FRONT PAGE

PROJECT FINAL REPORT

Grant Agreement number: 619633

Project acronym: ONTIC

Project title: Online Network Traffic Characterization

Funding Scheme: Collaborative Project

Period covered: from 01/02/2014 to 31/01/2017

Name of the scientific representative of the project's co-ordinator[1], Title and Organisation:

Dr. Alberto Mozo Velasco, Associate Professor, Universidad Politécnica de Madrid

Tel: (+34) 913 369 821

Fax: (+34) 913 365 485

E-mail: a.mozo@upm.es

Project website address:

http://ict-ontic.eu

[1] Usually the contact person of the coordinator as specified in Art. 8.1. of the Grant Agreement.

4.1 Final publishable summary report

Executive Summary

ONTIC (Online Network Traffic Characterization) is a small or medium-scale research project (STREP) that address the **Objective ICT-2013.4.2 Scalable data analytics of the EU's FP7 ICT Workprogram** by focusing on the development of tools and skills to deploy and manage robust and high performance data analytics processes over extremely large amounts of data.

Proactive and dynamic QoS management, network intrusion detection and early detection of network congestion problems among other applications in the context of network management and engineering, can benefit from the existence of an accurate and scalable mechanism for online characterization of evolving network traffic patterns. Nowadays, in a context of continuous exponential growth in network speed and number of devices interacting with each other through the network, current approaches for online network traffic characterization clearly lack scalability and accuracy. Therefore, a new generation of scalable mechanisms and techniques to characterize online network traffic are required. The ONTIC project investigated, implemented and tested a novel architecture of scalable online and offline mechanisms and techniques to characterize network traffic data streams. The goals of these mechanisms and techniques are twofold. The first goal is to identify network traffic patterns evolution and to proactively detect anomalies, when hundreds of thousands of packets are processed per second. The second goal is to characterize network traffic, applying a big data analytics approach and using distributed computation paradigms in the cloud on extremely large network traffic summary datasets consisting of trillions of records.

The ONTIC consortium is made of 7 partners: 3 from research/academics, Universidad Politecnica de Madrid (UPM), Politecnico di Torino (POLITO) and Centre Nationale de la Recherche Scientifique (CNRS), 3 large industry (ERICSSON Spain, SATEC, EMC Spain) and 1 SME (ADAPTIT) from 5 European countries.

ONTIC has integrated offline and online mechanisms and techniques into an autonomous supervised or unsupervised network traffic characterization framework to be used as cornerstone of a new generation of scalable and proactive network management and engineering applications. The final validation of the developed algorithms has been deployed using the Google Cloud platform.

ONTIC has proposed three paradigmatic use cases and implemented their corresponding prototypes. These prototypes have been validated with real data obtained from the core network of SATEC (a partner of the ONTIC consortium). On average, 1,5Gb of network traffic data was obtained each second and the capture process has been going on for a period of 24 months. The resulting dataset (0.5 Petabytes) has been anonymized and made publicly available to foster new research initiatives in the big data analytics area. In addition, a synthetic dataset called SynthONTS was generated and made publicly available in order to validate unsupervised network anomaly detectors. ONTIC has produced 30 scientific papers submitted to a wide range of high quality journals and international peer-reviewed conferences. Two of them were published in ISI/JCR Q2 journals and another one in the International Conference in Data Mining (ICDM), the premier data-mining conference in the world. Additionally, three innovations have been identified and three international workshops and education and training activities have been organized with the aim to foster the creation of skills in the data science and Big Data areas. All the partners have contributed to project dissemination attending and actively participating in different national and international workshops and

conferences. In order to provide an efficient project showcase, ONTIC released and setup a YouTube channel (https://www.youtube.com/channel/UCfDTUrnJ9lXBu-EHiExXb_A).

Project context and objectives

ONTIC Context

Accurate identification and categorization of network traffic according to application type is an important element of many network management tasks related with QoS such as flow prioritization, traffic shaping/policing, and diagnostic monitoring. For example, a network operator may want to identify and throttle (or block) in traffic from peer-to-peer (P2P) file sharing applications to manage its bandwidth budget and to ensure good performance of business critical applications. Similarly to network management tasks, many network engineering problems such as workload characterization and modeling, capacity planning and route provisioning also benefit from accurate identification of network traffic. Moreover, the detection of network attacks is an important task for network operators in today's Internet. Botnets, Distributed Denial of Service attacks (DDoS) and network-scanning activities are examples of the different threats that compromise the integrity and normal operation of the network every day.

Adaptive/Dynamic QoS implementations (based on online detection of evolving traffic patterns), proactive congestion control mechanisms (based on early problem detection), and Network Intrusion Detection Systems (based on anomaly detection), among others, are applications in the context of network management and engineering that will benefit from the existence of an accurate and scalable mechanism for online identification and characterization of network traffic patterns.

Due to the continuous growth in network speed, terabytes of data may cross the core network of a typical ISP every day. Moreover, in a medium term, the exponential growth of the M2M/IoT scenario is expected to generate more than 50 billion connected devices, each of them producing huge amounts of network traffic. Thus, two major issues hamper network data analysis in the short and medium term: (a) a huge amount of data coming from a huge number of sources can be collected in a very short time, and (b) it is hard to identify correlations and detect anomalies in real-time on network traffic traces that large.

In this context of continuous growth in network speed and number of interconnected devices, current approaches for online network traffic characterization clearly lack scalability and accuracy. Therefore, several **challenges** appear when considering the development of successful solutions.

Firstly, an offline mechanism for traffic characterization should be developed to provide initial network traffic identification. Considering the amounts of summary information that can be generated and stored at a typical ISP per day, and that more information available usually helps achieve better results; two significant needs appear: **i)** a big data solution is required to store and access summarized information of an ISP during a significant period of time (e.g. a one-year summary requires a petabyte-sized database); and **ii)** a true scalable traffic characterization mechanism is needed to efficiently process this huge amount of data.

Secondly, a scalable online mechanism for traffic characterization should be provided to both identify traffic patterns evolution and to detect anomalies in real time, minimizing the false positives rate. Given the high amount of data per second a medium size ISP faces, a tradeoff between the

expected accuracy, the computational complexity of the characterization mechanism, how and when to parallelize concrete subsystems, and the amount of data considered at each execution (i.e. temporal window size) have to be paid attention to. In this real time scenario, big data can provide a maximization of the overall system accuracy when identifying traffic patterns and their evolution over time.

ONTIC objectives

(Scientific Objective 1) Investigate, implement and test a novel architecture of scalable mechanisms and techniques to be able to (a) characterize online network traffic data streams, identifying traffic pattern evolution, and (b) proactively detect anomalies in real time when hundreds of thousands of packets are processed per second.

(Scientific Objective 2) Investigate, implement and test a completely new set of scalable offline data mining mechanisms and techniques to characterize network traffic, applying a big data analytics approach and using distributed computation paradigms in the cloud on extremely large network traffic summary datasets consisting of trillions of records.

(Technical Objective 3) Integrate offline and online mechanisms and techniques into an autonomous supervised or unsupervised network traffic characterization system to be used as cornerstone of a new generation of scalable and proactive network management and engineering applications.

(Technical Objective 4) Lead the dissemination and adoption of the ONTIC outcomes to other application domains where scalability, accuracy and in some cases real time response are a must. To this purpose, ONTIC will generate an open source, highly scalable offline/online analytics framework to be used by developers in other application domains.

Scientific and Technological results/foregrounds

Scientific results

We present in this section a summary of the scientific results obtained in the technical work packages WP2, WP3 and WP4.

WP2. Big Data Network Traffic Summary Dataset

Feature engineering is often a crucial step for the successful application of machine-learning and data mining techniques, especially when Big Data comes into play. Choosing a smaller subset of relevant features (attributes) helps to avoid the curse of dimensionality and over-fitting, reduce training times and improve model interpretability. In the context of this topic we have developed several feature selection algorithms:

- PPCSS: a parallelized version of the CSSP problem that focuses mainly on finding the minimizing column subset.
- PICS (Parallelized Independent Column Selection): an algorithm that introduces a random sampling strategy that is not dependent on the target rank and finds efficiently the best candidate subset.
- PPICS: a parallelized version of the Expectation Maximization algorithm to solve PPCA (probabilistic PCA)
- PRANKS (Parallel Ranking and Selection): an efficient, deterministic algorithm for providing feature ranking in Big Data-sets

WP3. Scalable Offline Network Traffic Characterization System

During the first and second year, WP3 released a set of innovative and scalable algorithms to address ONTIC use cases. For each algorithm we designed, the following information was provided: (1) actual software, by means of a pointer to the specific version in the code repository (https://gitlab.com/groups/ontic-wp3), and (2) the software documentation containing a short description of the algorithm, a how-to, some examples and the published papers related to the specific algorithms. The algorithms generated in WP3 are classified in three categories: unsupervised learning, supervised classification and correlation analysis.

Unsupervised learning

Contributions to this set of algorithms aimed at improving clustering techniques, in terms of scalability (new distributed approaches), flexibility (automatic parameter setting), and result evaluation (scalable clustering-quality index computation).

- FreeScan: scalable automatic parameter-free DBSCAN algorithm
- DiSiLike: a scalable distributed Silhouette-like computation for clustering result evaluation
- CLUSPLUS: a novel parallel subspace clustering algorithm on Apache Spark
- PSCEG: An unbiased Parallel Subspace Clustering algorithm using Exact Grids

Supervised classification

Contributions to this set of algorithms aimed at improving classification techniques, mainly addressing the scalability issue, but keeping the models human-readable to provide meaningful insights into the analyzed phenomena.

- BAC: [POLITO] scalable Bagged Associative Classifier

Correlation analysis

Contributions to this set of algorithms (developed in task T3.3) aimed at improving itemset mining and association rule extraction techniques, in terms of both technological scalability (new distributed approaches), and high-dimensional dataset scalability, besides an innovative flexible weight-based approach.

- PaMPa-HD: a Parallel MapReduce-based frequent Pattern miner for High-Dimensional data
- PaWI: Parallel Weighted Itemset Mining by Means of MapReduce

Furthermore, WP3 results went beyond the single task-separated contributions, by reaching additional targets, such as frameworks combining different techniques to better address the use cases. To this aim, WP3 led to algorithmic contributions that support or extend the specific targets of each task. A key contribution is represented by SaFe-NeC, a framework that shows how different WP3 algorithms can be exploited together to address the ONTIC challenge.

The aim of WP3 for the third year of the project was an experimental analysis of the algorithms designed to address the network traffic characterization domain. In this context we released SeLINA, the ONTIC framework for network traffic characterization, showing that the integration of different WP3 techniques and algorithms can properly address the ONTIC challenge. SeLINA aimed at specifically providing an all-in-one network traffic characterization tool. Overall, experimental results proved that our approaches addressed the ONTIC challenge, as SeLINA provided meaningful characterizations of network traffic. It has been able to detect known issues and changes in real-world network traffic traces, without prior knowledge, and with a completely automated approach. Human intervention on parameter settings is very limited. Furthermore, it proved to be scalable to Terabyte-size datasets (2-10 Terabytes), with billions of records, on hundreds of parallel executors, and exploiting state-of-the-art distributed platforms (Apache Spark).

WP4. Scalable Online Network Traffic Characterization System

During the first and second years, WP4 released a set of algorithms to characterize online large network traffic in terms of traffic pattern evolution and unsupervised network anomaly detection. These algorithms were designed to be run on the Big Data analytics system and the provisioning subsystem specified in the WP2. Experimental results of these algorithms were presented and discussed. These algorithms leverage large network traces and data mining techniques for helping the network administrator in its task. They offer them an insight on its network that could not have been possible with small network traces.

The key component produced in WP4 is an Unsupervised Network Anomaly Detection Algorithm (UNADA) capable of detecting network anomalies without relying on signatures training or labeled instances. UNADA is able to identify, in near real-time and on large network traffic, anomalies in an

unsupervised way, i.e., without previous knowledge on the anomalies. In addition, we released a parallel version of UNADA, called PUNADA. The obtained evaluation results showed that this solution improved UNADA execution time; however it displayed a limit in scalability. To overcome this issue, an Online and new Real-time Unsupervised Network Anomaly Detection Algorithm (ORUNADA) was then proposed. ORUNADA is an incremental version of UNADA relying on an incremental grid clustering algorithm and a sliding-time window. This solution speeds the execution time by a factor of 300 and allows online detection on large network traffic while preserving insuring a good quality is terms of anomaly identification. In the third year, we carried out the experimental design to test the platform and the algorithms, and provided detailed information about their configuration and parameterization. In addition, we generated a ground truth called SynthONTS for Synthetic Network Traffic Characterization of the ONTS Dataset. This ground truth was used to validate an unsupervised network anomaly detector. This unsupervised network anomaly detector is an improved version of ORUNADA and was tested using the Google cloud platform and more specifically the Google Dataproc and the Google Storage.

Complementary, this work package dealt with the problem of leveraging traffic patterns in order to make reliable forecasts. Driven by the goals of WP5 UC #2, which aimed to build a distributed system for bandwidth assignment to proactively control network congestion, an open source Network Traffic Forecasting Framework (NTFF) was proposed and developed. We designed a forecasting procedure that is shown to reliably predict the number of open sessions crossing a network link, which can be particularly useful for computing approximate bandwidth assignments effectively and avoid congestion. An exploratory analysis of the ONTS dataset was also carried out, drawing useful insights on the behavior of the traffic it represents. We conducted experiments on a sizeable sample of the ONTS dataset to validate the effectiveness of the forecasting procedure integrated in the NTFF, showing its ability to make 4-step forecasts within 2% of mean error. To the best of our knowledge, no existing network-oriented forecasting procedure has been trained and tested on a dataset of the scale of ONTS. During the third we addressed three different scenarios related to forecasting techniques and detection of anomalies. We released an enhanced version of the NTFF for network traffic behavior forecasting, as well as the obtained results when applied to the ONTS dataset. In particular, we showed the application of deep convolutional neural networks in order to exploit the temporal nature of forecasting scenarios. Additionally we developed SLBN++ a proactive congestion control protocol equipped with forecasting capabilities that outperforms current proposals. Finally, we released a novel machine-learning component for detecting anomalous behavior in cloud infrastructure based on deep neural networks.

Scientific publications

In the context of these technical work packages, ONTIC has produced 30 scientific papers submitted to a wide range of high quality journals and international peer-reviewed conferences. Below, we present the complete list of published papers classified by year.

Year 1

- A Survey on feature selection and reduction for network traffic characterization, B. Zhu and A. Mozo (First International Workshop in Big Data Applications and Principles, BIGDAP 2014, September, 2014, Madrid, Spain)

- A Telecom Analytics Framework for Dynamic Quality of Service Management, Guadamillas, M.A. López, N. Maravitsas, A. Mozo and F. Pulvirenti, (First International Workshop in Big Data Applications and Principles, BIGDAP 2014, September, 2014, Madrid, Spain)
- A.Bascuñana, F.Castro, D.Espadas, P.Sánchez, M.A Monjas. (2014) "Adaptive Quality of Experience (AQoE) control for Telecom Networks". Proceedings First international Workshop, BigDAP 14.Madrid, Spain, September 11-12 2014.
- P. Owezarski, "Unsupervised Classification and Characterization of Honeypot Attacks", International Conference on Network and System Management (CNSM'2014), Rio de Janeiro, Brazil, November 17-21, 2014
- P. Owezarski, P. Casas, J. Mazel, "Unsupervised detection of network attacks in the dark", 1st international workshop on big data applications and principles (BigDAP'2014), Madrid, Spain, September 11-12, 2014
- Misleading generalized itemset mining in the cloud. Elena Baralis, Luca Cagliero, Tania Cerquitelli, Silvia Chiusano, Paolo Garza, Luigi Grimaudo, Fabio Pulvirenti. IEEE ISPA 2014, Milan, Italy, 26-27-28 August 2014
- MGI-Cloud: discovery of misleading generalized itemsets. Elena Baralis, Luca Cagliero, Tania Cerquitelli, Silvia Chiusano, Paolo Garza, Luigi Grimaudo, Fabio Pulvirenti. SEBD 2014, Castellammare di Stabia, Italy, 16-17-18 June 2014
- Network traffic analysis by means of Misleading Generalized itemsets. Daniele Apiletti, Elena Baralis, Luca Cagliero, Tania Cerquitelli, Silvia Chiusano, Paolo Garza, Luigi Grimaudo, Fabio Pulvirenti, BigDap 2014, Madrid, Spain

Year 2

- Elena Baralis, Luca Cagliero, Paolo Garza, Luigi Grimaudo "PaWI: Parallel Weighted Itemset Mining by Means of MapReduce". IEEE International Congress on Big Data (BigData Congress), New York, NY, June 27-July 2, 2015, pp: 25-32, ISBN: 978-1-4673-7277-0,
- Daniele Apiletti, Elena Baralis, Tania Cerquitelli, Paolo Garza, Fabio Pulvirenti, and Pietro Michiardi, "PaMPa-HD: a Parallel MapReduce-based frequent Pattern miner for High-Dimensional data". IEEE International Conference on Data Mining (IEEE ICDM), 3rd International Workshop on High Dimensional Data Mining (HDM'15), November 14-17, Atlantic City (New Jersey, USA).
- Daniele Apiletti, Paolo Garza, Fabio Pulvirenti, "A Review of Scalable Approaches for Frequent Itemset Mining". 19th East-European Conference on Advances in Databases and Information Systems (ADBIS), 2nd International Workshop on Big Data Applications and Principles (BigDap 2015), September 8-11, 2015, Futuroscope, Poitiers, pp: 243-247.
- Daniele Apiletti, Elena Baralis, Tania Cerquitelli, Paolo Garza, Luca Venturini, "SaFe-NeC: a Scalable and Flexible system for Network data Characterization". IEEE/IFIP Network Operations and Management Symposium (IEEE NOMS), Istanbul, Turkey, 25-29 April 2016.
- B. Ordozgoiti, S. Gomez, A. Mozo, "Parallelized Unsupervised Feature Selection for Large-Scale Network Traffic Analysis". 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning ESSAN16, Bruges (Belgium), 27 - 29 April 2016. ESSAN is rated as Core B.
- B. Zhu, B. Ordozgoiti, A. Mozo, "PSCEG: An unbiased Parallel Subspace Clustering algorithm using Exact Grids". 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning ESANN16, Bruges (Belgium), 27 - 29 April 2016. ESANN is rated as Core B.
- B. Ordozgoiti, S. Gomez, A. Mozo, "Massively Parallel Unsupervised Feature Selection on Spark". 19th East-European Conference on Advances in Databases and Information Systems (ADBIS), 2nd International Workshop on Big Data Applications and Principles (BigDap 2015), September 8-11, 2015, Futuroscope, Poitiers. This work was selected as the best paper in BigDAP15.

- B. Zhu, A. Mara, A. Mozo , “CLUS: Parallel subspace clustering algorithm on Spark”. 19th East-European Conference on Advances in Databases and Information Systems (ADBIS), 2nd International Workshop on Big Data Applications and Principles (BigDap 2015), September 8-11, 2015, Futuroscope, Poitiers.
- B. Ordozgoiti, S. Gomez, A. Mozo, “NPEPE: Massive Natural Computing Engine for Optimally Solving NP-complete Problems in Big Data Scenarios”. 19th East-European Conference on Advances in Databases and Information Systems (ADBIS), 2nd International Workshop on Big Data Applications and Principles (BigDap 2015), September 8-11, 2015, Futuroscope, Poitiers.
- Juliette Dromard, Gilles Roudière and Philippe Owezarski,“Unsupervised Network Anomaly Detection in Real-time on Big Data”. 19th East-European Conference on Advances in Databases and Information Systems (ADBIS), 2nd International Workshop on Big Data Applications and Principles (BigDap 2015), September 8-11, 2015, Futuroscope, Poitiers

Year 3

- Ordozgoiti, Bruno, Sandra Gómez Canaval, Alberto Mozo, "A Fast Iterative Algorithm for Improved Unsupervised Feature Selection". IEEE International Conference on Data Mining, 2016 (Core A*). This paper has been selected as one of the best ICDM-16 papers and we have been invited for possible publication in the journal Knowledge and Information Systems (Q2 indexed in ISI/JCR)
- Ordozgoiti, Bruno, Sandra Gómez Canaval, and Alberto Mozo, "Feature Ranking and Selection for Big Data Sets". 3rd International Workshop on Big Data Applications and Principles, BigDAP 2016, co-located with the 20th East-European Conference on Advances in Databases and Information Systems, ADBIS 2016. Springer International Publishing, 2016.
- Zhu, Bo, and Alberto Mozo, "Spark2Fires: A New Parallel Approximate Subspace Clustering Algorithm". 3rd International Workshop on Big Data Applications and Principles, BigDAP 2016, co-located with the 20th East-European Conference on Advances in Databases and Information Systems, ADBIS 2016. Springer International Publishing, 2016.
- Ordozgoiti, Bruno, Alberto Mozo, Sandra Gómez Canaval, Udi Margolin, Elisha Rosensweig, Itai Segall, "Deep convolutional neural networks for detecting noisy neighbours in cloud infrastructure". ESANN 2017 (Core B).
- B. Ordozgoiti, S. Gomez Canaval, A. Mozo, “Probabilistic Leverage Scores for Parallel Unsupervised Feature Selection on Massive High-Dimensional Datasets”. 14th International Work-Conference on Artificial Neural Networks, IWANN 2017 (Core B).
- Apiletti, Baralis, Cerquitelli, Garza, Giordano, Mellia, Venturini, “SeLINA: a Self-Learning Insightful Network Analyzer”. IEEE Transactions On Network And Service Management (ISI/JCR Q2), vol. 13 n. 3, pp. 696-710, 2016.
- Apiletti, Baralis, Cerquitelli, Garza, Pulvirenti, “Frequent Itemsets Mining for Big Data: a comparative analysis”. Elsevier - Big Data Research (January 2017).
- Apiletti, Baralis, Cerquitelli, Garza, Venturini, “SaFe-NeC: A scalable and flexible system for network data characterization”. 2016 IEEE/IFIP Network Operations and Management Symposium, NOMS 2016, Istanbul (Turkey), 2016. pp. 812-816
- Ivanovic, Thalheim, Catania, Schewe, Kirikova, Saloun, Dahanayake, Cerquitelli, Baralis, Michiardi, “New Trends in Databases and Information Systems”. 3rd International Workshop on Big Data Applications and Principles, BigDAP 2016, co-located with the 20th East-European Conference on Advances in Databases and Information Systems, ADBIS 2016, Prague, Czech Republic, August 28-31, 2016, Proceedings.
- Venturini, Garza, Apiletti, “BAC: A bagged associative classifier for big data frameworks”. 3rd International Workshop on Big Data Applications and Principles, BigDAP 2016, co-located with the

20th East-European Conference on Advances in Databases and Information Systems, ADBIS 2016, Prague, Czech Republic, August 28, 2016. pp. 137-146.

- J. Dromard, G. Roudière, P. Owezarski, "Online and Scalable Unsupervised Network Anomaly Detection Method". IEEE Transaction on Network and Service Management (ISI/JCR Q2), 2016.
- M. A. López Peña, C. Area Rúa, S. Segovia, A "Fast Data" Architecture: Dashboard for Anomalous Traffic Analysis in Data Networks. IEEE 11th International Conference on Digital Information Management (ICDIM 2016), September 19-21, 2016, Porto (Portugal).

Technical results and foreground

In the context of network management and engineering, ONTIC identified three key scenarios to address network issues of increased importance. During the project's first year, the initial Use Cases were refined and assigned a specific slogan: (UC #1) Network Anomaly Detection; (UC #2) Adaptive Quality of Experience (QoE) Control; and (UC #3) Proactive Congestion Detection and Control Systems. During the project's second and third year, the Use Case requirements have been further refined and the functionalities required have been implemented at a system level.

UC#1 Network Anomaly Detection

Use Case #1 aimed at providing a system able to perform online network traffic monitoring for detecting in real-time network anomalies. For enhancing its applicability in realistic ISP/CSP network environments, Use Case #1 defined a scenario in which automatic anomaly detection is embedded within an administrator-oriented network supervision tool. This way, network traffic and potential anomalies can be analysed and traced back if required through a set of integrated visual interfaces fitting the every-day practices of network administrators.

UC #1 encompasses two subsystems: (a) an autonomous system for detecting and characterizing traffic anomalies, making it possible to autonomously and efficiently manage them: the Anomaly Detection Subsystem, and (b) a dashboard for enabling network operators to get detailed information about network traffic features and statistics, near real-time, anomalies detected and traffic behaviour during the periods in which the anomalies are detected: the Network Traffic Dashboard Subsystem. Both subsystems have been specified, designed, implemented and evaluated.

The results showed valid and scalable performance. And equally important, that system design and implementation can sufficiently resolve the critical challenge of timely synchronization –within real-time response constraints- between independent applications and processes that have to work on Big Data and share produced results and alarms.

UC#2 Adaptive Quality of Experience (QoE) Control

UC #2 implements a ML analytics-enhanced control loop so that it is possible to react to video quality of experience (QoE) degradation situations and apply appropriate alleviation measures. UC #2 has been designed under the umbrella of the so called AQoE (Adaptive Quality of Experience) framework. AQoE comprises several phases including measurement, analytics, policy decision and enforcement, all of them running in the form of a closed control loop. The UC #2 implementation shows how these tasks can be performed in an automated manner in order to cope with timely response requirements. The system detects and corrects deviations on network's performance automatically and hence, it is capable of delivering the best video customer experience possible.

The main challenge the implementation of this use case has faced, has been the actual detection of video QoE degradation patterns. However, the ONTS dataset –the dataset with actual traffic traces captured by the project- does not provide enough information to efficiently derive features related to the video services and their quality; payload has been removed from the captured packets for privacy reasons. Therefore, an alternative approach has been taken: a VLC-based test bed has been set up and a labeled TCP flow dataset has been generated. Labels have been assigned by human users and, following a training phase, a model is generated. The model is generated by means of clustering and afterwards clusters are mapped to existing quality labels. This way, it is possible to compute the average value of the quality of video experience associated to the clustered TCP flows for a time window. When it goes below a given threshold for a given time, an alarm is issued and actuation is triggered.

The proof-of-concept relies on several open-source off-the-self products and frameworks that enable stream processing, scalability and reliability. To name a few: Apache Spark (Streaming), DC/OS, Apache Mesos, Marathon and Docker. All the functionality has been developed using Java and Python. Source code and Docker images are available in Gitlab and Docker Hub. The proof-of-concept runs on Google Cloud services.

UC#3 Proactive Congestion Detection and Control Systems

This use case aimed to meet two complementary requirements: First, to detect in advance congestion problems that could occur in the network links by means of advanced forecasting mechanisms, and secondly, to avoid these congestion problems predicting max-min fair sending rates and assigning them to network sessions.

The proposed solution is included in a congestion control architecture that integrates a forecasting module into a proactive EERC (End-to-End Rate Control) system model. The implementation of this architecture relays in three functional blocks that work in cooperation in order to fulfill its requirements: (1) a scalable and distributed max-min fair optimization algorithm that computes max-min fair rates, (2) a proactive congestion control protocol that deploys the max-min fair algorithm, communicates sending rates to sessions and enforces the fulfillment of these rates at the edges of the network; and (3) a forecasting machine learning module that provides future predictions of protocol state values.

As it was unfeasible to setup a realistic deployment of the proposed solution involving hundreds of routers and thousands of nodes connected to a network, we demonstrated this solution by means of simulations run on top of a homemade discrete event simulator. Our simulator can (a) run experiments consisting of thousands of routers and up to a million of hosts and sessions, (b) import Internet-like topologies generated with the Georgia Tech gt-itm tool, and (c) model key network parameters, like transmission and propagation times in the network links, processing time in routers and limited size packet in link queues.

Finally, we carried out a thorough study of the ONTS dataset and conducted a stochastic characterization in this dataset. The aim of this research activity was to deliver a module that could generate realistic network traffic patterns of sessions joining and leaving the network to be used as input to simulations.

The ONTIC framework

The ONTIC Framework is provided as a number of software components under an open source license. Usually, the component source code is provided. When referring to the components implementing the use cases' proofs-of-concept, container images, are usually provided as well. Git has been selected as control version software as it is currently one of the most used free control version software, with support in the most popular development environments (IDE). As such, a Git server has been used as the ONTIC Framework software repository. On the other hand, some components in the ONTIC Framework, mainly those related to use case implementation, as provided as software containers as well. In particular, all the algorithms and software components generated in the technical workpackages (WP2, WP3, WP4 and WP5) have been released with an open source licence (Apache 2.0) and stored in a gitlab repository (https://gitlab.com/ontic/). In addition, the docker containers utilized in the Use case UC#2 have been uploaded at Docker Hub.

Data assets

One of the cornerstones of the ONTIC project is the provisioning of network traffic traces from the operational environment of an ISP over a long period of time −almost throughout project duration. The so-called ONTIC Network Traffic Summary (ONTS) dataset was not only used by the project for testing the developed on/off-line ML-based traffic analysis algorithms but it has also been made available to the wider research community for providing a contemporary, valid traffic dataset for reference testing.
ONTS has gathered traffic traces captured from the core network of a multi-service ISP over 760 days since October 2014, amounting to a total of 511TB corresponding to an average of 688GB of captured data per day. It can become available through an explicit electronic request procedure in text files containing time-stamped anonymized IP packet headers or flow-aggregated records adhering to current legislation for data protection.

In addition, SynthONTS was generated in order to validate the unsupervised network anomaly detector developed by a member of the project (LAAS-CNRS). This ground truth was generated by injecting anomalies created on CORE (link :https://www.nrl.navy.mil/itd/ncs/products/core) in the ONTS dataset. SynthONTS is also made publicly available as a part of the ONTS dataset. This ground truth is available under request for use by the academic researcher. It is delivered with a description of the ground truth and the injected anomalies (IP of attackers, victims and amplifiers).

Potential impact, Dissemination activities and Exploitation of results

ONTIC has pushed for an innovative and challenging way to foster research exploitation from the project. The exploitation strategy has been focused on the following ONTIC outcomes:

- ONTS (ONTIC Network Traffic) dataset
- The ONTIC framework
- Prototype for Network Intrusion Detection Use Case
- Prototype for Dynamic QoS management Use Case
- Prototype for Proactive Congestion Detection and Control Systems Use Case

The exploitation of the project's results was centred on four core principles: (1) User focused activities and user engagement, (2) Joint efforts by all consortium members, (3) Continuous exploitation activities, from the start of the project on, and (4) Activities and effects lasting well past the end of the project.

In the scope of this project, several types of users have been the main stakeholders for exploitation activities: Network operators, Service providers, Business users and End users. It was vital to address them all and integrate their needs into the research activities. Doing so not only produced market-relevant technologies but also ensured that all user groups obtained useful results out of the project. Complementary, ONTIC proposed the adoption of its outcomes to other application domains through the generation of a development framework. In this context it was important to include a new group of stakeholders formed by potential customers of those other areas where ONTIC results could be used.

User integration was accomplished with several instruments:

- The external user group comprised of industrial and commercial users, enabling them to give feedback and influence the development of the project, incorporating their actual problems and showing them possible solutions. They were the first users (early adopters) of the new technologies and thus also acted as multipliers and "beacons" for others to follow.
- The thorough and overarching market and use case research undertaken in WP5 outlined key stakeholders, business users and end-users and, as a major element of this task, also sought their needs and requirements, to incorporate in all future activities.
- The dissemination activities in WP6 alerted a broader range of potential users to the new developments, leading to feedback and further input and ideas for the project.
- The market studies carried out in WP6 addressed experts as well as end users, depending on the specific analyzed technology. This feedback influenced the further exploitation focus of the technologies. Furthermore, target groups were defined and analyzed to identify possible unique attributes and selling points of the inventions.

By paying close attention to the specific user needs and through the thorough analysis of their feedback, market adoption can be significantly increased or actually enabled.

The set of dissemination activities developed in ONTIC include scientific papers, journals and conferences of interest, workshop organization, IPR, press releases, and a list of relevant industrial associations that are interested in the project activities and outcomes. Both future and current to-date activities are presented, targeting different scientific communities, students, stakeholders and decision makers.

ONTIC has produced 30 scientific papers submitted to a wide range of high quality journals and international peer-reviewed conferences. The detailed list of these publications can be found in the former section "Scientific publications".

The BigDap international workshop has been organized and sponsored by ONTIC in 2014, 2015 and 2016. The key topics of the Workshop included the application of scalable Big Data analytics to network traffic characterization, the main theme of ONTIC, as well as other application domains. In order to promote the dissemination of the work activities and results of the ONTIC project, the specific goals of BigDAP were to provide a networking space for people in academic institutions and in the industry and to enhance the participation of graduate students, especially those enrolled in PhD programs related to Big Data, Big Analytics and Network Science.

For enhancing visibility and attendance, BigDAP was co-located with the East-European Conference on Advances in Databases and Information Systems (ADBIS) in the last two years and the proceedings volumes were published by Springer (www.springer.com/gp/book/9783319232003, www.springer.com/gb/book/9783319440651).

In order to promote the ONTS dataset and to wide audiences from both research and industrial areas interested in data analysis, at the end of the project, we setup and organized a Datathon (http://ict-ontic.eu/index.php/events/datathon). This event provided a "big" free choice of publicly accessible network traffic datasets to data analysts who are currently suffering from a lack of such datasets. The ONTS dataset was used in the datathon and different State-of-the-Art and novel approaches were proposed and applied.

Several training events were organized in order to foster the creation of skills in the data science and Big Data areas and in particular in the application of these topics in telecom scenarios:

- Course "Massively parallel machine learning"
  EIT Digital Master School's Master's Programme in ICT Innovation: Data Science.
- Creation of the Network Analytics Lab (Ericsson and Telefonica).
  In this lab, mentored Grant-holder students were involved in Data Science projects
- PhD, Master courses and undergraduate seminars were organized at POLITO and UPM in order to teach machine learning and data mining topics related to ONTIC algorithms
- A course called "Bootnets and anomaly detection in Internet Traffic" was taught by CNRS team in the Master TLS-SEC (Toulouse).

In addition, all the partners have contributed to project dissemination attending and actively participating in different national and international workshops and conferences.

Finally, and in order to provide an efficient project showcase, ONTIC released and setup a YouTube channel (https://www.youtube.com/channel/UCfDTUrnJ9lXBu-EHiExXb_A) containing three videos corresponding to each use case demo plus an additional introductory video in which some other outcomes not covered by the demo videos are presented.

From the start of the project, exploitation has played an important part in the project. Even from the outset, during project formation and partner search, the many results and their proper implementation had always very high priority on the agenda, which showed in the top-tier industrial partners and the special, sector-specific technology transfer knowledge brought on board. To adequately continue this process, a series of measures were carried out to ensure that the different exploitation aspects - scientific and commercial - received adequate treatment. On the scientific side these include publications, speeches and talks in congresses and workshops, doctorate theses and course integration, as well as training materials, posters, stands and functional prototype applications and frameworks. On the commercial side, the major industrial partners were certain that they could directly use the developed technologies in next-generation products and processes. For universities and those SMEs in need of it, a proven process was enacted to transfer the technology to commercial application, starting with screening talks and covering IPR management, market analysis, business development and license negotiation. A dual-pronged approach was used to streamline the technological development as well as keep it in sync with the technology transfer process, so that shorter times to market could be reached and usual innovation transfer problems could be alleviated.

A special focus of the ONTIC model was on remedying the so-called "valley of death" – the phase between finished research activities and market adoption. Throughout the runtime of the project, many activities were conducted to ensure that the existing exploitation potentials were actually realized - from active marketing of the results to hands-on business development and proactive securing of additional funding, both public and private. Several key stakeholders and decision-makers (through external member group, targeted dissemination and networking) were contacted as these are vital to any further and lasting use of the technologies, whether in science or in industry. Of course, the effectiveness of all these activities should be continuously measured, so that the work can be steered towards the most beneficial actions or areas. Lacking areas were identified quickly and counter measures were introduced.

The following table gives an overview over the key performance indicators (KPIs) achieved at the end of the project.

Area	KPI	Guaranteed	Current Value Year #3	Explanation
Screening	Number of screening talks	(10) ✓	**31** 18 UPM 9 CNRS 4 POLITO	Personal talks with the involved researchers to find out about current activities and stimulate idea disclosure notices.
	Validated Use cases	(3) ✓	**7** 3 ONTIC Use Cases 2 SATEC 1 ADAPTIT 1 CNRS	Ideas for a practical application of the technology.

	Business ideas	(4) ✓	**18** 6 SATEC 5 CNRS 4 ERC 3 ADAPTIT	Specific idea for a business based on the new technology.
Valorization	Patents	(3)	**N/A**	Number of patents; a specifically low amount is often a sign of research focused purely on science. Not possible to have a granted patent within the time frame of the project.
	Patent applications	(3)	**1** 1 ERC	Number of patent applications.
	Prototypes	(3)	**8** 1 ERC/EMC/POLITO 2 ERC 2 SATEC 2 CNRS 1 ADAPTIT	Working prototypes; if no prototypes can be produced, technologies are too far from market.
	Co-operation / license agreements	(2) ✓	**6** 6 CNRS	Agreements between research institutions and partners outside the project.
	Initiated start-up proposals	(1)		Plans for the establishment of a new business.
Contacts	Outreach to new R&D partners	(10) ✓	**44** 19 UPM 17 CNRS 1 ERC 1 EMC 6 POLITO	Additional scientific contacts.
	Outreach to new industrial partners	(10) ✓	**52** 23 UPM 5 SATEC 1 ADAPTIT 10 CNRS 1 EMC 2 ERC	Additional industrial contacts.

			3 POLITO 3 DELLEMC	
	Outreach to different user groups	(4)	**13** 1 UPM 6 POLITO 2 ADAPTIT 4 CNRS	Contact to different user groups in the scope of market or case studies.

Regarding the exploitation activities carried out in ONTIC, one of the main outcomes obtained at the end of the project is the identification of 3 innovations related to the use cases UC#1 and UC#2.

After the project has ended, its work and its results will continue to influence policy, business, science and society, whether through additional scientific or validation projects based on the technologies, through new products, services or companies, through new or improved international standards and through better and previously impossible experiences for every citizen.

Major factors in sustaining the results will be, on the one hand, the availability of new business models that are made available through the implementation of the technologies, such as service providers for specialized streaming services, virtual network providers and operators, brokers and exchanges for networks and transmission time slots and specialized solution providers for network automation and self-organizing (configuration, optimization and healing) technologies.

On the other hand, reduced market entry barriers for a variety of offerings, are a point where the continued influence of the project will be felt. It is expected that these two factors provide a key stimulation of commercial activity throughout Europe and worldwide, with European companies as first movers – they are most exposed to the new technologies and thus have a head start. The aforementioned new and interesting markets and the possibility to operate within them successfully will be proven by newly generated spin-offs, which build solutions based on the new technologies, thus providing examples for others and stimulating new ventures.

Overall, ONTIC will enhance the competitiveness of European firms on many levels, from lowered CAPEX and OPEX requirements, new IT and communication solutions, to new markets for new services and products. Additional standards, built on the work of the project, are expected to go into effect. The project will also offer rich possibilities for follow-up projects in specialised practical development or in national validation-type programmes. The public partners also see the possibility of applying for additional basic research grants from their individual states. Additional research potential also arises out of the co-operation with industrial partners, for which the project can lay the groundwork.

Finally, and in order to facilitate easy access to the results, Professor Dr. Alberto Mozo at Universidad Politécnica de Madrid will be in charge of maintaining and keeping alive the ONTIC web site (http://ict-ontic.eu) for at least two years after the end of the project.

Project website and relevant contact details

Website

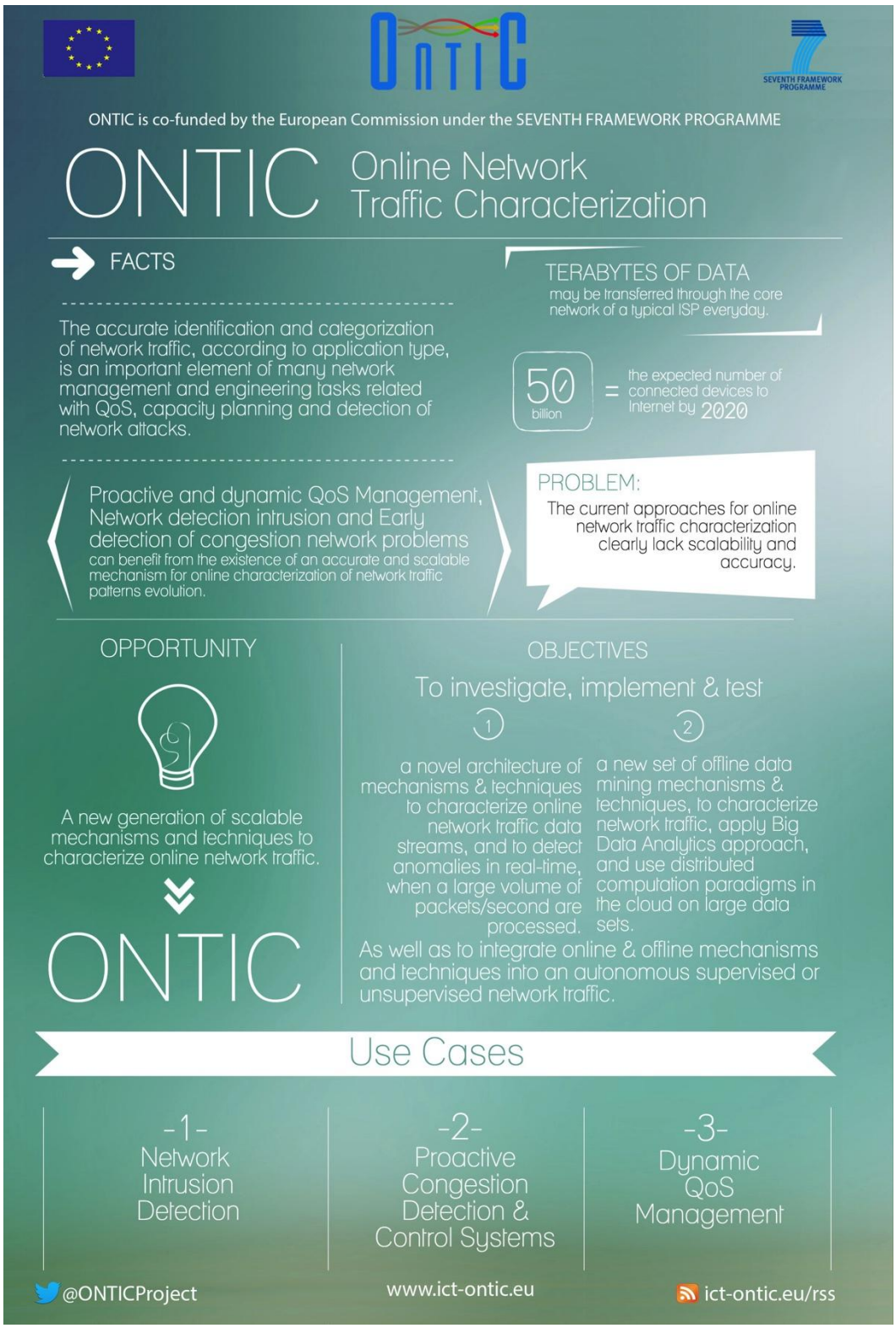
http://ict-ontic.eu

Contact

Dr. Alberto Mozo
Associate Professor
Technical University of Madrid
E-mail: **a.mozo@upm.es**

Logo

Fact-sheet


ONTIC is co-funded by the European Commission under the SEVENTH FRAMEWORK PROGRAMME
ONTIC Online Network Traffic Characterization
FACTS
The accurate identification and categorization of network traffic, according to application type, is an important element of many network management and engineering tasks related with QoS, capacity planning and detection of network attacks.
TERABYTES OF DATA
may be transferred through the core network of a typical ISP everyday.
50 billion = the expected number of connected devices to Internet by 2020
Proactive and dynamic QoS Management, Network detection intrusion and Early detection of congestion network problems
can benefit from the existence of an accurate and scalable mechanism for online characterization of network traffic patterns evolution.
PROBLEM:
The current approaches for online network traffic characterization clearly lack scalability and accuracy.
OPPORTUNITY
A new generation of scalable mechanisms and techniques to characterize online network traffic.
ONTIC
OBJECTIVES
To investigate, implement & test
1 a novel architecture of mechanisms & techniques to characterize online network traffic data streams, and to detect anomalies in real-time, when a large volume of packets/second are processed.
2 a new set of offline data mining mechanisms & techniques, to characterize network traffic, apply Big Data Analytics approach, and use distributed computation paradigms in the cloud on large data sets.
As well as to integrate online & offline mechanisms and techniques into an autonomous supervised or unsupervised network traffic.
Use Cases
-1- Network Intrusion Detection
-2- Proactive Congestion Detection & Control Systems
-3- Dynamic QoS Management
@ONTICProject
www.ict-ontic.eu
ict-ontic.eu/rss

4.2 Use and dissemination of foreground

The plan consists of:

- Section A

  This section describes the dissemination measures, including the scientific publications relating to foreground.

- Section B

  This section specifies the exploitable foreground and provides the plans for exploitation.

Section A (public)

TEMPLATE A1: LIST OF SCIENTIFIC (PEER REVIEWED) PUBLICATIONS, STARTING WITH THE MOST IMPORTANT ONES										
NO.	Title	Main author	Title of the periodical or the series	Number, date or frequency	Publisher	Place of publication	Year of publication	Relevant pages	Permanent identifiers[2] (if available)	Is/Will open access[3] provided to this publication?
1	*SeLINA: a Self-Learning Insightful Network Analyzer*	*Apiletti, Baralis, Cerquitelli, Garza, Pulvirenti,*	*IEEE Transactions On Network And Service Management, TNSM*		*.IEEE*		*2016*			*yes*
2	*Online and Scalable Unsupervised Network Anomaly Detection Method*	*Dromard, Owezarski*	*IEEE Transaction on Network and Service Management, TNSM*		*IEEE*		*2016*			*yes*
3	*A Fast Iterative Algorithm for Improved Unsupervised Feature*	*Ordozgoiti, Gomez Canaval, Mozo*	*IEEE International Conference on Data Mining, ICDM*		*IEEE*		*2016*			*yes*
4	*Deep convolutional neural networks for detecting noisy neighbours in cloud infrastructure*	*Ordozgoiti, Gomez Canaval, Mozo*	*European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN*		*I6DOC*		*2017*			*yes*
5	*Probabilistic Leverage Scores*	*Ordozgoiti,*	*International Work-*		*Springer*		*2017*			*yes*

[2] A permanent identifier should be a persistent link to the published version full text if open access or abstract if article is pay per view) or to the final manuscript accepted for publication (link to article in repository).

[3] Open Access is defined as free of charge access for anyone via Internet. Please answer "yes" if the open access to the publication is already established and also if the embargo period for open access is not yet over but you intend to establish open access afterwards.

	for Parallel Unsupervised Feature Selection on Massive High-Dimensional Datasets	*Gomez Canaval, Mozo*	*Conference on Artificial Neural Networks, IWANN,*							
6	*SaFe-NeC: A scalable and flexible system for network data characterization*	*Apiletti, Baralis, Cerquitelli, Garza, Pulvirenti,*	*IEEE/IFIP Network Operations and Management Symposium, NOMS*		*IEEE*		*2016*			*yes*
7	*PSCEG: An unbiased Parallel Subspace Clustering algorithm using Exact Grids*	*Zhu, Ordozgoiti, Mozo*	*European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN*		*I6DOC*		*2016*			*yes*
8	*Parallelized Unsupervised Feature Selection for Large-Scale Network Traffic Analysis*	*Ordozgoiti, Gomez Canaval, Mozo*	*European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning, ESANN*		*I6DOC*		*2016*			*yes*
9	*SaFe-NeC: a Scalable and Flexible system for Network data Characterization*	*Apiletti, Baralis, Cerquitelli, Garza, Pulvirenti,,*	*IEEE/IFIP Network Operations and Management Symposium, NOMS*		*IEEE*		*2016*			*yes*
10	*Misleading generalized itemset mining in the cloud*	*Baralis, Cagliero, Cerquitelli, Chiusano, Garza, Grimaudo, Pulvirenti.*	*IEEE International Symposium on Parallel and Distributed Processing with Applications, ISPA*		*IEEE*		*2014*			*yes*
11	*Unsupervised Classification and Characterization of Honeypot Attacks*	*Owezarski*	*International Conference on Network and System Management, CNSM*				*2014*			*yes*
12	*Frequent Itemsets Mining for Big Data: a comparative analysis*	*Apiletti, Baralis, Cerquitelli,*	*Big Data Research*		*Elsevier*		*2017*			*yes*

		Garza, Pulvirenti,,								
13	*PaMPa-HD: a Parallel MapReduce-based frequent Pattern miner for High-Dimensional data*	*Apiletti, Baralis, Cerquitelli, Garza, Pulvirenti, Michiardi,*	*IEEE International Conference on Data Mining (IEEE ICDM). International Workshop on High Dimensional Data Mining, HDM*		*IEEE*		*2015*			*yes*
14	*PaWI: Parallel Weighted Itemset Mining by Means of MapReduce*	*Baralis, Cagliero, Garza, Grimaudo*	*IEEE International Congress on Big Data, (BigData Congress)*		*IEEE*		*2015*			*yes*
15	*A "Fast Data" Architecture: Dashboard for Anomalous Traffic Analysis in Data Networks*	*López Peña, Area Rúa, Segovia,*	*IEEE International Conference on Digital Information Management, ICDIM*		*IEEE*		*2016*			*yes*
16	*Feature Ranking and Selection for Big Data Sets*	*Ordozgoiti, Gomez Canaval, Mozo*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*		*Springer*		*2016*			*yes*
17	*Spark2Fires: A New Parallel Approximate Subspace Clustering Algorithm*	*Zhu, Mozo*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*		*Springer*		*2016*			*yes*
18	*BAC: A bagged associative classifier for big data frameworks*	*Venturini, Garza, Apiletti*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*		*Springer*		*2016*			*yes*

19	*Unsupervised Network Anomaly Detection in Real-time on Big Data*	*Dromard, Owezarski*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*				*2015*			*yes*
20	*NPEPE: Massive Natural Computing Engine for Optimally Solving NP-complete Problems in Big Data Scenarios*	*Ordozgoiti, Gomez Canaval, Mozo*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*				*2015*			*yes*
21	*CLUS: Parallel subspace clustering algorithm on Spark*	*Zhu, Mara, Mozo*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*				*2015*			*yes*
22	*Massively Parallel Unsupervised Feature Selection on Spark*	*Ordozgoiti, Gomez Canaval, Mozo*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*				*2015*			*yes*
23	*A Review of Scalable Approaches for Frequent Itemset Mining*	*Apiletti, Garza, Pulvirenti*	*East-European Conference on Advances in Databases and Information Systems (ADBIS). International Workshop on Big Data Applications and Principles, BigDAP*				*2015*			*yes*
24	*Network traffic analysis by means of Misleading*	*Apiletti,*	*International Workshop on Big Data Applications and*		*BigDap, Universidad*		*2014*			*yes*

	Generalized itemsets	*Baralis, Cagliero, Cerquitelli, Chiusano, Garza, Grimaudo, Pulvirenti*	*Principles, BigDAP*		*Politecnica de Madrid*					
25	*MGI-Cloud: discovery of misleading generalized itemsets*	*Baralis, Cagliero, Cerquitelli, Chiusano, Garza, Grimaudo, Pulvirenti*	*Italian Symposium on Advanced Database Systems, SEBD*		*Sistemi Evoluti per Basi di Dati*		*2014*			*yes*
26	*Unsupervised detection of network attacks in the dark*	*Owezarski, Casas, Mazel*	*International Workshop on Big Data Applications and Principles, BigDAP*		*BigDap, Universidad Politecnica de Madrid*		*2014*			*yes*
27	*Adaptive Quality of Experience (AQoE) control for Telecom Networks*	*Bascuñana , Castro, Espadas, Sanchez*	*International Workshop on Big Data Applications and Principles, BigDAP*		*BigDap, Universidad Politecnica de Madrid*		*2014*			*yes*
28	*A Survey on feature selection and reduction for network traffic characterization*	*Zhu, Mozo*	*International Workshop on Big Data Applications and Principles, BigDAP*		*BigDap, Universidad Politecnica de Madrid*		*2014*			*yes*
29	*A Telecom Analytics Framework for Dynamic Quality of Service Management*	*Guadamillas, López, Maravitsas, Mozo Pulvirenti*	*International Workshop on Big Data Applications and Principles, BigDAP*		*BigDap, Universidad Politecnica de Madrid*		*2014*			*yes*

TEMPLATE A2: LIST OF DISSEMINATION ACTIVITIES								
NO.	Type of activities[4]	Main leader	Title	Date/Period	Place	Type of audience[5]	Size of audience	Countries addressed
1	*Workshop*	*UPM*	*1st International Workshop on Big Data Applications and Principles, BigDAP*	*11-13 September 2014*	*Madrid, Spain*	*Big Data researchers and industrial partners*	*40 attendees*	World Wide
2	*Workshop*	*POLITO, UPM*	*2nd International Workshop on Big Data Applications and Principles, BigDAP. Co-located with East-European Conference on Advances in Databases and Information Systems (ADBIS).*	*8 September 2015*	*Poitiers, France*	*Big Data researchers and industrial partners*	*35 attendees*	World Wide
3	*Workshop*	*POLITO, UPM*	*3rd International Workshop on Big Data Applications and Principles, BigDAP. Co-located with East-European Conference on Advances in Databases and Information Systems (ADBIS).*	*28 August 2016*	*Prague, Czechia*	*Big Data researchers and industrial partners*	*36 attendees*	World Wide
4	*Master Course*	*UPM*	*Massively parallel machine learning (3 ECTS) EIT Digital Master School's Master's Programme in ICT Innovation: Data Science*	*September to December 2016*	*Madrid, Spain*	*Master students*	*23 students*	*World Wide*
5	*Master Course*	*UPM*	*Services and protocols for Internet applications (6 ECTS)*	*September to December 2016*	*Madrid, Spain*	*Master students*	*25 students*	*Spain*

[4] A drop down list allows choosing the dissemination activity: publications, conferences, workshops, web, press releases, flyers, articles published in the popular press, videos, media briefings, presentations, exhibitions, thesis, interviews, films, TV clips, posters, Other.

[5] A drop down list allows choosing the type of public: Scientific Community (higher education, Research), Industry, Civil Society, Policy makers, Medias, Other ('multiple choices' is possible).

			Master's Degree in Software for Distributed and Embedded Systems					
6	*PhD Tesis*	*UPM*	*PhD Thesis mentoring*	*2014-2016*	*Madrid, Spain*	*PhD students*	*2 student*	*World Wide*
7	*Laboratory*	*ERICSSON*	*Network Analytics Labs*	*April to September, 2016*	*Madrid, Spain*	*Master and Posgraduate students*	*15 interns*	*Spain*
8	*PhD Course*		*Big Data and Data Mining*	*September-October 2016*	*Turin, Italy*	*PhD students*	*20 PhD students*	*Italy*
9	*PhD Thesis*	*POLITO*	*PhD Thesis mentoring*	*2014-2016*	*Turin, Italy*	*PhD students*	*1 student*	*World Wide*
10	*Master Course*	*CNRS*	*Course "Bootnets and anomaly detection in Internet Traffic" in Master TLS-SEC*	*September-December 2016*	*Toulouse, France*	*Master students*	*20 students*	*France*
11	*Presentation*	*SATEC*	*Presentation "Fast Data" in Semana del Impulso TIC (Event organized by the Official Associations of Computer Engineers (COIIPA) and Technical Engineers in Informatics (CITIPA)*	*November 2016*	*Avilés, Spain*	*ICT professionals*	*100 ICT professionals*	*Spain*
12	*Booth exhibition*	*Adaptit*	*Mobile World Congress*	*February 2016, 2015*	*Barcelona, Spain*	*ICT professionals*	*>2000 ICT professionals*	*World Wide*
13	*Presentation*	*EMC*	*DELL-EMC forum*	*29th November and 1st December 2015*	*Madrid and Barcelona, Spain*	*ICT professionals (customers and partners)*	*>200 ICT professionals*	*Spain*
14	*Sponsorship*	*UPM*	*Silver Sponsor of the European Data Forum (EDF) 2015*	*November 2015*	*Luxembourg*	*ICT professionals and researchers*	*>300 ICT professionals and researchers*	*World Wide*
15	*Presentation and Poster*	*ERICSSON*	*Ericsson Data Research Days*	*September 2014*	*Madrid, Spain*	*ICT professionals and researchers*	*>340 ICT professionals and researchers*	*World Wide*

16	*Presentation*	*ERICSSON*	*Ericsson Madrid R&D Innovation Days*	*November 2014*	*Madrid, Spain*	*ICT professionals and researchers*	*>700 ICT professionals and researchers*	*Spain*

Section B (public)

Part B1

TEMPLATE B1: LIST OF APPLICATIONS FOR PATENTS, TRADEMARKS, REGISTERED DESIGNS, ETC.					
Type of IP Rights[6]:	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Application reference(s) (e.g. EP123456)	Subject or title of application	Applicant (s) (as on the application)
Patent	NO	N/A	PCT/EP2014/070566	Congestion Mitigation by Offloading to non-3GPP Networks	Telefonaktiebolaget LM Ericsson

[6] A drop down list allows choosing the type of IP rights: Patents, Trademarks, Registered designs, Utility models, Others.

Part B2

Type of Exploitable Foreground[7]	Description of exploitable foreground	Confidential Click on YES/NO	Foreseen embargo date dd/mm/yyyy	Exploitable product(s) or measure(s)	Sector(s) of application[8]	Timetable, commercial or any other use	Patents or other IPR exploitation (licences)	Owner & Other Beneficiary(s) involved
General Advancement of Knowledge	*Adaptive Quality of Experience Control Proof-of-Concept*	*NO*	*N/A*	*Software components*	*Telecommunication*	*From 2015*	*NO*	*Ericsson*
General advancement of knowledge	*Practical applications of UC#3: multipurpose real-time analytic dashboard*	*NO*	*Not defined*	*Custom development*	*1.-Telecomunications 2.- Industry 3.- Public Administration*	*2017-2018*	*NO*	*SATEC*
General advancement of knowledge	*Practical applications of UC2: Video degradation detection*	*NO*	*Not defined*	*Custom development*	*1.- Telecommunications*	*2017->*	*NO*	*EMC*
General Advancement of Knowledge	*Techniques for a) stochastically characterising network traffic in terms of key statistical measures and their evolution over time, and b) generating synthetic traffic bearing input dynamics*	*NO*	*Not Defined*	*Enhanced analytics for offered products and customer solutions*	*1. Telecommunications*	*2017*	*Not currently foreseen*	*ADAPTIT*

[19] A drop down list allows choosing the type of foreground: General advancement of knowledge, Commercial exploitation of R&D results, Exploitation of R&D results via standards, exploitation of results through EU policies, exploitation of results through (social) innovation.

[8] A drop down list allows choosing the type sector (NACE nomenclature) : http://ec.europa.eu/competition/mergers/cases/index/nace_all.html