

Grant Agreement No.: 258378

FIGARO

Future Internet Gateway-based Architecture of Residential Networks



Instrument: **Collaborative Project**

Thematic Priority: **THEME [ICT-2009.1.1] The Network of the Future**

D2.1 Report on the Specification of metrics and data formats

Due date of deliverable: 30.09.2011

Actual submission date: 30.09.2011

Start date of project: October 1st 2010

Duration: 36 months

Project Manager: Henrik Lundgren, Technicolor R&D Paris

Revision: v1.0

Abstract

This document lists the functionality of the FIGARO monitoring module and the format to store and query data. We survey the monitoring requirements of all use-cases from the FIGARO project that require network and system monitoring. From these use-cases we derive a unified set of measurements and their configuration in terms of vantage points, the amount of history to keep, and types of aggregation of multiple measurements. We then discuss our reasoning to pick a standard format for data collection and define the meta-data to be used by the FIGARO monitoring module.

Project co-funded by the European Commission in the 7th Framework Programme (2007-2013)		
Dissemination Level		
PU	Public	✓
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

Document Revision History

Version	Date	Description of change	Editor	Authors
V1.0	30.09.2011	Final version submitted to the EC	UPMC	UPMC, TRDP, POLITO, GUAVUS, MARTEL (formal check and review)

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

Table of Contents

1	INTRODUCTION.....	3
2	MONITORING REQUIREMENTS PER USE CASE.....	4
2.1	NETWORK ACCESS BUNDLING & WIRELESS NEIGHBORHOOD OPTIMIZATIONS	4
2.2	CONTENT ACCESS AND HOME AUTOMATION	5
2.3	BACKUP SERVICE	5
2.4	GATEWAY-ASSISTED VIDEO-STREAMING OPTIMIZATIONS.....	6
2.5	TROUBLESHOOTING PERFORMANCE PROBLEMS.....	6
2.6	DETECTION OF POOR PERFORMANCE & PERFORMANCE PROFILES.....	7
3	MONITORING REQUIREMENTS	9
3.1	NETWORK PERFORMANCE METRICS	9
3.2	TRAFFIC CHARACTERISTICS.....	14
3.3	GATEWAY STATUS.....	15
4	DATA-FORMATS	17
A.	APPENDIX	21
1)	SURVEY.....	21

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

1 INTRODUCTION

In the FIGARO project the gateway-centric network monitoring, under the scope of WP2, has three main tasks:

- The monitoring of network characteristics of the Internet access link(s) and the home network interfaces (e.g. Wireless, Bluetooth, and Ethernet).
- The monitoring of services and applications, which exchange traffic through or with the gateway.
- Troubleshooting of performance and connectivity problems.

All of these tasks require multiple different types of measurements, which constitute the functionality of the monitoring module. This deliverable has two main goals:

1. List the functionality that the FIGARO monitoring module has to provide.
2. Define which data formats to use when storing and/or exchanging information.

In particular, the ultimate objective is to identify the measurement metrics specific for FIGARO. We therefore surveyed the FIGARO project partners to obtain a list of requirements, rather than undertaking a broad literature survey. We asked all partners working on use cases that require monitoring to answer a survey about the monitoring requirements of each use case. The survey consisted of three parts: (i) network performance metrics, (ii) characteristics of traffic traversing the gateway, and (iii) information about the gateway status and attached devices. In each part, we asked partners to elaborate on which networks should be monitored (the access link or the home network), the measurement frequency and granularity, and how long to keep the results. Please refer to Appendix A for the entire list of survey questions.

In FIGARO, all data collection and processing occurs in the gateway. It is out of the scope of this deliverable to explain how measurement techniques and algorithms work, or to discuss the implementation details for the identified measurements. We will design and implement the functionality defined in this deliverable in the rest of the project.

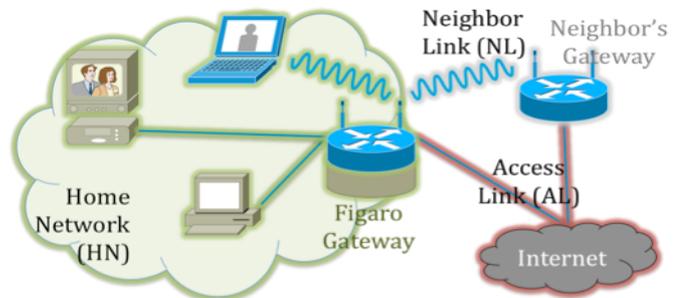
The remainder of this deliverable is organized as follows. Section 2 presents the survey results per use case. We consolidate the requirements of the different use cases into a unified list of requirements for the FIGARO monitoring module in Section 3. Finally, Section 4 justifies and presents the selected scheme for data formats.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

2 MONITORING REQUIREMENTS PER USE CASE

This section summarizes the monitoring requirements of each use case. We differentiate between different vantage points for the measurements:

- The Home Network (green, HN) often called LAN on home gateways, the
- Internet Access Link (red, AL) typically DSL or Cable, and the
- Link to a Neighbor (gray, NL), which is usually WLAN.



2.1 Network access bundling & Wireless Neighborhood Optimizations

FIGARO federated networking solutions (WP3) aims to utilize neighboring home gateways and their Internet access link. We envision three different cases of application: (i) Backhaul bandwidth aggregation: Because most of the time the access bandwidth is not used concurrently, the FIGARO home gateway will provide mechanisms to utilize unused access bandwidth of neighboring home gateways to increase the obtained bandwidth of a single user. (ii) Eco-Management: Focusing the load of multiple clients over one access point (or gateway) can help to reduce cost, as unused components can be turned off. (iii) Enabling Multipath at home: As one home can now use multiple access links services can use multipath solutions increase both performance and reliability at the same time.

In addition to using another gateway's Internet access FIGARO also aims to optimize the wireless neighborhood with the goal of increasing performance by interference mitigation and transmission power control. This use-case requires similar metrics to be measured.

All of these cases require a good understanding of the capacity and status of the local and the neighboring networks, and thus have almost the same requirements with slightly varying metrics, measurement frequencies and history:

- **Network Performance Metrics:** Available bandwidth, Achieved throughput, MAC-layer achieved goodput, and, Internet access synchronization rate for both uplink and downlink (if possible distinguish between local traffic and community backhauling), wireless physical bit rate (of the Gateway and averaged over all wireless users), (sampled) channel busy time, SNR, RSSI, IP-layer packet error rate, MAC-layer frame error rate, MAC-layer average frame payload size.
 - **Vantage Points:** HN, AL, NL (to neighboring gateways), AL at neighboring gateways
 - **Frequency:** Continuously, every few seconds
 - **History:** last 1 minute to last hour
 - **Data aggregation:** Aggregated (average, variance, median, min, max)

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

2.2 Content Access and Home Automation

The FIGARO Energy management, e-health and home automation solutions (in WP5) will retrieve information from devices attached to the gateway. These devices in turn manage their respective sensors and actuators. For a FIGARO-wide management solution it is necessary to know when these devices are operational. Likewise, the unified and remote content access as well as the backup service needs to know when gateways and attached storage systems are accessible and online.

Thus we need to capture the system behavior in terms of gateway availability, as well as device availability behind the GW. We also need to capture the amount of available disk space we can use for cloud storage purposes and how it evolves over time.

- **Gateway Status:** On-periods of the gateway, its HN interfaces, its AL and NLs; On-periods of attached devices available
 - **Frequency:** Continuously, update every minute
 - **History:** last year, or longer
 - **Data aggregation:** Raw
- **Gateway Status:** Available storage capacity
 - **Frequency:** Continuously, once a day
 - **History:** last year, or longer
 - **Data aggregation:** Raw

2.3 Backup Service

Content stored in the FIGARO unified content service (WP4) needs to be backed-up. In order to select which gateways to use for backup and when to perform the backups, it is important to know both network performance towards the backup gateways and the disk utilization of the candidate backup gateway. As backup is a fairly time-independent service, an extensive history of measurements enables locating the best timeframe for backup. This allows minimizing the impact of backup on regular user traffic. In addition to its own measurements, the backup service can take advantage of FIGARO's long-term gateway and storage availability measurements from use-case 2.2.

- **Network Performance Metrics:** Available bandwidth, (maximum) throughput
 - **Vantage Points:** HN, AL, NL
 - **Frequency:** Continuously (when backup is running), once a minute
 - **History:** last day
 - **Data aggregation:** Raw and aggregated (average, variance, median)
- **Network Performance Metrics:** Reachability
 - **Vantage Points:** HN, AL, NL
 - **Frequency:** On-demand (when backup is running), once a minute
 - **History:** last day
 - **Data aggregation:** Raw
- **Gateway Status:** Available storage (local and at candidate backup gateways)

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

- **Frequency:** On-demand, one-shot
- **History:** several days
- **Data Aggregation:** raw

2.4 Gateway-assisted video-streaming optimizations

Optimizing the in-home network performance and increasing the user quality of experience is key in home networking. Therefore, we consider content-aware optimization of the home network, with wireless video streaming as a prime use case. In this case, we perform optimization of the wireless network operation/configuration in order to best support the content currently being transported over the wireless network. This requires monitoring the state and performance of the wireless network as well as information about the content transported. The wireless network has to be continuously monitored to evaluate its varying performance and match it against the content requirements. In the particular case of video streaming with variable bit rate this requires continuous information about the near-future video streaming rate.

- **Network Performance Metrics:** Link capacity, available bandwidth, packet error rate, frame error rate, Channel Busy Time, wireless physical bit rate
 - **Vantage Points:** HN (wireless interface)
 - **Frequency:** Continuously, every second; Or per packet or per set of packets (frame error rate, bit rate)
 - **History:** none
 - **Data aggregation:** Raw and aggregated (average, variance, median, min, max)
- **Traffic characteristics:** Application/service performance
 - **Frequency:** Continuously, for detecting video transmissions
 - **History:** none
- **Traffic characteristics:** Application and service usage, video streaming info and status
 - **Frequency:** On-Demand (when video transmission is active), continuous monitoring
 - **History:** none

2.5 Troubleshooting performance problems

FIGARO will have a module to identify the network responsible for faults and performance problems (WP2). First, we need to distinguish whether the problem comes from inside or outside the home network. We will use network tomography techniques that combine passive measurements of active connections and triggered active measurements to infer the location of the problem. Second, when the problem is inside the home, we need to determine whether the problem originates from within the local home network or neighboring WiFi networks. For this step FIGARO implements an in-home network topology mapping system, which reveals the physical interconnectivity at home. If the problem involves WLAN, we can take advantage of several Wireless network performance metrics already collected for use-cases 2.1 and 2.4 to further pinpoint the problem.

As an extension to “on-demand” (i.e. after a user request) troubleshooting we also envision the monitoring module to detect degradation of application and service performance automatically (see 2.6). Once such degradation is detected, the troubleshooting procedure is launched.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

- **Network Performance Metrics:** Reachability, available bandwidth of up/down link and latency + Wireless metrics from use-cases 2.1 and 2.4.
 - **Vantage Points:** HN, AL, NL
 - **Frequency:** On-Demand, one-shot measurements to support failure localization
 - **History:** until problem is resolved
 - **Data aggregation:** raw
- **Traffic characteristics:** Application/service performance
 - **Vantage Point:** analyzing traffic passing through the gateway passively
 - **Frequency:** On-Demand, continuous monitoring
 - **History:** until problem is resolved
 - **Data aggregation:** extended flow-records including flow identifiers (IPs, Ports, transport protocol), the detected application/service and flow performance indicators (RTTs, data rates, retransmissions, resets, jitter)
- **Gateway Status:** Home Network Topology Mapping
 - **Vantage Point:** HN
 - **Frequency:** On-Demand
 - **History:** none
 - **Data aggregation:** none
- **Gateway Status:** CPU and memory utilization
 - **Frequency:** Continuously, every minute
 - **History:** 1 day
 - **Data aggregation:** Average, variance, minimum, maximum, median

2.6 Detection of poor performance & performance profiles

As mentioned in the previous use-case FIGARO aims to implement a continuous monitoring of application performance, both to inform the user about the current state of its services and to detect and performance degradation. For that FIGARO gateways will monitor traffic passing through them and build application and service performance profiles.

Correctly identifying the application/service for each connection is important for building the application usage profiles. Different applications have largely different performance requirements. Thus while huge delay does not affect large file-transfers as long as enough bandwidth is available, Internet telephony only requires minimal bandwidth but very low latency.

- **Traffic characteristics:** Application/service performance
 - **Frequency:** Continuously
 - **History:** one month, depending on the number of record sampling might be required
 - **Data aggregation:** extended flow-records including flow identifiers (IPs, Ports, transport protocol), the detected application/service and flow performance indicators (RTTs, data rates, retransmissions, resets, jitter)
- **Traffic characteristics:** Application/service performance (historic baseline)

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

- **History:** 6 months
- **Data aggregation:** per application/service profiles with average and percentiles in 5%-steps for the flow performance indicators
- **Traffic characteristics:** Failed connection requests
 - **Frequency:** Continuously
 - **History:** one month
 - **Data aggregation:** Fraction of failed vs. total connection requests in 1hour bins
- **Gateway Status:** CPU and memory utilization
 - **Frequency:** Continuously, every minute
 - **History:** 1 day
 - **Data aggregation:** Average, variance, minimum, maximum, median

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

3 MONITORING REQUIREMENTS

The use cases discussed in the previous section lead to the following monitoring requirements. Metrics, such as available bandwidth, reachability and latency require a list of targets as input. This list needs to be supplied by the use-case and will typically include a set of other FIGARO gateway or content sources in the Internet which are currently accessed. For metrics that are bi-directional both the source and the destination IP addresses need to be specified. Typically one of the IP addresses will be the gateways IP address. Swapping these IP addresses allows specifying the direction.

All of the continuous monitoring requirements include the history length H as an input parameter. This parameter will determine for how long measurements will be stored at the gateway before they can be evicted. As some use-cases might not be interested in the whole history, we foresee the option of specifying the desired history length HQ_i individually for each query. Of course replies will only return the minimum of H and HQ_i , this includes that aggregated results (i.e. average over multiple measurements) is also only calculated on the minimum of H and HQ_i .

3.1 Network performance metrics

Available Bandwidth

On-demand

- Unit: bits/s
- Input parameters: source IP, S ; destination IP, D ; duration for one measurement sample, T ;
- Output: average residual bandwidth of the bottleneck link on a path between S and D during T

Continuous

- Unit: bits/s
- Input parameters: set of pairs (S,D) , where S is the source IP and D is the destination IP; duration for one measurement sample T ; history length, H ; measurement frequency, F .
- Output: A list of tuples with one tuple $(num, avg_bw, var_bw, min_bw, median_bw, max_bw)$ per (S,D) , where num is the number of successful samples during H . Each sample measures the average available bandwidth between S and D during T . avg_bw is the average of all samples during H , similar for variance, minimum, median, and maximum, respectively.
- Notes: T has to be smaller than $1/F$; Parameters for 2.1: $H = 1h$ and $F = 1/(5sec)$; Parameters for 2.3: $H=1d$ and $F=1/60sec$; Possibility to start/stop a continuous measurement.

Reachability

On-demand

- Unit: Boolean
- Input parameters: source IP, S ; destination IP, D ;
- Output: True if a routed network path exists from S to D . False otherwise.
- Notes: for 2.3 the measurement should not take longer than 60 seconds

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

Latency

On-demand

- Unit: ms (milliseconds)
- Input parameters: destination IP, D;
- Output: Round-trip time for the path from the gateway to D and back.

Achieved Throughput (all data sent and received; need not be successful)

Continuous

- Unit: bits/s
- Input parameters: set of pairs (S,D), where S is the source IP and D is the destination IP; history length, H; measurement frequency, F.
- Output: A list of throughput values and a list of tuples with one tuple (num, avg_tput, var_tput, min_tput, median_tput, max_tput) per (S,D), where num is the number of successful samples during H. Each sample measures the average throughput between S and D for a duration of 1/F. avg_tput is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Parameters for 2.1: H = 1h and F = 1/5sec; Parameters for 2.3: H=1d and F=1/(60sec); Possibility to start/stop a continuous measurements.

MAC-layer Achieved Goodput

Continuous

- Unit: bits/s
- Input parameters: set of pairs (S,D), where S is the source MAC address and D is the destination MAC address; history length, H; measurement frequency, F.
- Output: A list of throughput values and a list of tuples with one tuple (avg_mgood_ul_el, avg_mgood_ul, avg_mgood_dl_el, avg_mgood_dl) per (S,D), where avg_mgood measures the average goodput between S and D for a duration of 1/F. The goodput is measured both uplink (ul) and downlink, and divided in elastic (el) and not elastic. The elastic goodput is given by transport protocol implementing congestion control (like TCP) while the inelastic goodput results from transport protocol not implementing it (like UDP) Notes: Parameters for 2.1: H = 1h and F = 1/5sec; Parameters for 2.3: H=1d and F=1/(60sec); Possibility to start/stop a continuous measurements. Goodput can be overestimated in the uplink direction due to incorrectly received ACKs leading to retransmissions. Information can be queried from system/interface driver.

Internet Access Synchronization Rate

Continuous

- Unit: bits/s
- Input parameters: Gateway Interface, I (e.g. DSL, 3G); history length, H; measurement frequency, F.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

- Output: A list of tuples with one tuple (num, avg_sync, var_sync, min_sync, median_sync, max_sync) per (S,D), where num is the number of samples during H. Each sample measures the current synchronization rate of I. avg_sync is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Parameters for 2.1: H = 1h and F = 1/(5sec); Information can be queried from system/interface driver.

Wireless Physical Layer Bit Rate

Continuous

- Unit: bits/s
- Input parameters: set of pairs (S,D), where S is the source MAC address and D is the destination MAC address; history length, H; measurement frequency, F.
- Output: A list of tuples with one tuple (num, avg_plbr, var_plbr, min_plbr, median_plbr, max_plbr) per (S,D), where num is the number of samples during H. Each sample measures the current wireless physical layer bit rate. avg_plbr is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Frequency can be specified per N packets or per time period. Parameters for 2.1 and 2.4: H = 1h and F = 1/pkt. 1/(10 pkts), or 1/(1sec); Information can be queried from system/interface driver.

Wireless Overall Average Physical Layer Bit Rate

Continuous

- Unit: bits/s
- Input parameters: history length, H; measurement frequency, F; WiFi-Network identifier, BSSid.
- Output: A list of tuples with one tuple (num, avg_oplbr, var_oplbr, min_oplbr, median_oplbr, max_oplbr) for the whole BSS; where num is the number of samples during H. Each sample measures the current wireless physical layer bit rate of successfully transmitted frames. avg_oplbr is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Frequency can be specified per N packets or per time period. Parameters for 2.1 and 2.4: H = 1h and F = 1/pkt. 1/(10 pkts), or 1/(1sec); Information can be queried from system/interface driver. In case the Gateway manages multiple BSSs, separate, per-BSS metrics can be obtained through the virtualization of the wireless card.

Channel Busy Time

Continuous

- Unit: Fraction or Percentage
- Input parameters: Gateway Interface, I (e.g. DSL, WLAN, 3G); history length, H; measurement frequency, F; WiFi-Network identifier, BSSid.
- Output: A list of tuples with one tuple (num, avg_cbs, var_cbs, min_cbs, median_cbs, max_cbs) per (S,D), where num is the number of samples during H. Both the proprietary

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

Atheros interface and IEEE 802.11k report the channel busy fraction, i.e., the fraction of time in which the wireless channel is sensed busy due to own or neighbor activity. Each sample measures the current channel busy time of I. avg_cbs is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.

- Notes: Parameters for 2.1: $H = 1h$ and $F = 1/(5sec)$; Information can be queried from system/interface driver. In case the Gateway manages multiple BSSs, separate, per-BSS metrics can be obtained through the virtualization of the wireless card.

IP-layer Packet Error Rate (PER)

Continuous

- Unit: Percentage or fraction.
- Input parameters: set of pairs (S,D), where S is the source MAC address and D is the destination MAC address; history length, H; measurement frequency, F.
- Output: A list of tuples with one tuple (num, avg_per, var_per, min_per, median_per, max_per) per (S,D), where num is the number of samples during H. Each sample measures the Packet Error Rate for a duration of $1/F$. avg_per is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Parameters for 2.1: $H = 1h$ and $F = 1/(5sec)$.

MAC-layer Frame Error Rate (FER)

Continuous

- Unit: Percentage or fraction.
- Input parameters: set of pairs (S,D), where S is the source MAC address and D is the destination MAC address; history length, H; measurement frequency, F.
- Output: A list of tuples with one tuple (num, avg_fer, var_per, min_fer, median_fer, max_fer) per (S,D), where num is the number of samples during H. Each sample measures the MAC-layer Frame Error Rate for a duration of $1/F$. avg_fer is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Typical parameters for 2.1 and 2.4: $H = 1h$ and $F = 1/(2sec)$, or per IP-layer packet. Information can be queried from system/interface driver.

MAC-layer Frame Error Rate due to channel error (FERc)

Continuous

- Unit: Percentage or fraction.
- Input parameters: set of pairs (S,D), where S is the source MAC address and D is the destination MAC address; history length, H; measurement frequency, F.
- Output: A list of tuples with one tuple (est_ferc) per (S,D). est_ferc is the estimated frame error rate due to bad channel conditions, and it has to be estimated given the bitrate at which S and D are communicating, their SNR, and assuming independent bit errors on the channel.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

- Notes: Typical parameters for 2.1 and 2.4: $F = 1/(2\text{sec})$, or per IP-layer packet. Information can be queried from system/interface driver.

MAC-layer Average Frame Payload size (P)

Continuous

- Unit: bits
- Input parameters: history length, H; measurement frequency, F; WiFi-Network identifier, BSSid.
- Output: One tuple (num, avg_p, var_p, min_p, median_p, max_p) for the whole BSS, where num is the number of samples during H. Each sample measures the size of the MAC-layer payload that have been successfully transmitted within the BSS for a duration of $1/F$. avg_p is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Typical parameters for 2.1 and 2.4: $F = 1/(2\text{sec})$. Information can be queried from system/interface driver. In case the Gateway manages multiple BSSs, separate, per-BSS metrics can be obtained through the virtualization of the wireless card.

Signal-to-Noise Ratio (SNR)

Continuous

- Unit: None (ratio)
- Input parameters: set of pairs (S,D), where S is the source MAC address and D is the destination MAC address; history length, H; measurement frequency, F.
- Output: A list of tuples with one tuple (num, avg_snr, var_snr, min_snr, median_snr, max_snr) per (S,D), where num is the number of samples during H. Each sample measures the SNR for a duration of $1/F$ for a specific S-D pair. avg_snr is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Typical parameters for 2.1 and 2.4: $H = 1\text{min}$ and $F = 1/(2\text{sec})$ or per packet; Information can be queried from system/interface driver.

Received Signal Strength Indicator (RSSI)

Continuous

- Unit: dBm.
- Input parameters: set of pairs (S,D), where S is the source MAC address and D is the destination MAC address; history length, H; measurement frequency, F.
- Output: A list of tuples with one tuple (num, avg_rssi, var_rssi, min_rssi, median_rssi, max_rssi) per (S,D), where num is the number of samples during H. Each sample measures the RSSI for a duration of $1/F$ for a specific S-D pair. avg_rssi is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Typical parameters for 2.1 and 2.4: $H = 1\text{min}$ and $F = 1/(2\text{sec})$ or per packet; Information can be queried from system/interface driver.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

3.2 Traffic characteristics

Application/Service Performance

Continuous

- Input parameters: history length, H; detailed history length, HR; storage, S; connection timeout, TO; connection idle timeout, ITO;
- Output:
 - A list of annotated bi-directional flow/connection records for flows in HR. This needs to be sampled when using more than S disk space.
 - A list of service/application profiles from flows in H.
- Annotated flow records contain the following information per flow/connection:
 - Flow/Connection identifiers: IP addresses and Port numbers of source and destination, transport protocol
 - Application/Service information:
 - Application Layer protocol (e.g. HTTP, FTP, ...)
 - Content-type or flow classification (e.g. gaming, video, chat)
 - Hostname when application layer protocol is HTTP
 - “Is the flow still active”-Flag (if so the following information only includes traffic seen up to “now”)
 - Flow/Connection statistics: Number of packets, volume in Bytes, start and end time (using a timeout of TO)
 - Performance indicators: Idle Time (using a timeout of ITO), RTT (avg/var/min/median/max), Data Rate (avg/var/min/median/max), packet loss (e.g. from number of TCP retransmissions)
- Service/Application profiles contain the following information per service/application:
 - Average and percentiles in 5%-steps of all flow statistics and performance indicators from all flows labeled as the respective service/application
- Notes: Suggested parameters for 2.5: H = 6months; HR = 1month; TO = 20 min; ITO = 10sec.

Application and Service Usage (Application Mix)

On-Demand

- Input parameters: Start and End Time, ST & ET; List of applications, APPS; Ranking Metric, M;
- Output: A list of applications (APPS + “other”) annotated and ranked by the overall contribution according to metric M for the period starting at ST and ending at ET. The contribution and rank is calculated from the collected application and service performance records.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

Failed connection requests

Continuous

- Unit: Fraction of failed connection attempts (over total connection attempts)
- Input parameters: Aggregation bin size, S; History length, H;
- Output: A list of fractions of connection attempts for each of the H/S bins. Count the number of failed and total connection attempts towards the Internet, which are passing through or originating from the gateway.
- Notes: Parameters for 2.6: S = 1hour, H = 1month

Video stream information and status

On-Demand

- Unit: bit/s
- Input parameters: Video stream identifier, VSI; video time offset, VTO;
- Output: Video-streaming rate for VSI for time period from VTO to VTO+1, recorded at application layer.

3.3 Gateway status

Available Storage Capacity

On-demand

- Unit: Bytes
- Input parameters: FIGARO gateway; G;
- Output: Available Storage Capacity of Gateway G.

Continuous

- Unit: Bytes
- Input parameters: Storage Device, D; history length, H; measurement frequency, F.
- Output: A list of storage capacity values. Each sample measures the currently available storage capacity of D.
- Notes: Parameters for 2.2: H = 1y and F = 1/1day;

On-Periods (of in-home devices, the gateway, all its HN interfaces, its AL and NLs)

Continuous

- Unit: none; time periods
- Input parameters: Target Device or Interface, T; history length, H; detailed history length, HR; update frequency, F.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

- Output: A list of periods when T is available/online during HR and one value for the uptime ratio UR during H. The list is updated by checking for the status with frequency F. The uptime ratio is the sum of the durations of the available/online periods divided by the total measurement duration.
- Notes: Parameters for 2.2: HR = 1 month, H = 1y, and F = 1/60sec;

CPU & memory utilization

Continuous

- Unit: Fraction or percentage
- Input parameters: History length, H; measurement frequency, F.
- Output: One tuple (cur_cpu, avg_cpu, var_cpu, min_cpu, median_cpu, max_cpu) for CPU utilization and one tuple (cur_mem, avg_mem, var_mem, min_mem, median_mem, max_mem). cur_cpu/cur_mem is the average CPU/memory utilization in the last measurement interval of length 1/F. avg_cpu/avg_mem is the average of all samples during H, similar for variance, minimum, median, and maximum, respectively.
- Notes: Parameters for 2.5 and 2.6: H = 1day and F = 1/60sec;

Home Network Topology mapping

On-Demand

- Input parameters: Detection Timeout, TO.
- Output: A list of tuples (home device, functionality) of all home devices and an adjacency matrix denoting the links between the home devices. Functionality is from the following list [HG, HNID, STA] (Home Gateway, Home Network Infrastructure Device, Station/Endhost).
- Notes: Suggested Parameter TO = 2sec;

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

4 DATA-FORMATS

The breadth in monitoring requirements as presented in the previous section motivates the need to have a principled design for data formats. Ideally, the optimal data format should satisfy the following design criteria:

1. **Extensibility:** The data format should be extensible so that if the monitoring metrics evolve over time or new metrics are added, the data formats do not need to be re-designed. This motivates the need for a template-based data format, where the meta-data dictionary is decoupled from the data encoding and data transmission primitives.
2. **Lightweight:** Because gateways have limited computation, the format should not incur much over-head to create. For example, we consider the computational overhead that is typically associated with XML-style encodings, as to high.
3. **Standards-based:** The format should be compliant with existing standards, so that we can borrow from a rich ecosystem of processing tools and be compliant with trends in industry.

Guavus has a consolidated experience working with a number of different network data format definitions and standards. Given the above design considerations, we found the IETF IPFIX standard to be a suitable fit for the following reasons: First, IPFIX was designed to be relatively lightweight and is used in modern routers today to capture high-speed flow data. Second, IPFIX has emerged as a popular standard, with a large body of RFCs, and a thriving ecosystem of data collection & processing tools. And third, most importantly, IPFIX is template-driven, and decouples data format, data encoding, and data dictionary, and therefore allows for maximum flexibility.

This flexibility of IPFIX is accomplished through careful abstraction. The key notions in IPFIX are *Template*, *Data Record*, and *Transport Session*. A *Template* can be thought of as a data dictionary – it is an ordered set of <type, length> pairs that specify both the structure & semantics of the information that needs to be transmitted from the IPFIX device (in this case the home gateway). The *Data Record* is a record that contains the values of the parameters that correspond to a template. Finally, the *Transport Session* specifies the transport protocol, e.g. UDP, TCP, or SCTP. In this manner, IPFIX separates data formats, data transmission and data dictionary. A detailed description of IPFIX can be found in RFC 5101.

Bringing the IPFIX format to the identified measurements we represent the measurement output definitions (see the Output Bullets) in the IPFIX format. Instead of reporting a complete list of IPFIX template we just give an example for the *continuous traffic characteristics* metric (Section 3.2). The output of the other measurement will be mapped alike.

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

As described in Section 3.2, each record of the *continuous traffic characteristic metric* would contain the following attributes:

Attribute	Format
Start Time	UTC
End Time	UTC
Application Protocol	HTTP, FTP, etc.
Content Type	MIME
URL	String
Source IP	IP v4/6
Source Port	Numeric
Destination IP	IP v4/6
Destination Port	Numeric
UpLink	Numeric (Bytes, Packets)
Downlink	Numeric (Bytes, Packets)
RTT	Numeric (Avg, Var, Min, Median, Max)
Packet Loss	Numeric (Avg, Var, Min, Median, Max)

Some of these attributes are already defined in in RFC 5102 (Information Model for IP Flow Information Export) and so can be used directly:

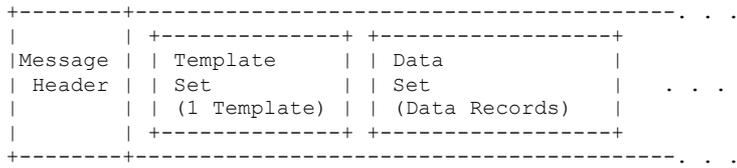
Attribute	IPFIX Element ID	IPFIX Element Name	Group	Type
StartTime	150	flowStartSeconds	timestamp	dateTimeSeconds
EndTime	151	flowEndSeconds	timestamp	dateTimeSeconds
SourceIPv4	8	sourceIPv4Address	ipHeader	ipv4Address
SourceIPv6	27	sourceIPv6Address	ipHeader	ipv6Address
Source Port	7	sourceTransportPort	transportHeader	unsigned16
DestinationIPv4	12	destinationIPv4Address	ipHeader	ipv4Address
DestinationIPv6	28	destinationIPv6Address	ipHeader	ipv6Address
Destination Port	11	destinationTransportPort	transportHeader	unsigned16

For the remaining attributes, Figaro *enterprise-specific*¹ elements need to be defined, as follows:

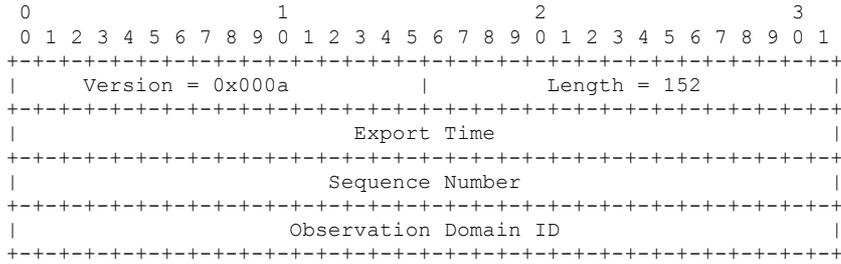
Attribute	Enterprise Element ID	Enterprise Element Name	Type
Application Protocol	2	applicationProtocol	unsigned8
Content Type	4	contentType	string
URL	6	url	string
Uplink	8	uplink	unsigned64
Downlink	9	Downlink	unsigned64
RTT	10	rtt	unsigned64
PacketLoss	11	packetLoss	unsigned64

¹ An enterprise ID can be obtained from IANA.

IPFIX Message Format:



IPFIX Message Header:



v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

Flow Records Template Set:

The following template is proposed to capture the continuous traffic metric of Section 3.2.

0								1								2								3													
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7
Version = 0x000a (10)																Length (148)																					
Export Time (UTC)																																					
Sequence Number																																					
Observation Domain ID																																					
Set ID (2) [Template]																Set Length (bytes)																					
Template ID (256)																Field Count (20)																					
0	flowStartSeconds (150)															Length (4)																					
0	flowEndSeconds (151)															Length (4)																					
0	sourceIPv4Address (8)															Length (4)																					
0	sourceIPv6Address (26)															Length (16)																					
0	sourceTransportPort (7)															Length (2)																					
0	destinationIPv4Address (12)															Length (4)																					
0	destinationIPv6Address (28)															Length (16)																					
0	destinationTransportPort (11)															Length (2)																					
[Figaro Enterprise ID#]																																					
1	applicationProtocol (2)															Length (1)																					
[Figaro Enterprise ID#]																																					
1	contentType (4)															Length (65535)																					
[Figaro Enterprise ID#]																																					
1	url (6)															Length (65535)																					
[Figaro Enterprise ID#]																																					
1	uplink (8)															Length (8)																					
[Figaro Enterprise ID#]																																					
1	downlink (9)															Length (8)																					
[Figaro Enterprise ID#]																																					
1	rtt (10)															Length (8)																					
[Figaro Enterprise ID#]																																					
1	packetLoss (11)															Length (8)																					
[Figaro Enterprise ID#]																																					

v1.0	<i>FIGARO</i> D2.1 Report on the Specification of metrics and data formats	
------	---	--

A. APPENDIX

1) Survey

UPMC conducted the following survey among the FIGARO partners. We received five answers from Eurecom, Polito, UPMC, Technicolor Rennes, and Technicolor Paris. Later we extended the use-cases from discussions in plenary and architecture meetings as well as from the requirements deliverable D1.1. These were the questions:

Case 1: your application needs network performance measurements

1. Which metrics of performance does your application need?

If your application uses more than one metric, please answer questions 2-5 for each metric.

2. For which networks do you need these metrics?
 - a. home network (wired)
 - b. home network (wireless)
 - c. home access link
 - d. network from home to an arbitrary destination connected to the Internet
 - e. neighbor wireless
 - f. neighbor access link
 - g. other. Please specify:
3. How often does your application need updates to the measurements?
4. Does your application require a history of measurements or only current performance? If it requires a history, for how long?
5. In which form you want your measurements reported? For example, an application that needs RTT measurements may require single or aggregated measurements (aggregation could be averages, variance, median, etc.)

Case 2: your application needs characteristics of the traffic traversing the home gateway

6. Which granularity of recording does your application need? Please mark all that suits your application needs.
 - a. per network interface byte and packet counters (SNMP-like)
 - b. flow-level statistics (NetFlow-like). Please give a definition of flow.
 - c. packet TCP/IP headers
 - d. full packets
 - e. application-specific information. Please list:
 - f. other. Please specify
7. What is the ideal way of summarizing network traffic for your application?
8. For how long does your application need this information to be stored at the gateway?
9. How often will your application access this information?

Case 3: Gateway status

10. Which information do you need?

Please answer questions 3-5 for each of them.