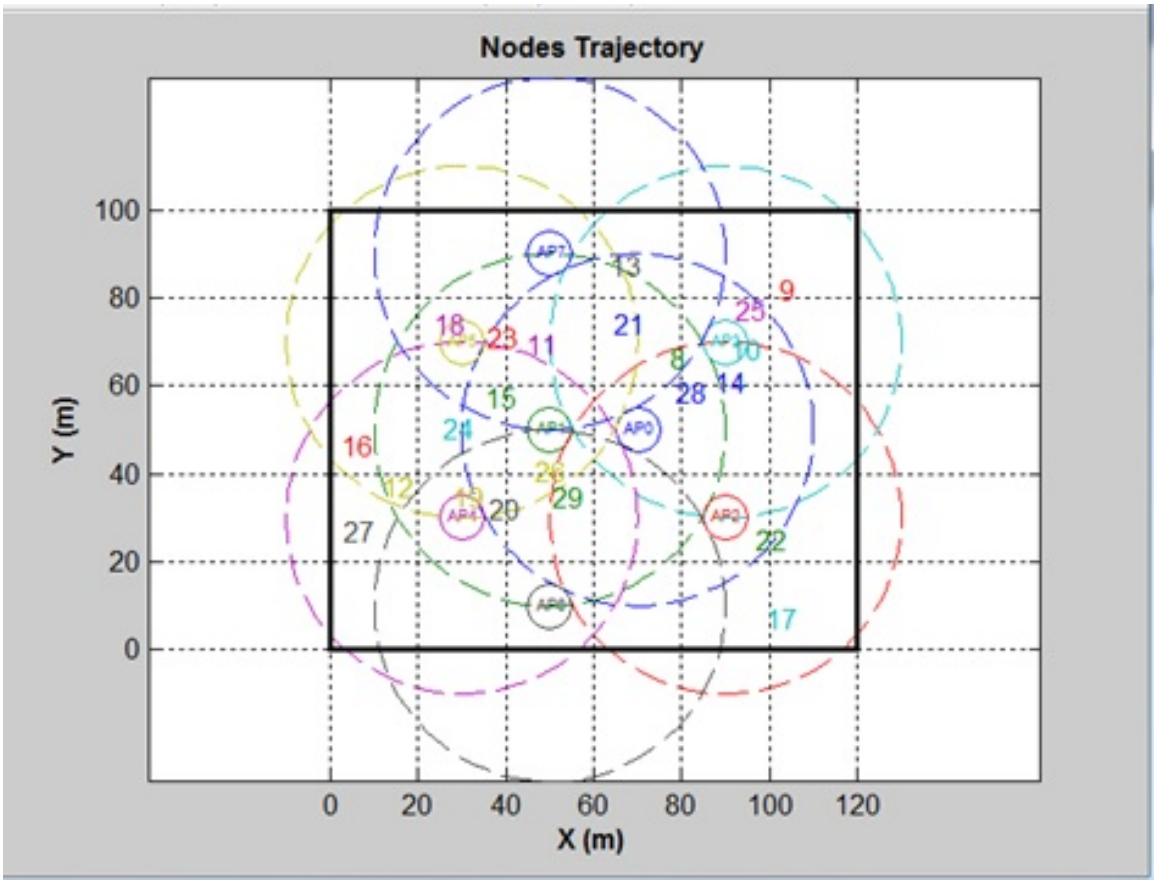


Figure 93: Simulation scenario

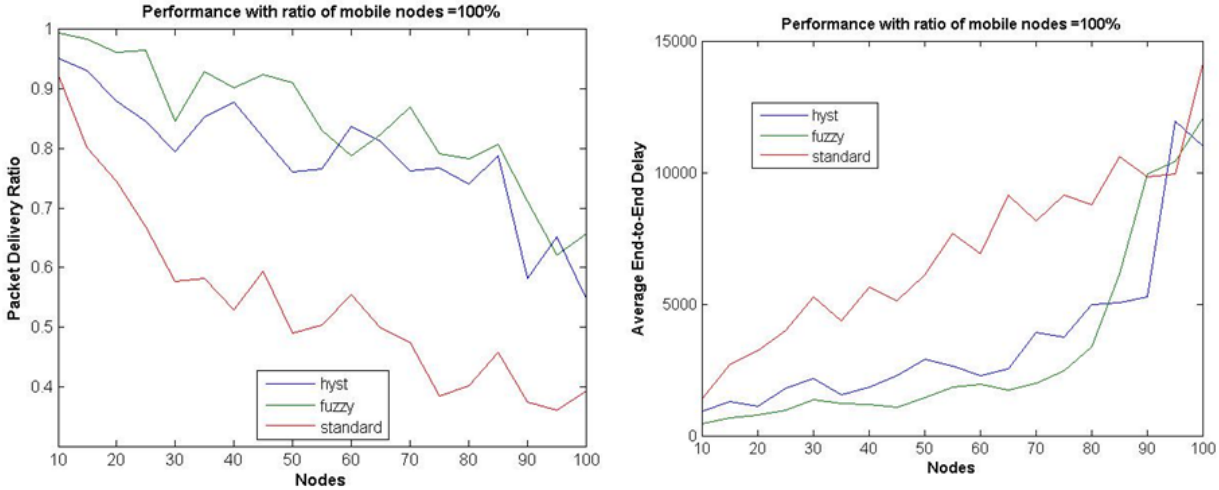


Figure 94: Simulation scenario

4 Challenges and Future Developments

4.1 Privacy and Security

This section discusses the security and privacy challenges and future possible developments. Starting with a summary of the challenges and issues, we deeply discuss business and marketing issues, the problem of the deployment of the security, the security issues related to 6LoWPAN technologies, the security of the devices, the security of the Physical Layer and some technologies related to the secure generation of credentials. For each involved topic, the section gives the challenges and gives some recommendations and future aspects that can be developed and/or studied in future IoT projects and/or solutions.

4.1.1 Challenges

According the overall security and privacy, the main challenges concern the compatibility of the requirements of the IoT stakeholders. Final users requires low cost solutions and user friendly solutions that automatically support security and privacy and add benefit. Solution providers need also low cost implementation but the privacy requirements is not the most important requirement - and sometimes privacy is not a requirement - they want to use data, perform data analytic to enhance the value of the data. These two stakeholders can have common interests but the requirements can be imcompatible. The current market is moving fast and the solutions will be very heterogenuous. With the hope that the regulation will provide framework and procedures to balance incompatible requirements.

Technically, the challenges concern the security of the Local Area Network (LAN) which can be deployed everywhere, the security of the devices which are the data provider (and/or actuator), the security of the Wide Area Network(WAN) which transports data between peers and the security of the applications. Applications may use intermediate technical entities (server , gateways etc...) to communicate with devices. Such intermediate entities could be a wake point where data can be retrieved and used without user consent. BUTLER addressed the security and privacy at design level and focused on architecture and communication; BUTLER marginaly addressed the security of the server and device implementations.

At application level the challenges concern the initialization of the security credentials allowing security bootstrapping in heterogenuous horizontal environment. At LAN level, challenges concern the concrete applicability of the security technics according to the device capabilities and the network environment.

4.1.2 Business and Market Issues

4.1.2.1 Challenge

The security and privacy of digital solution - in particular for the Internet of Things - is a great challenge according to the requirements of the Stakeholders. There are many roles in IoT ecosystem and IoT components and solutions may mix some roles according to business requirements.

IoT Application Provider (AP)

This entity provides application to end user. As example, it can be the monitoring of the location of trucks, the monitoring of goods to ensure the cold chain is not broken. For that purpose, the application uses some generic IoT services. These services are provided by Service Providers.

According Security and Privacy, the application shall be trusted and is the main recipient of clear business data. The business data shall be available in clear at this level. Application Provider may have to follow Personal Data related regulations such the European Data Protection Directive 95/46/EC EU Directive [27]. Even the most effective technology supporting Security and Privacy By Design cannot avoid fraudulent and/or usage of user data without user consent. When managing Personal Data [28] , the Application Provider shall follow privacy regulations and ethics.

From the Application Provider point of view, the challenge concerns the standardization of the API to access the Service provider (see below) and the format of the data either about the security and the clear data specifications.

IoT Service Provider (SP)

A service provider exposes some generic services for Applications. For instance, the Service Provider can support notification of events to application, local storage of data and other services like device management. IoT device must register to one or more Service Providers for providing data and Application must register to Service Providers for interacting with devices either for consuming device data or for actuating device state.

In current deployments, the Service Provider is often a technical entity where the data is available in clear. There are a secure links between the device and the Service Provider and another secure link between the Application and the Service provider. This poses a problem of Privacy because the Service Provider has knowledge of business data and such business data could be used in a fraudulent way and/or without user control or consent - for instance such data can be input of big-data behavioural engine.

The IoT Service Provider role enables device connectivity to applications. In case the IoT Service Provider plays only the SP role and if the applications and devices utilize a Security and Privacy Enabler framework - such as the BUTLER platform, they cannot make money on business data. Anyway, they can make money on data access and related services such as secure data temporary storage, device management, notifications of events etc. If the entities play only the “connectivity” and “management” roles, they do not have to follow personal data privacy related regulations such as European one. In consequence, it is important from the business perspective to provide secure data to many applications as possible. The IoT Service Provider can also be paid by providing device “management” feature to many applications.

Network Subscriber (NS)

The NS entity has a contract with the IoT Network Provider (see below) enabling WAN access of the IoT devices. This could be a consumer or an IoT Service Provider. It has to be noted that - in general for IoT solutions - the Network Subscriber is not the final consumer of data which/who consumes data through application.

Network Provider (NP)

This entity provides the Wide Area Network (WAN) access of the IoT devices. For instance, it can be a Mobile Network Operator (MNO) or an Internet Service Provider (ISP). According Security

[27] Data Protection Directive - http://ec.europa.eu/justice/policies/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf

[28] Any information relating to an identified or identifiable natural person (’data subject’); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity - EU Data Protection Directive 95/46/EC [art. 2(a)]

and Privacy, the NP needs to bootstrap the network and the associated security using Network Subscriber Identity - for instance the International Mobile Subscriber Identity (IMSI). In terms of Security and Privacy, it is up to the Network Provider for performing network security relying on lower level communication stack. Protecting the Identity is important for privacy perspective because this data can be used to track the subscriber. The data are sent over secure link from device to NP.

Device Manufacturer (DM)

The Device Manufacturer is responsible for building device interacting with the physical world. The device must support connectivity to/from one or more Service Providers - possibly using intermediate specific device like a gateway. The characteristic of the IoT devices can summarized as follow:

1. Lifespan of the deployed devices can be very long up to 10 or 20 years.
2. Generally, the devices do not propose any user interface and the user (human) involvement at installation and configuration phase must be as transparent as possible.
3. Computing resource such as CPU and memory is constrained.
4. Communication between device and application must be transparent to the user.
5. Devices can be deployed over unsecure physical environments.

The Device Manufacturer has a strong responsibility in terms of Security and Privacy. The main aspect is about the reliability of the generated data. This concerns the following aspects:

1. Quality of the sensing device. In some domains like the health care, it is important to be sure that the measured values are correct and (for actuator) that the related actuation is performed properly.
2. Resistance to hardware attacks. This aspect has not been specifically addressed in the BUTLER project. BUTLER Security Framework uses some credentials that need to be protected for instance using Secure Element.
3. Resistance to software attacks. This aspect has not been specifically addressed in the BUTLER project. The application running in unsecure device may be attacked and therefore the data can be corrupted.

Trust Enabler (TE)

This new emerging role refers to one or more entities enabling the trust and the security of the solution including the Devices, Service Providers and Applications and Network Provider. Previously supporting only one Device to Service Provider security schema, the Trust Enabler may support the following features:

1. Support of multi-actor device environment that allows different security and access settings for each actor.
2. Enable remote management of security parameters.
3. Post Issuance of service providers or end user specific services to a device after it has been deployed on the field.
4. Appoint an authorized party (Service Provider or Application) to act on the device.

Network Subscription Manager (NSM)

This new role is appearing in the IoT domain. The lifespan of the Device can be 10 years or more. Service Providers may need to be able to challenge the MNO(s) in term of price and/or quality of services. Therefore Service Provider may want to be able to dynamically and remotely update the Network Subscription on the devices. The Network Subscription Manager has contracts with Service Providers and MNO(s) for the management of the network subscriptions. In terms of Security and Privacy, The NSM shall respond to the same issues as the Network Providers.

4.1.2.2 Future

The Internet of Thing market is one of the most promising markets in Information Technology market. Many actors forecast that the market will grow rapidly. Different institute's estimations differ considerably- this highlights the difficulties on the definition of the IoT itself and the market evolution. The McKinsey Global Institute (2013) [29] estimates that the IoT can have economic impact of EUR 2.0 trillion to EUR 4.7 trillion by 2025 while Cisco (2013), referring to the analysis of 21 industry-specific and cross-industry use cases, forecasts that EUR 10.8 trillion will be at stake over the next 10 years (2013-2022).

The market is very promising. Anyway, as many other fast growing market, it grows without effective standardization. Most of IoT solutions are vertical solutions where the interoperability is not really yet considered as a problem for the actors. In consequence, this lake of interoperability implies that it is difficult - even impossible - to share assets between solutions and even to share software and hardware components among solutions.

One of the main issue is the secure management of the IoT assets and the security of the IoT applications at runtime. The IoT assets may have a long lifespan. In consequence, the Remote Management of the assets applications and security credentials will be a key factor of success of IoT based solutions. While the interoperability of the solutions is crucial, the number of expected devices over the world may have huge impact on energy consumption of IoT based solutions. Reducing the device required resources in term of storage and CPU power is an emerging requirement for the development of the IoT based solutions. Despite the cost of the energy, the green aspects of the IoT based solutions will be a major differentiator for user's acceptance.

From Static Vertical solution to Dynamic Horizontal IoT architecture

Currently the IoT solutions propose vertical solutions supporting security. The solutions use static security credentials allowing security between devices and specific applications. The schema does not support further usage of device data for new applications. For supporting sharing of device for new consuming applications, the future IoT Projects shall rely on horizontal architecture enabling the following mechanisms:

1. Remote Application Management. The architecture shall support management of device application including post issuance of device application and services in a secure way. For the market perspective, Device Manufacturers have to balance cost of the device with security (or low security) and the cost of the device supporting high level security. Support high level security in IoT horizontal solution may permit significant Return-On-Investment when the device can be used by many applications.
2. Relying on standards. Moving from Vertical market to Horizontal market requires high level of standardization. These standardization process shall include specification of interfaces between Service Provider / Application Provider and Application and Devices.

[29] http://www.mckinsey.com/insights/business_technology/disruptive_technologies

(a) Service Provider/Application Provider API. Application Provider shall be able to use another Service Provider without updating the format of the access request and/or the API of the notification callbacks. In the same way, supporting standardization at this level allows Service Provider to be interoperable with more applications.

(b) Application/Devices protocol. The device data (command and response) shall be specified to be understandable by each peers. End-to-end security communication protocol shall be standardized supporting peer authentication, data integrity and confidentiality and clear data format for each business data such as - energy, health, transport etc.

3. Secure initialization and bootstrapping of the security credentials allowing end-to-end security. This aspects is crucial for the deployment of horizontal IoT architecture in particular for constrained devices that can be used by different unrelated applications (assets sharing). Technologies such as Physically Unclonable Functions (PUF) can be investigated to support such secure initialization and bootstrapping.
4. Already deployed applications shall support secure access from potentially unknown but authorized (new) actors that may appear after deployment of the applications. Thanks to the Trust Enabler authorization mechanism and the related security protocol at application level.

4.1.3 Deployment of the Security

In section 3 - “Bootstrapping of Security” - we discussed the issues related to the initial deployment of security credentials in devices. Many techniques have been studied such as “Channel estimation” which requires that communication devices are closed enough, “Secure storage of device private key or shared symmetric key” or “Out-of band pairing” requiring another communication channel than the network one. All these techniques have drawbacks and cannot be used in all use cases.

4.1.3.1 Challenge

As described in Chapter 3, the commonly deployed security architectures are deployed according to 3 phases.

1. The Bootstrapping at low-level (LAN). It consists of securing the “small” data between the object and the gateway / proxy / modem linked to the WAN. It may consist in crediting the components of the LAN with a shared session key that we can call “local” session key.
2. The Bootstrapping at “high-level” (WAN). It consists in distributing and using the access rights which can take the form of security credentials such as login/password or the form of more sophisticated access-token.
3. The Session establishment. It addresses the problem of distribution ephemeral session credentials from the device to the application in order to implement “end-to-end” security.

UL used a pre-shared key mechanism to provide security in their 6loWPAN network composed by TelosB platforms. The size of RAM/ROM memories of these sensors is limited and does not permit to implement more complicated and secured schemes. The future deployment of the security will rely on new cryptography schemes based on ECC (Elliptic Curve Cryptography). Alternative motes will also enter into consideration with additional memory space.

The bootstrapping of security at “high level” in the WAN is provided by the BUTLER security framework which is an access token based security framework using and enhancing OAUTH 2.0 protocol. The BUTLER security framework also supports distribution of session key for providing end-to-end security between peers.

In an access token security concept, the resource (destination entity) shall verify the access token presented by the calling application. In case on IoT, the resource generally cannot securely reach the Authorization Server for checking the access token and allow the access. In these situations, the resource shall check itself the access token. Many techniques can be deployed such as pre-sharing of symmetric keys used to sign and encrypt the access token - the current implementation of the BUTLER security framework relies on symmetric cryptography - and/or signature and encryption performed using Public Key Infrastructure (PKI)model. Using the PKI model, the device can check the signature of the token using the serveur public key and decrypt the token using the device private key. All these techniques require pre-deployment of initial security credentials - pre-shared keys in case of symmetric cryptography or for PKI model the certificate of the certificate authority used to sign the certificate of the server public key and the device certificate related to the device encryption private key. Using symmetric cryptography, the commonly used technique is to personalize the device with a key derived from a Master key and a device identifier. The personalization process must be performed in a secure way - for instance at secure device factory. Using the device identifier, the serveur can retrieve the key either running the same algorithm or by retreiving it from a secure database. In the last case, the device factory must securely provide the server with the device identifier and the associated derived key.

When the security is deployed using access token paradigm provided by external Authorization Server like the BUTLER one, the relationship between device and Authorization Server can be considered as a vertical relationship where the Authorization Server and devices can be in a same ecosystem. If they are in a same ecosystem, they can share the keys using the symmetric technique described above. If they are not in the same ecosystem, they can use PKI model but it is costly and complex to deploy properly. In case the two solutions are not possible for any reasons (device constrains and/or business requirements - for instance for blank/unpersonalized device) a solution may consist of Out-of-band Pairing involving the user for setting/getting the keys both at the device side and server side. The out-of-band pairing is the current BUTLER solution for setting initial security credentials in the peers software.

Trusted Execution Environment or Secure Element do not fix the deployment issue of unpersonalized devices

Using Trust Execution Environment (TEE) or Secure Element (SE) on the device does not fix the problems of the deployment of the initial security credentials. TEE and SE insure protection and security of deployed credentials but the deployment system still requires initial security credentials. Both TEE and SE need external system in relation with device manufacturer for setting security credentials.

In case of the TEE, the device manufacturer generates unique key using the derivation mechanism decribed above and provides the information to the TEE provider for bootstrapping the security credentials. In case of SE, the silicon manufacturer provides the key (called factory key) to SE provider for bootstrapping the security credentials.

For Secure Element (smart card, embedded SE) and NFC related use cases (payment, transport, access control, etc), this technical relationship is provided by a Trusted Service Manager (TSM) role. Global Platform (GP) has defined the GP messaging allowing service management between the Service Provider and the TSM and specifying the interface between the SP and the Mobile Network Operator (MNO) and the Secure Element.

Currently, the main current disadvantages of the TEE related technologies concern is availability of integrated deployment architecture supporting secure deployment and personalization of Trusted Applications (TA) running in the TEE. Trusted Applications in the TEE environment must be securely loaded over the air and personalized according - for instance using the unique device identity. TEE

providers have contractual relationship with device manufacturer for defining the initial security credentials required to setup the security for a Service Provider (SP) related TEE Trusted Application. For this purpose, SP should have technical and commercial relationship with the TEE providers which can be complex and depends on the TEE providers. For now, The GP specification is not well adapted for the TEE based infrastructure which is user centric and where the MNO role is not always relevant.

4.1.3.2 Future developments

For enabling horizontal integration and sharing of device data for different applications, future initiatives should focus on secure deployment of initial security credentials. Due to the difficulties related to the deployment issue, the IoT solutions often do not support security or support low level security such as data integrity - thanks to hash algorithm. For enabling the security, user may interact with the device for setting/getting the security credentials and make the link with the server, but for now the processes are generally not very convenient and/or not very secure.

The main challenge is to provide a key management scheme that works in constrained environment. We proposed to design and simulate new key management protocols that are critical in an IoT deployment. The current security in a wireless sensor network is often based on pre-shared secret keys that should be efficiently managed over the time, with specific storage and distribution schemes. Simulation will enable the performance analysis of new schemes on different network topologies.

For future IoT projects using blank/unpersonalized devices, the projects should work on convenient and secure mechanims to setup initial security credentials. As these devices do not embedd any secure information allowing secure remote personalization, the initialisation mechanisms may involve the user in a convenient manner and/or may use context information for defining and securely publishing persistant credentials for the lifetime of the context.

For future IoT projects using TEE enabling devices, the projects should work in collaboration with Global Platform members for updating the Global Platform Messaging to be compatible with TEE model to deploy and personalize the Trusted Application(s) required for the secure execution of the Service related applications.

For all future IoT projects, a major focus must be on the usuability of the security bootstrapping process.

4.1.3.3 Deployment of the Security

A major challenge in 6LoWPAN wireless sensor networks is to secure single-hop but also multi-hop communications. The goal of key management is to establish secret keys between sensor nodes that need to communicate securely. One important characteristic of key management protocols is that they must be scalable, CPU and energy efficient. New nodes are periodically deployed in order to assure network connectivity. These new nodes must be able to establish secret keys with previously deployed nodes. Protocols that provide this property are known as Multi-Phase deployment protocols. The two presented algorithms, RPL and ARC (see 3.1.2.1.1) will be used to design and simulate new key management protocols that are critical in an IoT deployment. The security of a cryptography system is based on secret keys that should be efficiently managed over the time, with specific storage and distribution schemes. In fact efficient authentication mechanisms are required to ensure secure communications between IoT devices. Implanting, initialization, link establishment and operational phases will be added into the NARVAL module in order to support confidentiality and integrity protection, mote bootstrap, key negotiation and maintenance

(http://tools.ietf.org/html/draft-ohba-6tisch-security-01 and http://tools.ietf.org/html/draft-piro-6tisch-security-issues-01). The implementing phase consists of the installation of security credentials in each IoT node before its deployment. During the initialization phase (Phase-1), an authentication and key establishment protocol is performed between nodes or between each node and an authentication server (generally the border router of the 6loWPAN network). For authentication of nodes that need a multi-hop communication to reach the authentication server, intermediate nodes can play the role of the authentication server, acting as an authentication relay. During the link establishment phase (Phase-2), the Security Association (SA) of a link between two nodes is enabled according to 128-bit keys used in AES-CCM (IEEE 802.15.4) mode. AES/CCM is a quite efficient and very secure algorithm as long as the same nonce never occurs twice with the same key. Many applications will require re-keying within the lifetime of the 6LoWPAN network. The network key can be shared by all network nodes and used to encrypt and decrypt messages. There exist alternative models, where for instance each node uses a unique key, or where each pair of nodes has a unique key. Finally the nodes are assumed to be secured in the operational phase (Phase-3).

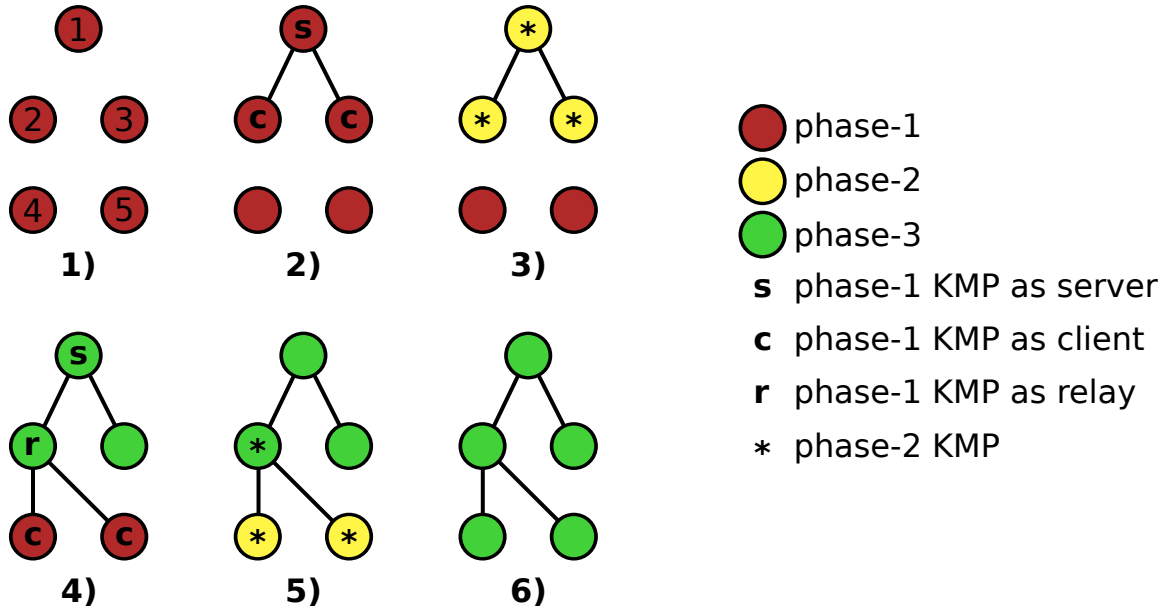


Figure 95: Security deployment in a small network topology

Figure 95 provides a simple example of the security deployment on a small network topology composed by 5 nodes. At the beginning of the experiment, all nodes start in Phase-1 (1). Thus the nodes 2 and 3 begin Phase-1 Key Management Protocol (KMP) with node 1 which is acting as the authentication server. They retrieve Phase-2 and Phase-3 credentials (2). Thereafter nodes 2 and 3 run Phase-2 KMP with node 1 (3). Nodes 4 and 5 perform Phase-1 KMP using node 2 as authentication relay (4). Alternatively node 5 may use node 3 as an authentication relay. Thus nodes 4 and 5 runs Phase-2 KMP with node 2 (5). Finally all nodes are operational (6). Simulations will permit to analyze the efficiency of new KMPs on different network topologies.

4.1.4 Security in 6LoWPAN-based IoT

4.1.4.1 Challenges

This section describes generic challenges that exist in providing security in 6LoWPAN. Also, it emphasizes specific challenges existed during some tests. The most important security challenges in any constrained environment were well described in security analysis and consideration for 6LoWPAN protocol in IETF [213, 214]. Also, few design considerations for security protocols in constrained environments was outlined in [215]. Wireless sensor networks based on IEEE 802.15.4

and 6LoWPAN share the frequency spectrum with Bluetooth and WiFi, therefore it suffers from their interference which directly influence reliability and connectivity in delay tolerant networks. IPSec and DTLS for CoAP are the existing state-of-art security solutions in any 6LoWPAN network, but this solution is not scalable as both needs more memory. Also, it has specific hardware requirements and don't fit into the nodes with small ROM. Keys management and exchange is considered to be an important issue in deploying large scale networks as described in the previous section 4.1.3 and it currently faces vulnerabilities such as passive eavesdropping and extraction of secret parameters by node capture,etc.
Further, Denial-of-Service (DoS) attacks on 6LoWPAN's is one of the most imminent threat to the networks' function, as an example: jamming; the routing protocol 'RPL' itself is vulnerable to attacks such as RPL-Rank attack, selective forwarding and common routing attacks.

4.1.4.2 Future Developments

To overcome the issues described above, light-weight scalable security solutions for key-exchange and encryption mechanisms should be developed while keeping in mind the design considerations for constrained environment. IPSec and DTLS should be tested for stable communication and energy efficiency within nodes. Already existing solutions such as tiny-DTLS, Lithe-Light weight IKev2 solution for IPSEC should be optimized. To defend and protect against interference and other DoS attacks, Intrusion Detection Systems for 6LoWPAN could be considered as an effective solution. [216, 217].

4.1.5 Security of Devices

4.1.5.1 Challenge

As already stated in section "Business and Market issues" the Device Manufacturer has a strong responsibility in terms of Security and Privacy. The main aspect is about the reliability of the generated data. This concerns the following aspects:

1. Quality of the sensing device. In some domains like the health care, it is important to be sure that the measured value are correct and (for actuator) that the related actuation is performed properly.
2. Resistance to hardware attacks.
3. Resistance to software attacks.

4.1.5.2 Future developments

Device Quality and Certification

In some use cases - for instance health or energy related use cases - it is important for the application to rely on device data. For this purpose, it could be important to refer to some certification process(es) giving accuracy of generated data according the physical world. In this sense, it could be useful for application to obtain ranges and reliability information on the values.

Another problem is the responsibility of the device manufacturer in case device provides wrong values which may be have non expected (and potentially) important impact at application level such as wrong alarm, unexpected notification, financial impact etc.

Future projects could focus on:

1. software security technology supporting protection of the device software against malwares.
2. device evaluation tool providing some evaluations according to the resistance to software and hardware attacks.
3. tools performing verification of the accuracy of the delivered data could also be useful for the applications point of view.

Secure Element

For now, the IoT devices do not embed secure element for performing security features. Some devices support cellular GPRS connectivity using SIM card for WAN connectivity. Even if the SIM card can run applications on behalf of services providers, most often, for business model reason; the IoT service providers do not accept having specific contract with Network Operator which could be dynamically updated during the lifespan of the devices. In consequence, the Service Providers may need Secure Element based security for implementing high secure services. Long lifespan of the devices and the exposure of the device in heterogeneous unsecure environments leave the devices vulnerable to physical and logical attacks. The reliability of the security implementation can be enhanced by using Secure Element for storing security credentials and for running cryptography algorithms. While the Secure Element can carry the network subscription and related security credential, it can also store and run specific secure applications for service providers.

The deployment of the security credentials on Secure Element is already available using Over The Air protocol specified by Global Platform organisation.

Future works could focus on:

1. updating security deployment protocols that can be supported by constrained devices either concerning the required network bandwidth, and the required CPU power and memory sizes.
2. providing Secure Element with extremely low energy requirements

Trusted Execution Environment

The Trusted Execution Environment (TEE) is an operating system which provides a way to execute some parts of an application in a secure way. TEE uses application processor architecture supporting secure mode execution such as ARM - TrustZone. TEE runs the application processor in a secure mode with a dedicated Virtual Memory that is not accessible when the processor is not running in secure mode. TEE provides security features such as secure storage of the sensitive information and secure execution of code like cryptography code. TEE offers more security than the rich OS, but is considered as less secure than a secure element.

For now, the IoT devices do not support Trusted Execution Environment. It is probably because the cost of the security deployment over a TEE is equivalent to the cost of the security deployment using Secure Element for a solution that is vulnerable to physical attacks on unsecure environment. The cost of the device is important and the TEE technology can provide a good balance between security and security related cost. It is expected that future device hardware will integrate TEE technologies.

4.1.6 Physical Layer Frameworks

4.1.6.1 Challenges

Embedded in all the aforementioned expressions in the Section 3.1.2.3 is the distance between legitimate nodes, and between those and eavesdroppers. The key role played by the relative location of devices, together with the lack of exact knowledge on the latter, lead to the widespread utilisation of stochastic-geometric approaches when analysing secrecy in random networks. In most of the works Probability Generating Functional (PGFL) of PPPs is utilised to obtain simple and closed-form for the secrecy outage probability in random networks with uniform random topologies. Clearly such results do not generalise to networks where a minimum distance between any pair of nodes must be maintained - as is the case of WiFi and Cellular Networks - which are known to require Hard-core Point Process (HCPPs) models instead. The fact that PPPs are not sufficiently accurate to capture the structure of various random networks of interest is a modern topic in wireless communications both within and without the particular question of secrecy. Literature on the impact of new Point process models in random networks is not vast, but the issue has not entirely escaped the attention of the community. Recently, focus has started to shift so as to address the somewhat naive assumption of uniformity, found in all aforementioned works in the Section 3.1.2.3. To clarify, it has been shown that the PPPs cannot accurately model the majority of wireless systems of interest, including cellular and WiFi networks.

4.1.6.2 Future Developments

We will study the inherent secrecy in random networks, with the particular aim of generalizing topological models. We will investigate the secrecy outage probability of cellular networks, employing emerging stochastic geometric model HCPP. Following this trend, we will look beyond the PPP model and study the secrecy of random networks under various spatial distributions. We should also point out that to the best of our knowledge, the analysis of secrecy of random networks outside the PPP model have not yet been attempted in currently literature. In that regard, therefore, we will bridge a gap between the progress made on the study of random networks outside and inside the secrecy question. Further study would be carried out to analyse more complicated scenarios such as fading, eavesdropper colluding while considering interference from other distributed nodes.

4.1.7 Challenges and Further Improvements of Secret Key Generation from IR-UWB Channels

4.1.7.1 Challenges

First of all, the high sampling frequencies needed to generate sufficiently long keys for cryptographic applications might be a limitation for implementing the proposed solutions in today's wireless devices. In order to address this issue, we could consider channel estimation algorithms that give access directly to infinite-bandwidth channel representations under lower sampling constraints [218], [219]. These estimations can then be convoluted with $x(t)$ or any other relevant waveform in order to produce a high-resolution convolved signal $y(t)$ that could serve for extracting a relatively large number of bits. This leads us back to the quantization of a signal with deterministic characteristics for which we can apply the HIST algorithm.

Secondly, an alternative to single-bit quantization of sampled signals can be the multi-bit quantization of an estimated UWB channel response $h[n]$. In order to find a trade-off between reciprocity (an indicator of robustness) and codeword diversity (an indicator of randomness) under a fixed-length constraint, new investigations have been performed regarding non-uniform quantization. The results show the interest of optimizing the quantization thresholds as a function of the signal-to-noise

ratio in the specific IR-UWB context [220], thus suggesting the adoption of adaptive thresholds as a function of the received channel excess delay.

4.1.7.2 Future Developments

Further investigations could be performed regarding the security of the scheme with respect to a passive eavesdropper that can observe the side channels between him and any of the legitimate users. The secrecy of the key generation protocol using a source model (i.e. the wireless channel) is based on the assumption of spatial decorrelation of radio signals, which does not hold in some scenarios (e.g. small distances, room symmetries etc.). This particular cases should be studied in order to understand how an attacker could exploit these advantages and his side observations in order to get an estimation of the legitimate channel. Furthermore, this could lead to improvements of the secret key generation scheme by avoiding the use of information that can be easily inferred by a potential attacker.

4.1.8 Challenges of Security at Low Level

4.1.8.1 Challenges

CEA has provided :

- A Threat Analysis Model that should be processed before the design of each application over a dedicated network architecture.
- A key management scheme deployed for a ZigBee components,
- A lightweight handshake enabling an easy and large scale deployment of heterogeneous devices.

The threat analysis is a powerful tool that enables to identify the vulnerabilities of a whole system handling various components. With this tool, the security can be envisaged without compromise for the values to protect and its cost is known at the beginning of the project. This tool should be systematically used for each new application to deploy over an IoT network and improved notably for quoting the impact of the threat. CEA has proposed a quotation based on qualitative considerations. Others quotations based on quantitative estimation exists in the litterature. We see by the use that the quotation of the impact of a threat in a given situation can vary significantly depending on the subject that performs the threat analysis. New quotation systems could be introduced in the future.

The key management scheme deployed for ZigBee components runs well. However, it is based on pre-shared symmetric keys. This kind of deployment may be very tedious at large scale. Moreover, this scheme is "'stand-alone"' and is not compliant with the trust server transactions designed at high level.

Keeping the goal to enable an easy deployment at large scale of the IEEE 802.15.4 constrained devices, a new lightweight handshake has been introduced. It enables the deployment of numerous nodes with the assumption that each node embeds a "'True"' random number generator. Such this API has been designed and realized for Butler project. The remaining challenge is to connect this security protocol involving the most constrained nodes of the LAN with the trust srver scheme at high level in the WAN.

4.1.8.2 Future Developments

In the future, CEA will pursue this work keeping the goal to provide Plug & Play secure bootstrapping techniques:

1) CEA will design short certificates for resource constrained devices to enable their security management from a trust server along time. The ideal will be that the certificate length does not exceed the length of the data payload of an IEEE 802.15.4 frame.

2) CEA plans to provide a semantic for Plug & Play secure bootstrapping techniques.

3) CEA will specify how the security at low level deployed over the LAN interacts with the security at high level based on the trust server. In particular, several techniques should be envisaged to push securely the token from the trust server to the leaf (resource constrained node) enabling both end-to-end security, dynamic security management and respecting the security by design principles.

4.2 Localization

4.2.1 Challenges

In comparison to the quality and omnipresence of satellite- and cellular-based systems (GPS and Mobile networks) in open outdoor spaces, wireless localization systems in indoor and urban scenarios [221, 222] are still quite fragile and under-deployed.

For currently deployed localization systems, the performance of the systems are evaluated in terms of their accuracy in target location and tracking, which is not sufficient as metric for positioning systems. An indoor positioning system which meets localization requirements and solves open challenges in wireless localization such as accuracy, precision, scalability, complexity, robustness, cost and power is necessary. Therefore, ranging and positioning techniques/algorithms that meets all the necessary performance metrics required for wireless localization.

For improvement in the performance and stability of ranging and positioning algorithms for target localization, having a priori the knowledge of the ranging statistics is necessary. Therefore, estimating the limit of localization error associated with each node is a fundamental problem within Wireless Sensor Network (WSN) context. In literature to this regards, the most widely used tools are the CRLB [223–225], describing the average estimation error (i.e. the *distance* between the estimated and actual node location) and the Position Error Bound (PEB) [226], depicting the *region* where the node should be estimated within a certain confidence.

However, CRLB and PEB both rely on the knowledge of the true target location and the distribution of the ranging errors; which depends on various environmental factors such that obtaining their formulation *a priori* is almost impossible.

The only practical solution is therefore to estimate this statistic directly on-site during the network deployment, collecting samples from each link and then obtaining the limit on localization error even before obtaining the location estimates.

To this end, the well known *maximum likelihood parametric approach* is going to fail, given the lack of a priori knowledge on the error distribution. A truly non-parametric approach is therefore required; in particular the *kernel method* is very appreciated for its capability to reconstruct empirical distributions from samples.

Also, as it is well known, localization systems could fail when there are not enough position-related measurements. Cooperative localization and integration of low-cost inertial sensors (INS) could be good aid solutions for the current RF-based localization systems, but there are still some open challenges, such as efficient localization algorithms and practical implementations.

From a practical point of view one challenge is to design non-parametric distributed algorithms that, at a low computational complexity, can recover the targets' locations as well as their confidence. One step towards such a result was accomplished though the CIS algorithm that in the future will be implemented in a distributed fashion.

4.2.2 Novel Cooperative Localization Algorithms

Currently, research activities focused on cooperative positioning approaches are still on the theoretical phase. They are usually high in complexity, lack of real implementation and testing. The cooperative localization approaches could be adopted to enable various applications in the field of the IoT, which often require high position accuracy and low energy consumption. Therefore, the future research topics should include the adoption of more terrestrial communication technologies, the study of less complex positioning algorithms and the realizations of cooperative P2P networks.

At present, cellular communication, UWB devices, inertial sensors are adopted separately to augment the GNSS-based positioning system, but they can be combined together to further improve the localization performance. Furthermore, other future positioning techniques might be considered to support Intelligent Transportation Systems (ITS) applications, in combination with the adoption of the IEEE 802.11p protocol, which is an approved amendment to the IEEE 802.11 standard to add wireless access in vehicular environments (WAVE). The hybrid and cooperative positioning algorithms, like HC-PF and H-SPAWN, which show large computation complexity, may not be suitable to be implemented in cheap devices. Therefore, simpler algorithms, such as improved least square and belief propagation based on Kalman filter, should be studied to reduce the complexity while keeping the same accuracy.

4.2.3 Indoor Localization with Low Cost Inertial Sensors

Indoor environment is complex from localization point of view. In fact, it is full of obstacles, furniture, walls, people, which may cause the interruption of range measurements for a period of time, making the localization process fail and location-based service become unavailable. In order to overcome this challenge, two typical approaches can be adopted. The first one is the deployment of redundant devices to provide additional range measurements. This solution has increased hardware and maintaining cost. Another positioning solution is the adoption of low-cost hybrid solution that combines radio frequency localization with intertiao sensors (INS).

INS includes motion sensors (accelerometers) and rotation sensors (gyroscopes) and via dead reckoning they can continuously help calculate the position, orientation, and velocity of a mobile object without the external references. Traditional INS are expensive and large in size, and are usually used on vehicles such as ships, aircraft, submarines. Recently, small Micro-Mlectro-Mechanical Sensors (MEMS) have become affordable and available, enabling low-cost positioning solutions for indoors.

INS can help indoor localization as follows: on one hand, they can work together with measurements from other technologies (i.e., UWB, ZigBee, WiFi) to improve the position estimate; on the other hand, they can be used in dead reckoning to localize the mobile objects when other system fails.

Future research activities should be focused on the integration of low-cost ins with the current UWB indoor localization system. There would be three main steps to follow: first, design hybrid UWB/INS localization architecture and positioning algorithm and perform computer simulation to get some basic knowledge of the achievable performance; then, perform real experiments and refine the designed architecture and algorithm; finally, apply the design hybrid UWB/INS approach to provide LBS for real IoT applications.

4.2.4 Heterogeneous and Distributed Positioning Algorithms

As wireless networks evolves, multimode operation of networks are more and more common. That is networks will be supporting and running simultaneously several radio technologies. Heterogeneous and ubiquitous localization service in a modern IoT environment requires to solve a question

how to select and handle a different type of information (e.g. possessing same spatial correlation properties) in a unified manner. Furthermore, handling different measurement latencies and to preserve the quality of the location estimates is a great challenge. Thus designing a solution that is capable to combine and handle data of different nature is asked. In addition, a solution needs to be optimized regarding the requirement of positioning accuracy for specific user case versus energy consumption in data collection, computation and transmitting information amongst users parallel with other user applications.

Future studies will focus on extending a set of types of information used in a positioning application. A traditional positioning system relies on information such as range and range difference measurements or angle information between agents in a network and lately *hybrid solutions* combining these informations (although still under research).

More recent updates for positioning systems are taking into account semantic information (*i.e.* semantic similarity of an user surroundings compared against a priori recorded database of an environment or real time against other users). Taking into account other users and especially, regarding and sensing other users as a crowd is a rising research topic in various fields in information technologies [227, 228]. Mobile Crowd Sensing (MCS) is a new paradigm to provide information from the set of various mobile devices (e.g. smartphones, wearable sensors, etc.) [227] opposite to measurements recorded by single user per single device.

Furthermore, as capabilities of personal devices are coming greater, a methods of Multicriteria Optimization and Decision Analysis (MODA) should be exploited in future developments. MODA deals with the various aspects of finding optimal decision or decisions with multiple alternatives and conflicting objectives.

Compared to single-objective optimization, in MODA there exist no single optimum solution but several isolated solutions, and it requires a search over the set of valid solutions. Thus additional level of optimization is required to have efficient algorithms. The research would aim at the investigation for multicriteria optimization strategies to handle heterogeneous information.

4.2.5 Non-parametric Estimation of Error Bounds in LoS and NLoS Environments

We seek to propose an efficient and accurate method to evaluate on-site the fundamental error bounds for WSN localization. While there exist efficient tools like CRLB and PEB to estimate error limits, in their standard formulation they all need an accurate knowledge of the statistic of the ranging error. This requirement, especially under NLoS environments, is impossible to be met a-priori. We will show therefore that collecting a number of samples from each link and applying them to a non-parametric estimator, like the Gaussian Kernel (GK) and Edgeworth Expansion (EE), could lead to a quite accurate reconstruction of the error distribution and then, in turn, of the error bounds. The EE method will then be employed to reconstruct the error statistic in a much more efficient way – less number of samples required – with respect to the GK. We would finally show that with the proposed EE method, it will be possible to get fundamental error bounds almost as accurate as the theoretical case, i.e. when perfect a priori knowledge of the error distribution is available.

4.3 Behavior Modelling and Synthesis

4.3.1 Challenges

Although behavior modeling and synthesis had progressed significantly in the scope of the horizontal architecture of BUTLER, there are still many open challenges as listed in this section. In line with Section 3.3, we have broadly divided the challenges into two categories - i.e. enhanced algorithms and distributed systems realizing these algorithms.

In general, IoT will further challenge the big data paradigm, as the interconnected devices become smaller while grow in numbers. This will further complicate efficient handling of relevant data both from an algorithmic and distributed computing perspective, especially in the areas of collecting, aggregating, enriching, mining, storing and sharing data in IoT applications effectively while considering security and privacy constraints. The next research focus should be on practical, scalable and distributed streaming-oriented software solutions to better understand the value and veracity of data, as well as with the ability to deal with the growing velocity, volume and variety of streaming data. This will require new and horizontally scalable approaches towards stream mining and learning, and contemporary approaches on ontologies and semantic interoperability will have to be revisited as well.

From the algorithmic perspective, we foresee a need for 'soft' user identification techniques that would inherently identify the users without requiring explicit user inputs, in order to make IoT applications truly ubiquitous and personalized without having the user overwhelmed with large amounts of data and violating his privacy. Of particular interest are techniques to identify changes with respect to established patterns in common behavior. We foresee new research initiatives to tackle the challenge of differentiating concept drifts from noise, detect and learn such evolutionary concept drifts in order to improve adaptive learning with temporal windows. Another way to maximize the benefits of learned knowledge and experience is to transfer knowledge learned from one task to the other. In the future, we will aim for improved algorithms, modeling, verification and validation methods that would enable effective knowledge transfers across smart domains and users.

Similar to concept drifts, in real-world applications the causal dependencies of context-aware behavior will have to be dynamic and evolve over time. Hence, in addition to improving the accuracy of directed information techniques, we should support dynamic adaptation of the structure of the causal networks. As resource provisioning is still a major problem in IoT devices, we foresee algorithmic improvements for efficient cryptographic applications that supports better bandwidth utilization and support adaptive thresholds. Also, as in the IoT everything will be connected on a more dense scale, the attack service of IoT applications will grow. Therefore, security mechanisms and assessments of new proposed schemes and protocols remains a continuous and growing cross-cutting concern.

4.3.2 Software Engineering Perspective

4.3.2.1 Evaluating and improving the scalability of the framework

In our current preliminary evaluation of SAMURAI, we have benchmarked the performance of the behavioral smart server

1. under centralized and distributed deployment settings
2. for processing time as a function of number of REST requests
3. for horizontal scalability as a function of number of nodes

In the future, we would like to extend the evaluation for both vertical and horizontal scalability especially with respect to the number of users under the multi-tenant settings. Improving upon our current context models, we would like to evaluate on more complex scenarios with larger user base. We will explore large scale data processing frameworks such as Spark, in order to improve the throughput of the behavioral smart server. Also, we will compare the domain-knowledge driven aggregation and data driven distribution by state-of-the-art data processing engines.

4.3.3 Algorithmic Aspects of User Context Recognition

4.3.3.1 Continuous User identification

Modern intelligent environments thrive to provide personalized and real-time user centric services both cheaply and non-intrusively. Continuous user identification is indispensable for context-aware applications embedded in such an intelligent environment. Although the user identification techniques in most systems such as identification using a RFID tag or his/her mobile device provide reliable identification, it is both cumbersome for the user to carry some physical device always and in case the user has misplaced them then the system will end-up erroneously tracking the tag/device instead of the user. For example, in BUTLER we assume inputs from a set of devices correspond to a particular user.

We propose an implicit user identification system with contextual information inferred from sensors available in a smart phone along with other context sources, to improve the overall user experience. Such a user identification system can enable intuitive and seamless information (or) service access mechanisms without requiring explicit user attention. The proposed user identification system will be built on top of HARD-BN framework which can model, infer and keep track of multiple user-contexts and correlations between them. The consistencies between the user contexts in the underlying Bayesian framework will be capitalized for implicit user identification. In case of inconsistencies, it will prompt the users to identify themselves manually with their credentials. Some of the typical scenarios where manual identification would be requested include,

1. Deviations from the expected default user context which is already learnt by the framework. For instance, an estimate of the current location of the mobile device (e.g., in the train or city) is different from the expected default location of the user (e.g. at work). The challenge is to differentiate the noises from the actual deviations in the ground truth.
2. Any inconsistencies between different localization modules available in the framework. E.g., work computer usage estimates user presence at work, whereas the WiFi based localization estimate the location as home. The challenge is to replace the deterministic classical identification systems with probabilistic identification systems which can reliably combine multiple sources of information along with their uncertainties.

4.3.3.2 Advanced context-model adaptation to handle concept drifts

Incremental learning or online learning, is an important enabler for stream mining and classification where the classifiers can learn with experience. Currently, HARD-BN framework supports three types of incremental learning -

1. updating context models on new data
2. adding new class values (i.e., re-structuring the output parameter space) which were not available at design time
3. incorporating new contextual sources or sensors

A major open challenge yet to be addressed is to incrementally learn any changes in the output distribution and conditional distribution between the input feature and output, which is popularly addressed in the machine learning community as concept drifts. The challenges are manifold - to detect any changes in the distribution, learn new distribution on top of the current models, determining which part of the previous model (knowledge) is relevant and preserve it while learning the new knowledge. We plan to support the detection and learning of the following types of concept drifts,

- a sudden or abrupt drifts such as change in 'work' location

- gradual drifts such as increased daily physical activity levels of the user
- recurring concept drifts such as week day/week end specific difference in user behaviors

4.3.3.3 Enhanced support for transfer learning

Although the recent works in sensor technologies and machine learning have considerably improved the recognition accuracy for activities of daily living, their robustness is highly limited even in such simple use cases and requires complete re-learning of the models. We categorized the transfer learning problem in to multiple smaller problems:

- Finding suitable abstractions and specifications to model the objective and subjective parts of contextual information in order to facilitate a partial transfer of models for the new tasks/users.
- Analyzing the various sources for subjective parts of the model and how to address them. Some typical sources are devices, pre-processing algorithms for feature extraction and the user itself. In the previous section on contextual networking, we addressed the differences in the models arising from the difference in the devices and input features.

In the future, we would like to investigate how a trained model for one individual can be applied for another individual - (transfer learning) - to reduce the overall training effort. First, we will analyze user details, associated contexts and learned models of different individuals (collecting data using a crowd sourcing-like approach). For clusters of similar models, techniques will be developed to elicit the corresponding similarities in the user details and context. These similarity requirements will be used to adapt the model parameters to ensure that transfer learning can be applied without adversely affecting the activity classification accuracy for any individual user within the cluster.

4.3.4 Contextual Networking

A context-aware protocol needs to adapt automatically and dynamically to its environment requiring the difficult task of combining a multitude of metrics. Thus, a mobility and multi-context management module has been implemented in an event-driven and packet-oriented simulation tool named WSNet. This module adapts in real-time the intra-cluster communications depending on the detected local and temporal context patterns (i.e. in terms of nodes activity, mobility, energy autonomy, traffic, etc.) and provides a handover mechanism ensuring the stability of the whole network. During this work on BUTLER, several challenges were observed and are described hereafter.

Firstly, inside the proposed module, the weight values for each metric inside a group and for each group were defined manually, which can be non-optimal. Therefore, further investigation on both the optimization of the weight value of each metric inside their groups and the definition of profiles (i.e. specific sets of group weights) according to the specificity of the application are required.

Secondly, another open issue is the definition of a mechanism for inter-cluster handover. The inter-cluster handover is complex once it requires informations being exchanged between different clusters, operating on different communication channels.

Finally, future research must focus on the integration of both intra and inter-cluster handover inside the CLUBCROM protocol along side with the optimized weight values for metrics and profiles.

4.3.5 Contextual Management

The definition of entities and their associated contexts by the end-users have proven to be a valuable feature provided by BUTLER. The automatic synthesis of the context contents by means of Complex Event Processing technologies eases the way end-users can take advantage of systems that

mediate IoT data streams such as those provided by BUTLER. It also can complement the "classical" context management architectures very much modelled from the Open Mobile Alliance (OMA) Next Generation Service Interfaces (NGSI) proposals, as those are much more focused on providing context contents to consumers than in the creation and synthesis of the actual contents of a given context. However, some limitations and drawbacks have been already described in section 3.3.1.9.

On the other hand, the unbounded possibility of creating virtual entities and associated context contents (even with appropriate tagging for subsequent discovery) from the resources exposed by said virtual entities is not enough when it comes to the modelling of the behavior. CEP is quite powerful when it comes to creating new data streams from existing ones (even considering different time windows in order to compute averages, maximums and minimums over a period of time, or comparisons with data stored in a static database). Although the theoretical expressivity of what could be obtained in terms of context contents from a full-fledged use of a business rule language such as Drools Rule Language (DRL) could be enough to infer behavior associated to an entity, the limiting factor is not the rule engine, but the lack of an inference engine, as the DRL rules are applied not to one of said engines but to a CEP infrastructure. Moreover, ad hoc creation of virtual entities and associated context makes it difficult to map them to existing ontologies and therefore to benefit from the inference mechanisms the semantic technologies provide.

In that sense, future research and experimentation activities should focus on the following challenges:

- From a purely implementation point of view, we propose to migrate the CEP core from Apache S4 to Apache Storm. The latter is gaining momentum and will be possibly the "winner" among the open source data streaming frameworks.
- We propose to add stream analytics tools such as SAMOA to the CEP core in order to enlarge the possibilities for automatic synthesis of context contents provided to end-users.
- We propose to provide a basis set of built-in virtual entities and associated contexts to end-users, re-using established ontologies when available, in order to ease the entity and context definition process for end-users.
- We propose to provide basic inference tools that can work with predefined entities in order to offer inferred knowledge as part of the context associated to any type of virtual entities.

We propose to provide an enhanced web-based user interfaces that allow end-users a simple access to the functionality.

4.3.6 Non Linear and Time Varying Dependent Processes

In sections 3.3.1.5 and 3.3.1.6 we have described an convenient framework to estimate multi-variate stochastic processes. However, in order to estimate the processes, after inferring their causal structure, we assume their dependency remains constant over time. Many real-world systems, instead, consist of processes interacting in a way that may change every time step. Therefore one challenge is addressing such processes and finding an appropriate metric to infer local causality. Although DI is proven to infer the causal structure between N-length processes, this structure may change dynamically over the time.

Another direction the research in is to improve the causal structure, on one hand, by increasing the estimation of DI accuracy in larger spaces, and the linear parameters of the ARX on the other hand.

4.4 IoT Architectures

The main challenges to the BUTLER architecture have been already identified and briefly outlined in section 6 in deliverable D3.2 [1]. However, it is obvious that a 3-year long project does not work in a closed environment, but must be opened to the inputs of the market, the industry and the society. It means that as the project evolves, the environment does it so and therefore, the partners have to be able to scout it in order to align their own developments with the evolution of the technology outlook.

As the result of the work done in the development and integration of the BUTLER during the third year of the project, we have come out with some conclusions related to the BUTLER architecture:

- The scenarios that BUTLER addresses (as any other in the real life) cannot deal exclusively with IoT, even if they are enabled by devices. Any IoT architecture applied to real scenarios has to deal not only with IoT architectures and standards but also consider functional components supporting functionalities that not even related to IoT. Architecture design must take a best-of-breed approach that considers not only IoT inputs.
- IoT-A[30] has created an Architectural Reference Model (ARM) [229]. Although IoT-A ARM blueprints, guidelines and design choices provide a superb input to any IoT-related project (in fact the ARM has been the main input to the BUTLER architectural work, especially with regard to BUTLER Information Model, see section 4.3 in deliverable D3.2 [1]), the fact that it does not provide an applied architecture makes it difficult to take an off-the-shelf architecture and use it as the basis of BUTLER's own architecture
- IoT is driven by the emergence of devices. However, the power of IoT lies on the ubiquity of devices and on the information they sense. Therefore, the emphasis put on the connectivity of devices cannot hide that IoT means also Data Management and that any IoT architecture must include a wide set of functional components dealing with data/context management.
- The large amount of SmartObjects that can be connected to any IoT environment, and the heterogeneity of the data the sense and generate, joined with the fact that many of the features exposed by any IoT environment rely on almost real-time reaction to events bring the need to process data streams as they are generated to the table. That is the reason to introduce a Complex Event Processing Functional Component (CEPFC) (which has been anticipated by FI-WARE [230]) within the Data/Context Management layer, as it supports the functionality of several functional components.
- Although it cannot be considered a finding or conclusion, as it was one of the main requirements and cornerstone of the project, security management must be handled across all the BUTLER architectural layers and not in silos.

As mentioned before, a number of challenges were listed and briefly described in deliverable D3.2 [1]. Although IoT (as any other area related to the Internet) is quickly evolving and approaching very fast to the inflection point of massive acceptance and deployment, and therefore it can involve the emergence of new architectural challenges, we think that the three relevant challenges we identified remain: **semantic support**, **big data and analytics support** and **fragmentation of standards and industry efforts**.

Semantics define a globally interpretable significance to data. Although not specifically tied to the IoT, adding semantics to it would allow data originated from different SmartObjects to be unambiguously accessible and processable across different domains. Semantic descriptions are particularly useful in M2M environments where a high level of autonomy is desired [231]. The industry and the standardization bodies are actively working in providing semantic support to IoT, however in vertical areas. For instance, the EC has sponsored, under the oneM2M umbrella, a study on "Available

[30] Internet of Things - Architecture. Web site at http://www.iot-a.eu/public.

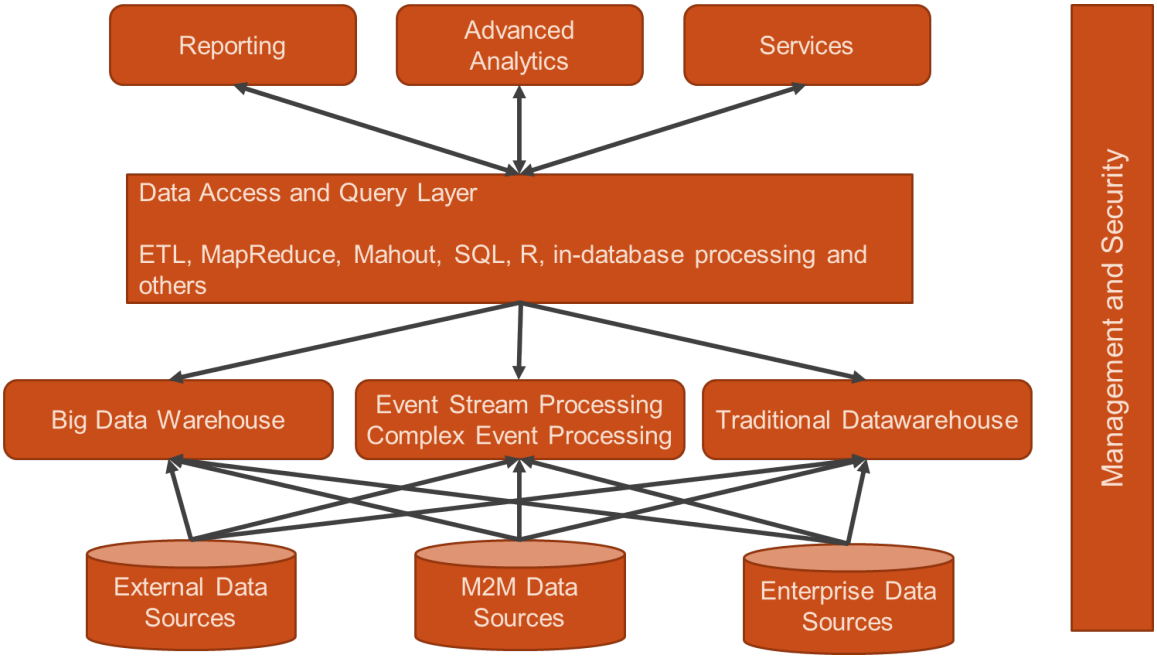


Figure 96: Analytics Architectural Overview [232].

semantics assets for the interoperability of Smart Appliances. Mapping into a common ontology as a M2M application layer semantics". It aims to propose a common ontology for smart appliances and will be due for February 2015.

Semantics has not been explicitly supported in BUTLER but it would provide straightforward advantages. For instance, exposition of SmartObject data would ease discovery and subsequent use of said data. The use of semantic technologies would also help to aggregate information from diverse sources and to extract knowledge from them. Thus, the introduction of **semantic support** in BUTLER would impact on the Data / Context Management and Services Layer.

Another obvious challenge would come from the **big data and analytics support** that could be added to BUTLER, especially considering mediated SmartObject data. As Holler et al. [232] highlight, some of the key characteristics of M2M data include, among several ones: Big Data ("huge amounts of data are generated, capturing detailed aspects of the processes where devices are involved"), Heterogeneous Data ("The data is produced by a huge variety of devices and is itself highly heterogeneous, differing on sampling rate, quality of captured data..."), Real-World Data ("The overwhelming majority of the M2M data relates to real-world processes and is dependent on the environment they interact with"), and Real-Time Data ("M2M data is generated in real-time and overwhelmingly can be communicated also in a very timely manner. The latter is of pivotal importance since many times their business value depends on the real-time processing of the information they convey"). Moreover, they highlight the importance of data analysis within the IoT technologies [232]. A proposal of an IoT analytics architecture can be seen in the Figure 96.

Although not explicitly focused on IoT, FI-WARE provides enablers for offline analytics (using the now commonplace Hadoop paradigm) [233] and complex event processing [230].

In the same line, the market is sending clear messages on the relevance of big data analytics, as applied to IoT. In September 2014, Goldman Sachs stated that "As the IoT will by definition generate voluminous amounts of unstructured data, the availability of big data analytics is a key enabler" [234]. Although BUTLER already provides some basic support in the form of the CEPFC, additional components should be added in order to provide both online and offline analytics. Thus, the challenge would be how to integrate big data analytics enablers into the BUTLER architec-

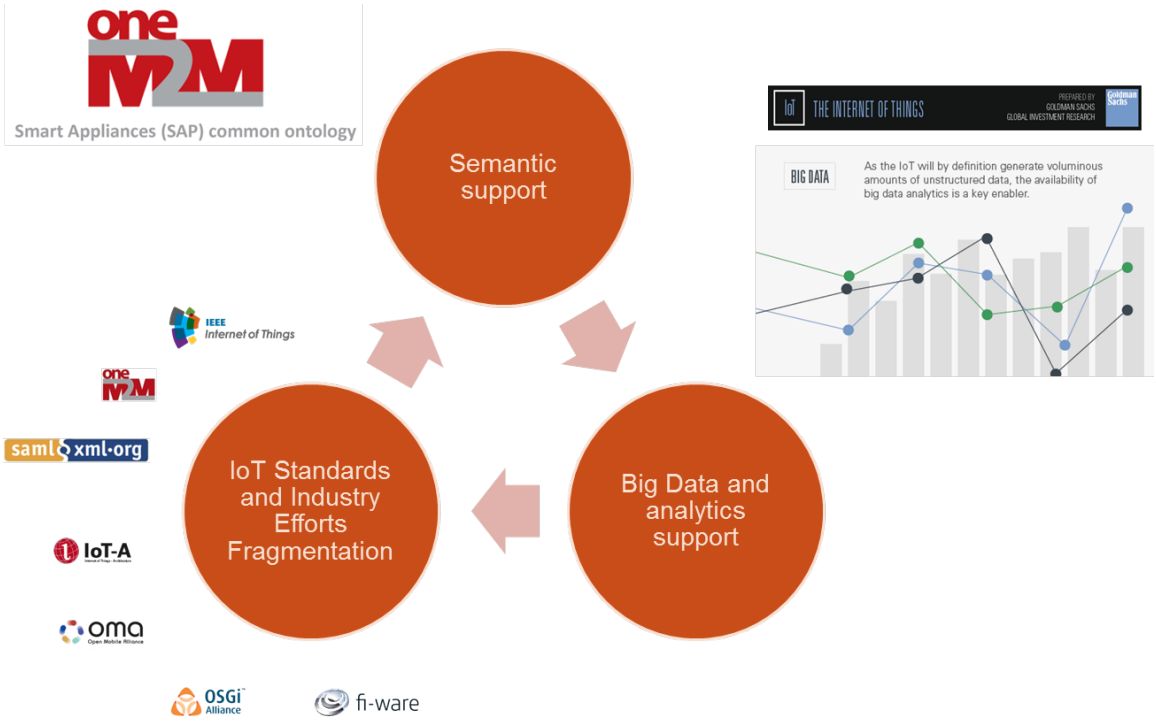


Figure 97: Open challenges in the BUTLER IoT architecture.

ture and also how to add online analytics support to massive SmartObject data mediated through BUTLER SmartServers.

Finally, another relevant challenge that the BUTLER Architecture would have to face is the increasing **fragmentation of standards**, industry efforts and EU projects related to IoT. As mentioned below, IoT-A has come out with an Architectural Reference Model (ARM) [229] that have been followed by BUTLER in order to device its own architecture. However, said ARM actually plays more the role of a meta-architecture than an actual applied architecture and therefore interoperability is far from being obtained. Compliancy between the BUTLER Architecture and other IoT-A-based architectures would have an obvious impact on the BUTLER architecture. Besides, there are already other additional activities in the IoT area in standards bodies (ETSI SmartM2M[31], oneM2M[32], TM Forum[33]...) and very recently IEEE has created the P2413 Work Group, Standard for an Architectural Framework for the IoT[34]. It is unclear what aspects P2413 intends to cover and if IEEE actually has the right (technical) competence profile to tackle an entire architecture, or if they intend to stay on the lower layers of the stack. Anyway, it means another player proposing its own architecture (members of the BUTLER consortium are currently negotiating to join the IEEE WG, so that the BUTLER Architecture will be provided to them as a valid input for architectural discussions). Anyway, BUTLER challenge is to keep on providing its requirements and architectural solutions to the aforementioned standardization bodies and, on the other hand, adapt its architecture to the solutions already being proposed by them. It is necessary to also acknowledge that each of said standardization bodies is stronger in a different area of the IoT field. For instance, oneM2M is more involved in the Network M2M APIs to the Gateways, while the IEEE P2413 WG is expected to have a stronger role in the device & capillary network side (thus, compliancy within each of the BUTLER Architectural Layers would have a different focus).

[31] ETSI SmartM2: http://portal.etsi.org/tb.aspx?tbid=726&SubTb=726
[32] oneM2M: http://www.onem2m.org/
[33] TM Forum IoT Hub: http://www.tmforum.org/IoT/16362/home.html
[34] Standard for an Architectural Framework for the Internet of Things (IoT): http://grouper.ieee.org/groups/2413/

As Figure 97 summarizes, the areas for growth and further development of the BUTLER architecture are clearly identified and fully aligned with both the IoT industry developments and the research and standardization efforts: **semantics support**, **big data analytics** and **standard alignment**.

5 Conclusions

This deliverable presented an overview of the advances achieved, within the BUTLER project, in the development of the integrated IoT enabling technologies. In addition, the discussion was extended to an outlook of the challenges opened by the current research trends and, consequently, to the future work expected for the next years.

Three main IoT enabling technologies, namely privacy and security, localization, and behavior modeling, were identified and the technical achievements associated with their development in the context of IoT were summarized. Such achievements were detailed by making use of extensive citation to both relevant scientific publications and their integrations in the BUTLER platform.

BUTLER has designed a Security Framework which is able to deal with the majority of IoT solutions, supporting an authorization paradigm and end-to-end security procedures between devices and applications. This report presented the results and issues of the different security technologies, prototypes and experimentation within this BUTLER Security Framework and introduced the main questions related to the implementation of the low level security and the application level security, as well as the problems of the security bootstrapping.

Wireless localization has suffered from multipath and NLoS propagation which negatively affect the reliability of data and measurements collected on the basis of wireless propagation. This document reported the description of some advanced algorithms able to mitigate the effects of multipath and NLoS conditions. Concretely, an accurate ranging algorithm using super-resolution techniques over phase measurements for distance estimation among devices was described along with a likewise accurate, robust and efficient positioning algorithm for target localization using algebraic confidence via circular interval scaling. Moreover, this report also analyzed the possibility of performing multipoint ranging using orthogonal set of Golomb rulers and presented novel hybrid cooperative positioning algorithms and cooperative NLoS detection techniques for challenging environments.

Behavior modeling and synthesis has brought out a wide variety of novel applications in IoT context-aware networks with a potential for exhibiting the progress of sophisticated intelligent behaviors. This document found space for the description of the algorithms and tools originally developed/enhanced in the scope of the horizontal BUTLER architecture and aimed at behavior modeling. Similarly, the advanced integration between deterministic and probabilistic modelling of human behaviors was presented too. In this research field the attention is also focused on some work on contextual networking, which pays more attention to mobility support for masses by providing context-aware adaptation of various networking mechanisms, and the user perspective context synthesis and management, which is on the basis of CEP to define user contexts of user's interest even without much technical skills.

After over-viewing the work actually done, the report concentrated on the challenges posed by integration of the mentioned enabling technologies in the IoT and possible future developments were discussed and envisioned. For what concerns privacy and security, aspects related to business and marketing were primarily considered, starting from several problems like the security deployment, the security open issues related to 6LoWPAN, the devices security, the security of the physical layer and, finally, some technologies related to the secure generation of credentials. Moving the focus on localization, the open challenges behind such types of systems such as accuracy, precision, scalability, complexity, robustness, cost and required power, were faced and the fundamental limit of localization error associated with each node within WSN context was analyzed. Finally, in the field of behavior modelling and synthesis, potential interest was foreseen for the enhancement of current algorithms and the distributed systems realizing these algorithms under the challenges of big data and 'soft' user identification, with the aim to enable effective knowledge transfers across smart domains and users based on the improved algorithms, modeling, verification and validation methods.

Basically, the floating result of the investigation analysis carried out in this report is that the BUTLER project demonstrated the strategic importance and centrality of the identified enabling technologies in the context of IoT field. More than a mere list of achievements around the presented issues, the main contribution of this document consists of providing new hints and opening new ground for future developments that it will be possible to address in next, original and ambitious, research projects.

References

[1] BUTLER Consortium, “D3.2 - Integrated System Architecture and Initial Pervasive BUTLER proof of concept,” October 2013.

[2] N. Koblitz, “Elliptic curve cryptosystems,” in *Mathematics of Computation 48, (177): 203-209*, 1987.

[3] V. Miller, “Use of elliptic curves in cryptography,” in *CRYPTO 85: 417-426*, 1985.

[4] A. Liu and P. Ning, “Tinyecc: A configurable library for elliptic curve cryptography in wireless sensor network,” in *7th International Conference on Information Processing in Sensor Networks (ISPN 2008), SPOTS Track, p245-256*, 2008.

[5] Certicom Research, “Standards for efficient cryptography - sec1: Elliptic curve cryptography,” 2000. [Online]. Available: http://www.secg.org/download/aid-385/sec1_final.pdf

[6] ——, “Standards for efficient cryptography - sec2: Recommended elliptic curve domain parameters,” 2000. [Online]. Available: http://www.secg.org/collateral/sec2_final.pdf

[7] S. Raza, T. Chung, S. Duquennoy, D. Yazar, T. Voigt, and U. Roedig, “Securing internet of things with lightweight ipsec,” in *SICS Technical Report*, 2012.

[8] S. Raza, T. Voigt, and V. Jutvik, “Lightweight ikev2: A key management solution for both compressed ipsec and ieee 802.15.4 security,” in *IETF Workshop on Smart Objects Security, Paris, France*, 2012.

[9] O. Bergmann, “Tinydtls.”

[10] BUTLER Consortium, “D5.1 - BUTLER Platforms and Pervasive Functionalities,” March 2014.

[11] V. Jutvik, “IPSec implementation for Contiki,” https://github.com/vjutvik/Contiki-IPsec, 2014, [Online; accessed 1-July-2014].

[12] “Calipso: Connect all ip-based smart objects!” [Online]. Available: http://www.ict-calipso.eu/

[13] A. Steffen, “The OpenSource IPsec-based VPN Solution,” http://www.strongswan.org/, 2011, [Online; accessed 2-July-2014].

[14] ZigBee Specification, “Document 053474r17,” January 2008.

[15] J. Voris, N. Saxena, and T. Halevi, “Accelerometers and randomness: perfect together,” in *Fourth ACM Conference on Wireless Network Security (WiSec)*, June 2011, pp. 115–126.

[16] A. Francillon and C. Castelluccia, “Tinyrng: A cryptographic random number generator for wireless sensor network nodes,” in *Symposium on Modeling and Optimization in Mobile, Ad Hoc and Wireless Networks, IEEE WiOpt*, April 2007.

[17] C. Hennebert, H. Hossayni, and C. Lauradoux, “Entropy harvesting from physical sensors,” in *6th ACM Conference on Security and Privacy in Wireless and Mobile Networks (WiSec)*, April 2013.

[18] ——, “Entropy from wireless statistics,” in *EuCNC 2014*, June 2014.

[19] T. Q. Duong, D. B. da Costa, K. J. Kim, K.-H. S. Liu, and V. N. Q. Bao, “Secure physical layer communications,” Special Issue of the IET Communications, 2014.

[20] A. D. Wyner, “The wire-tap channel,” *Bell System Technical Journal*, vol. 54, no. 8, pp. 1355 –1367, Oct. 1975.

[21] L. Y. Cheong and M. Hellman, “The gaussian wire-tap channel,” *IEEE Trans. Inform. Theory*, vol. 24, no. 4, pp. 451 – 456, Jul. 1978.

[22] I. Csiszár and J. Körner, “Broadcast channels with confidential messages,” *IEEE Trans. Inform. Theory*, vol. 24, no. 3, pp. 339 – 348, May 1978.

[23] P. K. Gopala, L. Lai, and H. El-Gamal, “On the secrecy capacity of fading channels,” *IEEE Trans. Inform. Theory*, vol. 54, no. 10, pp. 4687 – 4698, Oct. 2008.

[24] M. Bloch, J. Barros, M. R. D. Rodrigues, and S. W. McLaughlin, "Wireless information-theoretic security," *IEEE Trans. Inform. Theory*, vol. 54, no. 6, pp. 2515 – 2534, Jun. 2008.

[25] M. Haenggi, "A geometric interpretation of fading in wireless networks: Theory and applications," *IEEE Trans. Inform. Theory*, vol. 54, no. 12, pp. 5500 – 5510, Dec. 2008.

[26] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousee, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications.*, vol. 27, no. 7, pp. 1029 – 1046, Sep. 2009.

[27] P. C. Pinto, J. Barros, and M. Z. Win, "Secure communication in stochastic wireless networks—part i: Connectivity," *IEEE Trans. Information Forensics and Security*, vol. 7, no. 1, pp. 125 – 138, Feb. 2012.

[28] P. Gupta and P. Kumar, "The capacity of wireless networks," *IEEE Trans. Information Theory.*, vol. 46, no. 2, pp. 388– 404, 2000.

[29] S. Weber, J. G. Andrews, and N. Jindal, "An overview of the trans- mission capacity of wireless networks," *IEEE Trans. Communications.*, vol. 58, no. 12, Dec. 2010.

[30] O. O. Koyluoglu, C. E. Koksal, and H. E. Gamal, "On secrecy capacity scaling in wireless networks," *IEEE Trans. Information Theory.*, vol. 58, no. 5, pp. 3000 – 3015, May. 2012.

[31] S. Goel, V. Aggarwal, A. Yener, and A. R. Calderbank, "The effect of eavesdroppers on network connectivity: A secrecy graph approach," *IEEE Trans. Information Forensics and Security*, vol. 6, no. 3, pp. 712 – 724, Sep. 2011.

[32] S. Vuppala and G. Abreu, "Secrecy rate, outage and transmission capacity of random networks with colluding eavesdroppers," submitted to IEEE Trans. Information Forensics and Security.

[33] X. Zhou, R. K. Ganti, J. G. Andrews, and A. Hjørungnes, "On the throughput cost of physical layer security in decentralized wireless networks," *IEEE Trans. Wireless Comm.*, vol. 10, no. 8, pp. 2764–2775, Aug. 2011.

[34] Z. Shu, Y. L. Yang, Y. Qian, and R. Q. Hu, "Impact of interference on secrecy capacity in a cognitive radio network," in *Proc. IEEE Global Telecommunications Conference (Globecom'11)*, 2011.

[35] D. Daley and D. V. Jones, *An introduction to the theory of point processes.* NewYork: Springer, 1988.

[36] S. Vuppala and G. Abreu, "Secrecy outage in random wireless networks subjected to fading," in *Proc. IEEE Personal Indoor Mobile Radio Communication*, London, UK., Sept8-11 2013, pp. 441 – 445.

[37] H. A. David and H. N. Nagaraja, *Order Statistics*. Wiley, 2003.

[38] S. Vuppala and G. Abreu, "Unicasting on the secrecy graph," *IEEE Trans. Information Forensics and Security*, vol. 8, no. 9, pp. 1469 – 1481, Sep. 2013.

[39] S. Akoum and R. W. Heath, "Interference coordination: Random clustering and adaptive limited feedback," *IEEE Trans. Signal Processing*, vol. 61, no. 7, pp. 1822–1834, Apr. 2013.

[40] M. Haenggi, "Mean interference in hard-core wireless networks," *IEEE Commun. Lett.*, vol. 15, no. 8, pp. 792–794, Aug. 2011.

[41] H. Jeon, N. Kim, J. Choi, H. Lee, and J. Ha, "Bounds on secrecy capacity over correlated ergodic fading channels at high snr," *IEEE Trans. Inform. Theory*, vol. 57, no. 4, pp. 1975–1983, April 2011.

[42] J. Zhu, X. Jiang, O. Takahashi, and N. Shiratori, "Effects of channel correlation on outage secrecy capacity," *Journal of Information Processing*, vol. 21, no. 4, pp. 640–649, Oct. 2013.

[43] W. Liu, S. Vuppala, G. Abreu, and T. Ratnarajah, "Secrecy outage in correlated nakagami fading," June 2014, submitted to IEEE Sensor Array and Multichannel Signal Processing Workshop.

[44] S. Vuppala, W. Liu, , T. Ratnarajah, and G. Abreu, "Secrecy outage analysis of cognitive wireless sensor networks," June 2014, submitted to IEEE Sensor Array and Multichannel Signal Processing Workshop.

[45] S. Binnikov and R. Moessner, "Expansions for nearly gaussian distributions," *Astron. Astrophys. Suppl. Ser.*, no. 130, pp. 193–205, Oct. 1998.

[46] A. Rabbachin, M. Z. Win, and A. Conti, "Interference engineering for network secrecy in nakagami fading channels," in *Proc. IEEE International Conference on Communications*, 2013.

[47] M. Z. Win, A. Rabbachin, J. Lee, and A. C. and, "Cognitive network secrecy with interference engineering," *IEEE Networks*, 2014.

[48] S. Vuppala and G. Abreu, "Secrecy transmission capacity of random networks," in *IEEE Fourty-Seventh Asilomar Conference Conference on Signals, Systems, and Computers*, 2013.

[49] P. C. Pinto, J. Barros, and M. Z. Win, "Secure communication in stochastic wireless networks—part ii: Maximum rate and collusion," *IEEE Trans. Information Forensics and Security.*, vol. 7, no. 1, pp. 139 – 147, Feb. 2012.

[50] X. Zhou, R. K. Ganti, and J. G. Andrews, "Secure wireless network connectivity with multi-antenna transmission," *IEEE Trans. on Networking.*, vol. 10, no. 2, pp. 425 – 430, Feb. 2011.

[51] H. Wang, X. Zhou, and M. C. Reed, "Physical layer security in cellular networks: Physical layer security in cellular networks: Physical layer security in cellular networks:a stochastic geometry approach," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2776–2787, June 2013.

[52] H. ElSawy, E. Hossain, and M. Haenggi, "Stochastic geometry for modeling, analysis, and design of multi-tier and congnitive cellular wireless networks:a survey," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 3, pp. 996 – 1019, July 2013.

[53] P. C. Pinto and M. Z. Win, "Communication in a Poisson field of interferers – part i: Interference distribution and error probability," *IEEE Trans. Wireless Commun.*, (to appear).

[54] P. Neelakanta, "Designing robust wireless communications for factory floors," in *Industrial Informatics, 2006 IEEE International Conference on*, August 2006, pp. xxviii–xxix.

[55] S. K. Timalsina, R. Bhusal, and S. Moh, "NFC and its application to mobile payment: Overview and comparison," in *Information Science and Digital Content Technology (ICIDT), 2012 8th International Conference on*, June 2012, pp. 203– 206.

[56] S. Biswas, R. Tatchikou, and F. Dion, "Vehicle-to-vehicle wireless communication protocols for enhancing highway traffic safety," *Communications Magazine, IEEE*, vol. 44, no. 1, pp. 74 – 82, January 2006.

[57] L. Ying-Chang, S. Sumei, P. Peng, and F. Chin, "Tutorial 2: Emerging wireless standards for wran, wifi, wimedia and zigbee," in *Communication systems, (ICCS) 10th IEEE Singapore International Conference on*, October 2006, pp. nil27 – nil29.

[58] D. Jun, Z. Rongqing, S. L. H. Zhu, and J. Bingli, "Truthful mechanisms for secure communication in wireless cooperative system," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 9, pp. 4236 – 4245, September 2013.

[59] G. Pasolini, D. Dardari, G. T. F. de Abreu, and S. Severi, "The effect of channel spatial correlation on physical layer security in multi-antenna scenarios," in *IEEE Fourty-Seventh Asilomar Conference on Signals, Systems and Computers, (Asilomar 2013)*, November 2013, to appear.

[60] K. Ren, H. Su, and Q. Wang, "Secret key generaton exploiting channel characteristics in wireless communications," *IEEE Wireless Communications*, vol. 18, no. 4, pp. 6–12, August 2011.

[61] H. Liu, J. Yang, Y. Wang, and Y. Chen, "Collaborative secret key extraction leveraging received signal strength in mobile wireless network," in *31st Annual IEEE International Conference on Computer Communications (INFOCOM)*, March 2012, pp. 927 – 935.

[62] N. Patwari, J. Croft, S. Jana, and S. K. Kasera, "High-rate uncorrelated bit extraction for shared secret key generation from channel measurements," *IEEE Transactions on Mobile Computing*, vol. 9, no. 1, pp. 17 – 30, Jan. 2010.

[63] Q. Wang, K. Xu, and K. Ren, "Cooperative secret key generation from phase estimation in narrowband fading channels," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 9, pp. 1666–1674, October 2012.

[64] A. Sayeed and A. Peerig, "Secure wireless communications: Secret keys through multipath," in *International Conference on Acoustic, Speech and Signal Processing (ICASSP)*, March 2008, pp. 3013–3016.

[65] K. Xu, Q. Wang, and K. Ren, "Wireless key establishment with asynchronous clocks," in *Military Communications Conference (Milcom)*, November 2011, pp. 1410– 1415.

[66] S. T. B. Hamida, J.-B. Pierrot, and C. Castelluccia, "An adaptive quantization algorithm for secret key generation using radio channel measurements," in *New Technologies, Mobility and Security (NTMS), 3rd International Conference on*, December 2009, pp. 1– 5.

[67] Q. Wang, H. Su, K. Ren, and K. Kim, "Fast and scalable secret key generation exploiting channel phase randomness in wireless networks," in *30th IEEE International Conference on Computer Communications (INFOCOM)*, 2011, pp. 1422– 1430.

[68] K. Hassan and W. Henkel, "Unequal error protection with eigen beamforming for partial channel information mimo-ofdm," in *Sarnoff Symposium, 2007 IEEE*, April 2007, pp. 1–5.

[69] G. T. F. de Abreu, "On the generation of Tikhonov variates," *Communication, IEEE Transactions on*, vol. 56, no. 7, pp. 1157–1168, July 2008.

[70] H. Fu and P. Y. Kam, "Exact phase noise model and its application to linear minimum variance estimation of frequency and phase of a noisy sinusoid," in *Personal, Indoor and Mobile Radio Communications, (PIMRC). IEEE 19th International Symposium on*, September 2008, pp. 1–5.

[71] E. Ip, "Coherent detection and digital signal processing for fiber optic communications," Ph.D. dissertation, Stanford University, 2009.

[72] D. E. Amos, "Computation of modified bessel functions and their ratios," *Mathematics of Computation*, vol. 28, no. 125, pp. 239–251, January 1974.

[73] S. Tmar-Ben Hamida, J.-B. Pierrot, and C. Castelluccia, "An Adaptive Quantization Algorithm for Secret Key Generation Using Radio Channel Measurements," in *Proc. NTMS'09*, Cairo, Egypt, Dec. 2009.

[74] N. Patwari, J. Croft, S. Jana, and S. K. Kasera, "High-Rate Uncorrelated Bit Extraction for Shared Secret Key Generation from Channel Measurements," *IEEE Trans. on Mobile Computing*, vol. 9, no. 1, pp. 17–30, Jan. 2010.

[75] I. Tunaru, B. Denis, and B. Uguen, "Random Patterns of Secret Keys from Sampled IR-UWB Channel Responses," in *Proc. ICUWB'14*, Paris, France, Sept. 2014.

[76] ——, "Public Discussion Strategies for Secret Key Generation from Sampled IR-UWB Channel Responses," in *Proc. COMM'14*, Bucharest, Romania, May 2014.

[77] A. Molisch, D. Cassioli, C.-C. Chong, S. Emami, A. Fort, B. Kannan, J. Karedal, J. Kunisch, H. Schantz, K. Siwiak, and M. Win, "A Comprehensive Standardized Model for Ultrawideband Propagation Channels," *IEEE Trans. on Antennas and Propagation*, vol. 54, no. 11, pp. 3151–3166, Nov. 2006.

[78] N. Amiot, M. Laaraiedh, and B. Uguen, "PyLayers: An Open Source Dynamic Simulator for Indoor Propagation and Localization," in *Proc. IEEE ICC'13*, Budapest, Hungary, Jun. 2013.

[79] S. Bradner, "The end of end-to-end security? [internet security]," *Security Privacy, IEEE*, vol. 4, no. 2, pp. 76–79, March 2006.

[80] "Oauth 2.0." [Online]. Available: http://oauth.net/2/

[81] "Bespoon." [Online]. Available: http://spoonphone.com/en/

[82] "Decawave." [Online]. Available: http://decawave.com/

[83] C. Hennebert and J. D. Santos, "Security Protocols and Privacy Issues into 6LoWPAN Stack: A Synthesis," *IEEE Journal of Internet of Things*, vol. 1, no. 5, Sept. 2014.

[84] D. Macagnano, G. Destino, and G. Abreu, "A comprehensive tutorial on localization: Algorithms and performance analysis tools," *International Journal of Wireless Information Networks*, vol. 19, no. 4, pp. 290–314, July 2012. [Online]. Available: http://dx.doi.org/10.1007/s10776-012-0190-4

[85] BUTLER Consortium, "D2.2 - Requirements, Specifications, Localization and Context-acquisition for IoT Context-aware Networks," October 2012.

[86] P. Harrop and R. Das, "Wireless sensor networks (WSN) 2012-2022: Forecasts, technologies, players - the new market for ubiquitous sensor networks (USN)," Dec., 2012. [Online]. Available: www.IDTechEx.com/ips

[87] L. Wirola, T. Laine, and J. Syrjrinne, "Mass-market requirements for indoor positioning and indoor navigation," in *Indoor Positioning and Indoor Navigation (IPIN), 2010 International Conference on*, 2010, pp. 1–7.

[88] P. Harrop and R. Das, "Mobile phone indoor positioning systems (IPS) and real time locating systems (RTLS) 2014-2024 - forecasts, players, opportunities," July, 2013. [Online]. Available: www.IDTechEx.com/ips

[89] S. Azzouzi, M. Cremer, U. Dettmar, T. Knie, and R. Kronberger, "Improved AoA based localization of UHF RFID tags using spatial diversity," in *IEEE International Conference on RFID-Technologies and Applications (RFID-TA'11)*, 2011, pp. 174–180.

[90] T. E. Tuncer and B. Friedlander, *Classical and Modern Direction-of-Arrival Estimation*. Elsevier Science, 2009. [Online]. Available: http://books.google.de/books?id=1aQbxKJI2CsC

[91] M. Scherhaufl, M. Pichler, E. Schimback, D. J. Muller, A. Ziroff, and A. Stelzer, "Indoor localization of passive UHF RFID tags based on phase-of-arrival evaluation," *IEEE Trans. on Microwave Theory and Techniques*, vol. 61, no. 12, pp. 4724–4729, 2013.

[92] A. Barabell, "Improving the resolution performance of eigenstructure- based direction-finding algorithms," in *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '83)*, vol. 8, 1983, pp. 336–339.

[93] O. Oshiga, S. Severi, and G. T. F. de Abreu, "Optimized super-resolution ranging over ToA measurements," in *IEEE Wireless Communications and Networking Conference (WCNC'14)*, 2014.

[94] N. N. Tayem, "2d DOA estimation of multiple coherent sources using a new antenna array configuration," in *Proc. IEEE The 46^{th} Asilomar Conference on Signal, Systems and Computers (ASILOMAR'12)*, 2012, pp. 212–216.

[95] W. T. Rankin, "Optimal golomb rulers: An exhaustive parallel search implementation," Ph.D. dissertation, Duke University, 1993.

[96] A. H. Dewdney, "Computer recreations," *Scientific American Magazine*, pp. 16–26, Dec. 1985.

[97] O. Oshiga, S. Severi, and G. T. F. de Abreu, "Superresolution multipoint ranging with optimized sampling via orthogonally designed golomb rulers," *IEEE Trans. on Wireless Communications*, submitted for publication.

[98] G. T. F. de Abreu, "On the generation of tikhonov variates," *IEEE Trans. on Communications*, vol. 56, no. 7, pp. 1157–1168, July 2008.

[99] C. Chien-Sheng, C. Yi-Jen, J.-M. Lin, and L. Chi-Hsien, "Geometrical positioning schemes for MS location estimation," in *IEEE International Symposium on Computer, Consumer and Control (IS3C'12)*, June 2012, pp. 487–490.

[100] G. T. F. de Abreu and G. Destino, "Super MDS: Source location from distance and angle information," in *IEEE Wireless Communications and Networking Conference (WCNC'07)*, 2007, pp. 4430–4434.

[101] H. Wymeersch, J. Lien, and M. Win, "Cooperative localization in wireless networks," *Proceedings of the IEEE*, vol. 97, no. 2, pp. 427 –450, Feb 2009.

[102] J. Saloranta, D. Macagnano, and G. Abreu, "Interval-scaling for multitarget localization," in *Positioning Navigation and Communication (WPNC), 2012 9th Workshop on*, Mar 2012.

[103] W. Torgerson, "Multidimensional scaling 1: Theory and method," *Psychometrica*, vol. 17, pp. 401 – 419, 1952.

[104] D. Macagnano and G. de Abreu, "Gershgorin analysis of random gramian matrices with application to mds tracking," *IEEE Transactions on Signal Processing*, vol. 59, no. 4, pp. 1785 –1800, Apr 2011.

[105] I. Borg and P. Groenen, *Modern multidimensional scaling: theory and applications*, ser. Springer series in statistics. Springer, 1997.

[106] A. Beck, P. Stoica, and J. Li, "Exact and approximate solutions of source localization problems," *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1770 –1778, May 2008.

[107] P. J. F. Groenen, *The Majorization Approach to Multidimensional Scaling: Some Problems and Extensions*. DSWO Press, Leiden University, 1993.

[108] T. Cox and M. Cox, *Multidimensional scaling*, ser. Monographs on statistics and applied probability. Chapman & Hall/CRC, 2001.

[109] J. de Leeuw, *Some Majorization Techniques*. Department of Statistics Papers, UCLA., 2006. [Online]. Available: http://www.escholarship.org/uc/item/1kp3t79r

[110] G. J. Miel, "Majorizing sequences and error bounds for iterative methods," *Mathematics of Computation*, vol. 34, no. 149, pp. 185–202, 01 1980. [Online]. Available: http://www.jstor.org/stable/2006227

[111] P. J. F. Groenen, S. Winsberg, O. Rodríguez, and E. Diday, "I-scal: Multidimensional scaling of interval dissimilarities," *Comput. Stat. Data Anal.*, vol. 51, pp. 360–378, Nov 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1647967.1648294

[112] D. Torrieri, "Statistical theory of passive location systems," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-20, no. 2, pp. 183 –198, Mar 1984.

[113] S. M. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1993.

[114] N. Patwari, J. Ash, S. Kyperountas, I. Hero, A.O., R. Moses, and N. Correal, "Locating the nodes: Cooperative localization in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 54 – 69, Jul 2005.

[115] G. Destino and G. Abreu, "On the maximum likelihood approach for source and network localization," *IEEE Transactions on Signal Processing*, vol. 59, no. 10, pp. 4954 –4970, Oct 2011.

[116] M. Spirito, "On the accuracy of cellular mobile station location estimation," *IEEE Transactions on Vehicular Technology*, vol. 50, no. 3, pp. 674–685, 2001.

[117] N. Patwari, A. Hero, M. Perkins, N. Correal, and R. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003.

[118] J. A. Costa, N. Patwari, and A. O. Hero, III, "Distributed weighted-multidimensional scaling for node localization in sensor networks," *ACM Trans. Sen. Netw.*, vol. 2, pp. 39–64, Feb 2006. [Online]. Available: http://doi.acm.org/10.1145/1138127.1138129

[119] H. David and H. Nagaraja, *Order Statistics*, 3rd ed., ser. Wiley Series in Probability and Statistics. Wiley, 2004.

[120] S. Nadarajah and S. Kotz, "Exact distribution of the max/min of two gaussian random variables," *Very Large Scale Integration (VLSI) Systems, IEEE Transactions on*, vol. 16, no. 2, pp. 210–212, 2008.

[121] A. Azzalini, "The skew-normal distribution and related multivariate families," *Scandinavian Journal of Statistics*, vol. 32, no. 2, pp. 159–188, May 2005.

[122] P. Forero and G. Giannakis, "Sparsity-exploiting robust multidimensional scaling," *IEEE Transactions on Signal Processing*, vol. 60, no. 8, pp. 4118–4134, 2012.

[123] P. J. Huber, "Robust regression: asymptotics, conjectures and monte carlo," *The Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.

[124] BUTLER Consortium, "D2.4 - Selected Technologies for the BUTLER Platform," April 2013.

[125] Z. Xiong, F. Sottile, R. Garello, and C. Pastone, "A Cooperative NLoS Identification and Positioning Approach in Wireless Networks," in *proc. ICUWB 2014*, Sept. 2014.

[126] S. Van de Velde, H. Wymeersch, and H. Steendam, "Comparison of message passing algorithms for cooperative localization under nlos conditions," in *9th Workshop on Positioning Navigation and Communication (WPNC)*, Mar. 2012, pp. 1–6.

[127] H. Wymeersch, S. Marano, W. M. Gifford, and M. Z. Win, "A Machine Learning Approach to Ranging Error Mitigation for UWB Localization," *IEEE Transactions on Communications*, vol. 60, no. 6, pp. 1719–1128, June 2012.

[128] I. Guvenc, C.-C. Chong, F. Watanabe, and H. Inamura, "NLOS Identification and Weighted Least-Squares Localization for UWB Systems Using Multipath Channel Statistics," *EURASIP Journal on Advances in Signal Processing*, no. 1, 2008.

[129] T. Minka, "Expectation propagation for approximate bayesian inference," in *17th Conference in Uncertainty in Artificial Intelligence*, Aug. 2001, pp. 362–369.

[130] F. Sottile, H. Wymeersch, M. A. Caceres, and M. A. Spirito, "Hybrid GNSS-terrestrial cooperative positioning based on particle filter," in *IEEE Global Telecommunications Conference (GLOBECOM)*, Dec. 2011, pp. 1–5.

[131] M. Caceres, F. Penna, H. Wymeersch, and R. Garello, "Hybrid cooperative positioning based on distributed belief propagation," *Selected Areas in Communications, IEEE Journal on*, vol. 29, no. 10, pp. 1948–1958, December 2011.

[132] D. Heckerman, "A tutorial on learning with bayesian networks," Microsoft Research, Tech. Rep. MSR-TR-95-06, 1995.

[133] A. Doucet, "On Sequential Monte Carlo Methods for Bayesian filtering," University of Cambridge, Cambridge, UK, Tech. Rep. CB2 1PZ, 1998.

[134] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 174–188, Feb 2002.

[135] F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund, "Particle filters for positioning, navigation, and tracking," *Signal Processing, IEEE Transactions on*, vol. 50, no. 2, pp. 425–437, Feb 2002.

[136] BUTLER Consortium, "D2.3 - Requirements, Specifications and Behavioral Modelling and Synthesis Technologies for IoT Context-Aware Networks," October 2012.

[137] A. Ramakrishnan, D. Preuveneers, and Y. Berbers, "A loosely coupled and distributed bayesian framework for multi-context recognition in dynamic ubiquitous environments," in *IEEE 10th International Conference on Ubiquitous Intelligence and Computing*, Dec 2013, pp. 270–277.

[138] ——, "Enabling self-learning in dynamic and open iot environments," in *accepted at 5th International Conference on Ambient Systems, Networks and Technologies (ANT-2014)*, June 2014.

[139] S. N. Z. Naqvi, A. Ramakrishnan, D. Preuveneers, and Y. Berbers, "Walking in the clouds: deployment and performance trade-offs of smart mobile applications for intelligent environments," in *Proceedings of the 9th International Conference on Intelligent Environments (IE13)*. IEEE Computer Society, July 2013, pp. 212–219. [Online]. Available: https://lirias.kuleuven.be/handle/123456789/411259

[140] A. Ramakrishnan, S. N. Z. Naqvi, Z. W. Bhatti, D. Preuveneers, and Y. Berbers, "Learning deployment trade-offs for self-optimization of Internet of Things applications," in *Proceedings of the 10th International Conference on Autonomic Computing, ICAC 2013, ICAC '13, the 10th International Conference on Autonomic Computing, San Jose, CA, U.S.A., 26-28 June 2013*. ACM, Jun. 2013, pp. 213–224. [Online]. Available: https://lirias.kuleuven.be/handle/123456789/411273

[141] A. Ramakrishnan, D. Preuveneers, and Y. Berbers, "A bayesian framework for life-long learning in context-aware mobile applications," in *Context in Computing*, ser. Lecture Notes in Computer Science, P. Brezillon and A. Fonzalez, Eds. Springer, 2014 (To be published).

[142] R. Agrawal and R. Srikant, "Fast algorithms for mining association rules in large databases," in *Proceedings of the 20th International Conference on Very Large Data Bases*, ser. VLDB '94, 1994, pp. 487–499.

[143] N. Vasconcelos and A. Lippman, "A unifying view of image similarity," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 1, 2000, pp. 38–41 vol.1.

[144] N. Z. Naqvi, D. Preuveneers, W. Meert, and Y. Berbers, "The right thing to do: Automating support for assisted living with dynamic decision networks," in *Ubiquitous Intelligence and Computing, 2013 IEEE 10th International Conference on and 10th International Conference on Autonomic and Trusted Computing (UIC/ATC)*. IEEE, 2013, pp. 262–269.

[145] R. Ashley, C. W. Granger, and R. Schmalensee, "Advertising and aggregate consumption: an analysis of causality," *Econometrica: Journal of the Econometric Society*, pp. 1149–1167, 1980.

[146] J. Armstrong, "Illusions in regression analysis," *Available at SSRN 1969740*, 2011.

[147] P. Spirtes, C. Glymour, and R. Scheines, *Causation, Prediction, and Search*. The MIT Press, 2000, vol. 81.

[148] J. Aldrich, "Correlations genuine and spurious in pearson and yule," *Statistical Science*, pp. 364–376, 1995.

[149] G. U. Yule, "On the time-correlation problem, with especial reference to the variate-difference correlation method," *Journal of the Royal Statistical Society*, vol. 84, no. 4, pp. 497–537, 1921.

[150] ——, "An introduction to the theory of statistics," *London,*, vol. 346, 1922.

[151] C. Quinn, T. Coleman, N. Kiyavash, and N. Hatsopoulos, "Estimating the directed information to infer causal relationships in ensemble neural spike train recordings," *Journal of Computational Neuroscience*, vol. 30, pp. 17–44, 2011, 10.1007/s10827-010-0247-2.

[152] H. Marko, "The bidirectional communication theory–a generalization of information theory," *Communications, IEEE Transactions on*, vol. 21, no. 12, pp. 1345 – 1351, dec 1973.

[153] L. A. Baccalá and K. Sameshima, "Partial directed coherence: a new concept in neural structure determination," *Biological Cybernetics*, 2001.

[154] Y. Saito, H. Harashima, N. Yamaguchi, and K. Fujisawa, "Recent advances in eeg and emg data processing," *Chap. Tracking of information within multichannel EEG record-causal analysis in EEG*, pp. 133–146, 1981.

[155] J. L. Massey, "Causality, feedback and directed information," 1990.

[156] G. Kramer, "Directed information for channels with feedback," Ph.D. Dissertation, Swiss Federal Institute of Technology (ETH) Zurich, 1998.

[157] S. A. Abdallah, M. D. Plumbley, and Q. Mary, "Information dynamics: patterns of expectation and surprise in the perception of music," *Connection Science*, vol. 21, no. 2, pp. 89 – 117, 2009.

[158] ——, "Predictive information, multiinformation and binding information," Technical Report C4DMTR-10-10, Queen Mary University of London, Tech. Rep., 2010.

[159] J. Pearl, *Causality: models, reasoning and inference*. Cambridge Univ Press, 2000, vol. 29.

[160] M. Pinsky and S. Karlin, *An introduction to stochastic modeling*. Academic press, 2010.

[161] G. A. Darbellay and I. Vajda, "Estimation of the information by an adaptive partitioning of the observation space," *Information Theory, IEEE Transactions on*, vol. 45, no. 4, pp. 1315–1321, 1999.

[162] The MathWorks Inc , *MATLAB and Statistics Toolbox Release 2012b*, Natick, Massachusetts, United States, 2012.

[163] Q. Zhang and A. Benveniste, "Wavelet networks," *Neural Networks, IEEE Transactions on*, vol. 3, no. 6, pp. 889–898, 1992.

[164] A. Neumaier and T. Schneider, "Estimation of parameters and eigenmodes of multivariate autoregressive models," *ACM Transactions on Mathematical Software (TOMS)*, vol. 27, no. 1, pp. 27–57, 2001.

[165] H. Lütkepohl, *New introduction to multiple time series analysis*. Springer, 2007.

[166] M. Feki, F. Kawsar, M. Boussard, and L. Trappeniers, "The Internet of Things: The Next Technological Revolution," *Computer*, vol. 46, no. 2, pp. 24–25, 2013.

[167] H. Ning, H. Liu, and L. Yang, "Cyberentity Security in the Internet of Things," *Computer*, vol. 46, no. 4, pp. 46–53, 2013.

[168] E. Tapia, S. Intille, and K. Larson, "Activity Recognition in the Home Using Simple and Ubiquitous Sensors," in *Pervasive Computing*, ser. Lecture Notes in Computer Science, A. Ferscha and F. Mattern, Eds. Springer Berlin Heidelberg, 2004, vol. 3001, pp. 158–175.

[169] C. Berge, *Graphs and Hypergraphs*. Elsevier Science Ltd, 1985.

[170] X. Jiang, Y. Yao, H. Liu, and L. Guibas, "Compressive Network Analysis," *Journal of Machine Learning Research (preprint)*, 2011.

[171] D. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, pp. 1289–1306, 2006.

[172] M. Asif and J. Romberg, "On the LASSO and Dantzig selector equivalence," in *IEEE* 44*th Annual Conference on Information Sciences and Systems*, 2010, pp. 1–6.

[173] S. Salvador and P. Chan, "Determining the number of clusters/segments in hierarchical clustering/segmentation algorithms," in 16*th IEEE International Conference on Tools with Artificial Intelligence*, 2004, pp. 576–584.

[174] B. Turlach, W. N. Venables, and S. J. Wright, "Simulataneous Variables Selection," *Technometrics*, vol. 47, no. 3, pp. 349–363, 2005.

[175] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan 2011. [Online]. Available: http://dx.doi.org/10.1561/2200000016

[176] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, no. 1, pp. 49–67, 2006. [Online]. Available: http://dx.doi.org/10.1111/j.1467-9868.2005.00532.x

[177] D. Preuveneers et al., "Design for failure: Intelligent systems learning from their mistakes," in *Workshop Proceedings of the 10th International Conference on Intelligent Environments, Workshop on the Reliability of Intelligent Environments (WoRIE)*, 2014 (To be published).

[178] L. Neumeyer, B. Robbins, A. Nair, and A. Kesari, "S4: Distributed stream computing platform," in *Data Mining Workshops (ICDMW), 2010 IEEE International Conference on*, Dec 2010, pp. 170–177.

[179] M. Proctor, N. Michael, L. Peter, and F. Michael, "Drools documentation," JBoss. org, Tech. Rep., 2008.

[180] "A storm is coming: more details and plans for release," September 2011. [Online]. Available: http://web.archive.org/web/20110810053658/http://engineering.twitter.com/2011/08/storm-is-coming-more-details-and-plans.html

[181] M. De Francisci and Gianmarco, "Samoa: A platform for mining big data streams," in *Proceedings of the 22nd international conference on World Wide Web companion*, May 2013, pp. 777–778.

[182] "S4 project incubation status," March 2014. [Online]. Available: http://incubator.apache.org/projects/s4.html

[183] S. Sharwood, "Apache Foundation embraces real time big data cruncher 'Storm'," September 2013.

[184] VSE, "Wege in die neue stromzukunft," 2012. [Online]. Available: http://www.strom.ch/uploads/media/VSE_Wege-Stromzukunft_Gesamtbericht_2012.pdf

[185] K. C. Armel, A. Gupta, G. Shrimali, and A. Albert, "Is disaggregation the holy grail of energy efficiency? the case of electricity," *Energy Policy*, vol. 52, no. 0, pp. 213–234, 2013.

[186] G. Hart, "Nonintrusive appliance load monitoring," *Proceedings of the IEEE*, vol. 80, no. 12, pp. 1870–1891, 1992.

[187] M. Zeifman, "Disaggregation of home energy display data using probabilistic approach," *Consumer Electronics, IEEE Transactions on*, vol. 58, no. 1, pp. 23–31, 2012.

[188] J. Liang, S. Ng, G. Kendall, and J. Cheng, "Load signature study v part i: Basic concept, structure and methodology," in *Power and Energy Society General Meeting, 2010 IEEE*, July 2010.

[189] ——, "Load signature study v part ii: Disaggregation framework, simulation and applications," in *Power and Energy Society General Meeting, 2010 IEEE*, July 2010.

[190] A. Zoha, A. Gluhak, M. A. Imran, and S. Rajasegarar, "Non-intrusive load monitoring approaches for disaggregated energy sensing: A survey," *Sensors*, vol. 12, no. 12, pp. 16 838–16 866, 2012.

[191] A. Reinhardt, D. Burkhardt, M. Zaheer, and R. Steinmetz, "Electric appliance classification based on distributed high resolution current sensing," in *Local Computer Networks Workshops (LCN Workshops), 2012 IEEE 37th Conference on*, Oct 2012, pp. 999–1005.

[192] "Plotwatt, 04 2014." [Online]. Available: https://plotwatt.com/

[193] "Bidgely, 04 2014." [Online]. Available: http://www.bidgely.com/

[194] M. E. Berges, E. Goldman, H. S. Matthews, and L. Soibelman, "Enhancing electricity audits in residential buildings with nonintrusive load monitoring," *Journal of industrial ecology*, vol. 14, no. 5, pp. 844–858, 2010.

[195] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explor. Newsl.*, vol. 11, no. 1, pp. 10–18, Nov. 2009. [Online]. Available: http://doi.acm.org/10.1145/1656274.1656278

[196] J. LaMarche, K. Cheney, S. Christian, and K. Roth, “Home energy management products & trends,” Fraunhofer Center for Sustainable Energy Systems. Cambridge, Massachusetts, Tech. Rep., 2011.

[197] K. Anderson, M. Berges, A. Ocneanu, D. Benitez, and J. Moura, “Event detection for non intrusive load monitoring,” in *IECON 2012 - 38th Annual Conference on IEEE Industrial Electronics Society*, Oct 2012, pp. 3312–3317.

[198] MathWorks, “Matlab 2013b,” 2013.

[199] J. Z. Kolter and M. J. Johnson, “Redd: A public data set for energy disaggregation research,” in *proceedings of the SustKDD workshop on Data Mining Applications in Sustainability*, 2011.

[200] D. Preuveneers, Y. Berbers *et al.*, “Samurai: A streaming multi-tenant context-management architecture for intelligent and scalable internet of things applications,” in *Intelligent Environments (IE), 2014 10th International Conference on (Accepted for publication)*. IEEE, 2014.

[201] D. Preuveneers, A. D. Landmark, and L. W. M. Wienhofen, “Probabilistic event processing for situational awareness,” in *Lecture Notes in Informatics (LNI) - 12th International Conference on Innovative Internet Community Systems (I2CS 2012)*, vol. P-204. Köllen Druck+Verlag GmbH, June 2012, pp. 96–107. [Online]. Available: https://lirias.kuleuven.be/handle/123456789/350718

[202] A. Ramakrishnan, Z. Bhatti, D. Preuveneers, Y. Berbers, A. Andrushevich, R. Kistler, and A. Klapproth, “Behavior modeling and recognition methods to facilitate transitions between application-specific personalized assistance systems,” in *Ambient Intelligence*. Springer, 2012, pp. 385–390.

[203] M. Van Assche, A. Ramakrishnan, D. Preuveneers, and Y. Berbers, “Towards a transfer learning-based approach for monitoring fitness levels,” in *Evolving Ambient Intelligence*. Springer International Publishing, 2013, pp. 33–43.

[204] M. S. Kang, J. W. Chong, H. Hyun, S. M. Kim, B. H. Jung, and D. K. Sung, “Adaptive interference-aware multi-channel clustering algorithm in a zigbee network in the presence of wlan interference,” in *Wireless Pervasive Computing, 2007. ISWPC ’07. 2nd International Symposium on*, Feb 2007.

[205] “Specification of the bluetooth system v2.1 + edr,” july 2007.

[206] J. Suhonen, M. Kuorilehto, M. Hannikainen, and T. Hamalainen, “Cost-aware dynamic routing protocol for wireless sensor networks - design and prototype experiments,” in *IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sept 2006.

[207] C. E. Perkins and E. M. Belding-Royer, “Ad-hoc on-demand distance vector routing,” in *WMCSA*. IEEE Computer Society, 1999, pp. 90–100.

[208] R. E. Kalman, “A new approach to linear filtering and prediction problems,” *J. Basic Eng.*, vol. 82, pp. 35–45, 1960.

[209] G. Welch and G. Bishop, “An introduction to the kalman filter.”

[210] P. Bellavista, A. Corradi, and C. Giannelli, “Evaluating filtering strategies for decentralized handover prediction in the wireless internet,” in *Computers and Communications, 2006. ISCC ’06. Proceedings. 11th IEEE Symposium on*, June 2006, pp. 167–174.

[211] L. Xia, J. Ling-ge, H. Chen, and L. Hong-wei, “An intelligent vertical handoff algorithm in heterogeneous wireless networks,” in *Neural Networks and Signal Processing, 2008 International Conference on*, June 2008, pp. 550–555.

[212] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, “A performance comparison of multi-hop wireless ad hoc network routing protocols,” in *The fourth annual ACM/IEEE international conference on Mobile computing and networking*, 1998, pp. 85–97.

[213] O. Garcia-Morchon, et.al, “Security Considerations in the IP-based Internet of Things,” http://tools.ietf.org/html/draft-garcia-core-security-06, [Online; accessed 7-September-2014].

[214] e. Park S, "IPv6 over Low Power WPAN Security Analysis," http://tools.ietf.org/html/draft-daniel-6lowpan-security-analysis-05, 2011, [Online; accessed 7-September-2014].

[215] e. L. Seitz, "Design Considerations for Security Protocols in Constrained Environments," draft-seitz-ace-design-considerations-00, 2014, [Online; accessed 7-September-2014].

[216] S. Raza, L. Wallgren, and T. Voigt, "Svelte: Real-time intrusion detection in the internet of things," *Ad Hoc Networks*, vol. 11, no. 8, pp. 2661 – 2674, 2013. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1570870513001005

[217] P. Kasinathan, G. Costamagna, H. Khaleel, C. Pastrone, and M. A. Spirito, "Demo: An ids framework for internet of things empowered by 6lowpan," in *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*. ACM, 2013, pp. 1337–1340.

[218] I. Maravic, M. Vetterli, and K. Ramchandran, "Channel Estimation and Synchronization with Sub-Nyquist Sampling and Application to Ultra-Wideband Systems," in *Proc. IEEE ISCAS'04*, vol. 5, Vancouver, Canada, May 2004.

[219] J. Paredes, G. Arce, and Z. Wang, "Ultra-Wideband Compressed Sensing: Channel Estimation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 1, no. 3, pp. 383–395, Oct. 2007.

[220] I. Tunaru, B. Denis, and B. Uguen, "Reciprocity-Diversity Trade-off in Quantization for Symmetric Key Generation," in *Proc. PIMRC'14*, Washington DC, US, Sept. 2014.

[221] D. E. ingeniuertechnik gmbh, "Dresden elektronik and ZIGPOS announce real-time locating system kit with new ATMEL ranging technology," March, 2013. [Online]. Available: www.prlog.org/12092596

[222] A. V. Medina, J. A. Gómez, J. A. Ribeiro, and E. Dorronzoro, "Indoor position system based on a zigbee network," *Communications in Computer and Information Science*, vol. 362, pp. 6–16, 2013.

[223] H. V. Poor, *An introduction to signal detection and estimation (2nd ed.)*. New York, NY, USA: Springer-Verlag New York, Inc., 1994.

[224] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. Wiley, 2004, no. v. 1.

[225] S. Kay, *Fundamtls Statstcl Processg V1& Signal V2 Pk*. Prentice Hall, 2001.

[226] J. Saloranta, S. Severi, D. Macagnano, and G. Abreu, "Algebraic confidence for sensor localization," in *Proc. Asilomar Conf. Signals, Systems, and Computers*, 2012.

[227] B. Guo, F. Calabrese, E. Miluzzo, and M. Musolesi, "Mobile crowd sensing: Part 1 [guest editorial]," *Communications Magazine, IEEE*, vol. 52, no. 8, pp. 20–21, Aug 2014.

[228] H. Ma, D. Zhao, and P. Yuan, "Opportunities in mobile crowd sensing," *IEEE Communications Magazine*, vol. 52, no. 8, pp. 29–35, Aug 2014.

[229] IoT-A Consortium, "D1.4-converged architectural reference model for the iot v2.0," November 2012. [Online]. Available: http://www.iot-a.eu/public/public-documents/documents-1/1/1/D1.4/at_download/file

[230] FI-WARE, "Complex Event Processing (CEP) - IBM Proactive Technology Online," March 2014. [Online]. Available: http://catalogue.fi-ware.org/enablers/complex-event-processing-cep-ibm-proactive-technology-online

[231] I. Ishaq, D. Carels, G. K. Teklemariam, J. Hoebeke, F. V. D. Abeele, E. D. Poorter, I. Moerman, and P. Demeester, "IETF standardization in the field of the internet of things (IoT): a survey," *Journal of Sensor and Actuator Networks*, vol. 2, no. 2, pp. 235–287, 2013. [Online]. Available: http://www.mdpi.com/2224-2708/2/2/235/htm

[232] J. Holler, V. Tsiatsis, C. Mulligan, S. Avesand, S. Karnouskos, and D. Boyle, *From Machine-to-machine to the Internet of Things: Introduction to a New Age of Intelligence*. Academic Press, 2014.

[233] FI-WARE, “Bigdata analysis - cosmos.” [Online]. Available: http://catalogue.fi-ware.org/enablers/bigdata-analysis-cosmos

[234] Goldman Sachs, “Macroeconomic insights: What is the internet of things?” September 2014. [Online]. Available: http://www.goldmansachs.com/our-thinking/outlook/iot-infographic.html