

ONTIC: D6.6: Progress on Exploitation and Dissemination Plans – Part II

Miguel Angel López Peña, Miguel-Angel Monjas Erllorente, Alejandro Bascuñana, Bruno Ordozgoiti, Alberto Mozo, Sandra Gomez, Daniele Apileti, Athanasios Tzaferis, Philippe Owezarski, Fernando Arias, et al.

▶ To cite this version:

Miguel Angel López Peña, Miguel-Angel Monjas Erllorente, Alejandro Bascuñana, Bruno Ordozgoiti, Alberto Mozo, et al.. ONTIC: D6.6: Progress on Exploitation and Dissemination Plans – Part II. ONTIC Project (GA number 619633), LAAS-CNRS; SATEC; Ericsson Spain; UPM; Dell-EMC; Politecnico di Torino; ADAPTIT. 2017. hal-01476217

HAL Id: hal-01476217 https://laas.hal.science/hal-01476217

Submitted on 24 Feb 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





Online Network Traffic Characterization

Progress on Exploitation and Dissemination Plans - Part II

ONTIC Project (GA number 619633)

Deliverable D6.6 Dissemination Level: PUBLIC

<u>Authors</u> Miguel Angel López Peña SATEC

Miguel-Angel Monjas Llorente, Alejandro Bascuñana ERICSSON

Bruno Ordozgoiti, Alberto Mozo, Sandra Gomez UPM

> Daniele Apileti POLITECNICO DI TORINO

> > Athanasios Tzaferis ADAPTIT

Philippe Owezarski CNRS

Fernando Arias, Vicente Martín EMC

Version ONTIC_D6.6_2017.02.03_1.0.DOCX





Version History

Previous version	Modification date	Modified by	Summary
0.0	2016-09-29	SATEC	Structure proposal
0.1	2016-10-16	SATEC	First draft version
0.2	2016-12-02	SATEC CNRS	Contributions from partners
0.3	2017-01-30	ERICSSON ADAPTIT EMC UPM POLITO SATEC	Contributions from partners
0.4	2017-01-31	ADAPTIT UPM	Second Draft: First review
1.0	2017-02-03	SATEC	QA Review Final version

Quality Assurance:

	Name
Quality Assurance Manager	Miguel Angel López (SATEC)
Reviewer #1	Miguel-Angel Monjas (ERC)
Reviewer #2	Panos Georgatsos (ADAPTIT)





Copyright © 2017, ONTIC Consortium The ONTIC Consortium (http://www.http://ict-ontic.eu/) grants third parties the right to use and distribute all or parts of this document, provided that the ONTIC project and the document are properly referenced.

THIS DOCUMENT IS PROVIDED BY THE COPYRIGHT HOLDERS AND CONTRIBUTORS "AS IS" AND ANY EXPRESS OR IMPLIED WARRANTIES, INCLUDING, BUT NOT LIMITED TO, THE IMPLIED WARRANTIES OF MERCHANTABILITY AND FITNESS FOR A PARTICULAR PURPOSE ARE DISCLAIMED. IN NO EVENT SHALL THE COPYRIGHT OWNER OR CONTRIBUTORS BE LIABLE FOR ANY DIRECT, INDIRECT, INCIDENTAL, SPECIAL, EXEMPLARY, OR CONSEQUENTIAL DAMAGES (INCLUDING, BUT NOT LIMITED TO, PROCUREMENT OF SUBSTITUTE GOODS OR SERVICES; LOSS OF USE, DATA, OR PROFITS; OR BUSINESS INTERRUPTION) HOWEVER CAUSED AND ON ANY THEORY OF LIABILITY, WHETHER IN CONTRACT, STRICT LIABILITY, OR TORT (INCLUDING NEGLIGENCE OR OTHERWISE) ARISING IN ANY WAY OUT OF THE USE OF THIS DOCUMENT, EVEN IF ADVISED OF THE POSSIBILITY OF SUCH DAMAGE



619633 ONTIC. Deliverable D6.6 SEVENTH FRAMEWORK Progress on Exploitation and Dissemination Plans Part II



Table of Contents

1. ACRONYMS AND DEFINITIONS	8
2. Executive Summary	9
3. SUGGESTED PREVIOUS READINGS	10
4. DISSEMINATION ACTIVITIES DURING THIRD PERIOD	11
 4.1 Conferences, workshops and journals 4.1.1 UPM scientific dissemination 4.1.2 POLITO scientific dissemination 4.1.3 CNRS scientific dissemination 4.1.4 SATEC scientific dissemination 	. 11 . 11 . 12
4.2 BigDAP2016 workshop	. 13
4.3 Public Availability of the ONTS dataset4.3.1 ONTS dataset dissemination4.3.2 SynthONTS ground truth	. 17
4.4 ONTIC Datathon	
 4.4.1 Description of ONTIC Datathon	. 18 . 19 . 19 . 20
4.5 Other activities	20
 4.5.1 UPM dissemination activities 4.5.2 Ericsson dissemination activities 4.5.3 POLITO dissemination activities 4.5.4 CNRS dissemination activities 4.5.5 SATEC dissemination activities 4.5.6 ADAPTIT dissemination activities 4.5.7 EMC dissemination activities 	20 21 22 22 22 23
5. Exploitation Activities During Third Period	25
5.1 UPM exploitation activities during the Third period	25
5.2 Ericsson exploitation activities during the Third period	25
5.3 POLITO exploitation activities during the Third period	27
5.4 CNRS exploitation activities during the Third period	. 27
5.5 SATEC exploitation activities during the Third period	. 28
5.6 Adaptit exploitation activities during the Third period	. 29





5.7 EMC exploitation activities during the Third period	30
6. REFERENCES	31



619633 ONTIC. Deliverable D6.6 SEVENTH FRAMEWORK Progress on Exploitation and Dissemination Plans Part II



List of figures

Figure 1: BigDAP16 web page	15
Figure 2: Access from ONTIC portal to BigDAP16 web	
Figure 3: Ericsson Spain Network Analytics Labs launch internal press note	
Figure 4: Paper in a Spanish business newspaper	
Figure 5: Paper and presentation in Spanish event	
Figure 6: Dell-EMC Forum	
Figure 7: Ericsson Machine Learning Summit leaflet	





List of tables

Table 1: Acronyms	8
Table 2: BIGDAP 2016 Important dates	. 14





1. Acronyms and Definitions

Acronyms:

Acronym	Defined as
ACM	Association for Computing Machinery
ADBIS	Advances in Databases and Information Systems
BigDAP	Big Data Applications and Principles
EDF	European Data Forum
FOA	First Office Application
ICDM	International Conference on Data Mining
ICDIM	International Conference on Digital Information Management
IEEE	Institute of Electrical and Electronics Engineers
ISBN	International Standard Book Number
IPR	intellectual property rights
ISP	Internet Service Provider
IXP	Internet eXchange Point
KPI	Key Performance Indicator
NOMS	Network Operations and Management Symposium
ONTIC	Online Network Trafflc Characterization
QoE	Quality of Experience
QoS	Quality of Service
SAPC	smart Policy Control
SDN	Software Defined Network
SLA	Service Level Agreement
R&D&I	Research and Development and Innovation

Table 1: Acronyms





2. Executive Summary

This deliverable presents the progress in the exploitation and dissemination plans for the project [1] [2] [3] during the third period (third year) of the project by describing the activities performed in order to implement these plans and it is and update of deliverable D6.5 [3]

The set of activities include scientific papers, journals and conferences of interest, 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.







3. Suggested previous readings

The contents of this document are the development of the exploitation strategy, the exploitation plans and the dissemination plans. Those plans and strategy have been described in the following documents and therefore its reading is recommended:

- ONTIC. Deliverable D6.3. "Exploitation Strategy". May 2014.[1]
- ONTIC. Deliverable D6.4, "Exploitation and Dissemination Plans". February 2015.[2]
- ONTIC. Deliverable D6.5, "Progress on Exploitation and Dissemination Plans". January 2016. [3]





4. Dissemination Activities During Third Period

4.1 Conferences, workshops and journals

4.1.1 UPM scientific dissemination

Regarding to scientific dissemination, the following list represents the accepted and submitted papers presented by UPM during the third period:

• 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."

East European Conference on Advances in Databases and Information Systems. Springer International Publishing, 2016.

• Zhu, Bo, and Alberto Mozo.

"Spark2Fires: A New Parallel Approximate Subspace Clustering Algorithm."

East European Conference on Advances in Databases and Information Systems. 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).

- A. Mozo, J.L. Lopez-Presa, A. Fernandez Anta
 "A distributed and quiescent max-min fair algorithm for network congestion control". Submitted to Expert Systems With Applications (Q1 ISI/JCR indexed journal).
- B. Ordozgoiti, S. Gomez Canaval, A. Mozo "Probabilistic Leverage Scores for Parallel Unsupervised Feature Selection on Massive High-Dimensional Datasets" Submitted to the 14th International Work-Conference on Artificial Neural Networks, IWANN 2017 (Core B).

4.1.2 **POLITO scientific dissemination**

Specifically addressing scientific dissemination, Polito in the third year has focused on publishing longer and deeper experimental results on international peer-reviewed journals, besides conferences.





In particular, Polito has prepared 3 accepted journal publications:

- Apiletti, Baralis, Cerquitelli, Garza, Giordano, Mellia, Venturini.
 "SeLINA: a Self-Learning Insightful Network Analyzer" IEEE TRANSACTIONS ON NETWORK AND SERVICE MANAGEMENT, vol. 13 n. 3, pp. 696-710 This is a joint work with Marco Mellia, "mPlane" FP7-IP project coordinator.
- Apiletti, Baralis, Cerquitelli, Garza, Pulvirenti
 "Frequent Itemsets Mining for Big Data: a comparative analysis" Elsevier - Big Data Research (accepted, end of January 2017)
- Apiletti, Baralis, Cerquitelli, Garza, Michiardi, Pulvirenti (under review)
 "A Parallel Map-Reduce Algorithm to Efficiently Support Itemset Mining on High Dimensional Data"
 Elsevier Big Data Research (submitted July 2016).

Furthermore, Polito published 3 papers in international conferences:

- 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"

ADBIS 2016 Short Papers and Workshops, BigDAP 2016, DCSA, DC, 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, colocated with the 20th East-European Conference on Advances in Databases and Information Systems, ADBIS 2016, Prague, Czech Republic, August 28, 2016. pp. 137-146.

4.1.3 CNRS scientific dissemination

Regarding scientific dissemination, CNRS has published the following paper in the third period:

 J. Dromard, G. Roudière, P. Owezarski
 "Online and Scalable Unsupervised Network Anomaly Detection Method" IEEE Transaction on Network and Service Management (TNSM).

Philippe Owezarski from CNRS was also co-editor of:

 "Big data for management" Special issue of the IEEE Transaction on Network and System management, October 2016.





Philippe Owezarski was also invited to give the following talk:

"Unsupervised Anomaly Detection in Internet Traffic"
 France-USA workshop on cybersecurity, Atlanta, GA, USA, 12-14 October 2016.

4.1.4 SATEC scientific dissemination

Regarding to scientific dissemination SATEC has presented the following paper in the third period:

 Miguel Angel López Peña, Carlos Area Rúa, Sergio 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). Indexed by IEEE Xplore (<u>http://ieeexplore.ieee.org/document/7829756/</u>)

4.2 **BigDAP2016** workshop

The third International Workshop on Big Data Applications and Principles, BigDAP 2016, is the third Workshop in the row organized by ONTIC. 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.

For enhancing visibility and attendance, BigDAP 2016 was co-located with the 20th East-European Conference on Advances in Databases and Information Systems (ADBIS 2016) at Prague, Czech Republic. It took place on Sunday August 28, 2016 with 5 presentations, including a keynote speech and an invited presentation.

For promoting the dissemination of the work activities and results of the ONTIC project, the specific goals of BigDAP 2016 were:

- To provide a networking space for people in academic institutions and in the industry.
- To enhance the participation of graduate students, especially those enrolled in PhD programs related to Big Data, Big Analytics and Network Science.

The intended audience of the workshop can be grouped as follows:

- Industrial sector: professionals involved in R&D&I (Research, Development and Innovation) activities.
- Enterprise sector: representatives of SMEs and large corporations interested in improving their organizational structure and business intelligence by leveraging novel Big Data technologies and analytics.
- Academics and researchers whose research interests focus on activities carried out by ONTIC.
- PhD students who are working on a thesis related to the ONTIC activities.

BigDAP 2016 consisted of one track that gathered academics perspectives. The session was structured in three parts separated by breaks. The first one was a keynote speech, the second one featured an invited paper presentation, while the third included 3 paper presentations.





Part II

The topics of interest of BigDAP 2016 included but not limited to the following:

- Big Data applications
- Case studies on Big Data management and analytics
- Experiences with Big Data project development and deployment
- Big Data models, algorithms, and architectures
- Cloud computing techniques for Big Data
- Big Data management
- Programming models and environments to support Big Data
- Parallel, distributed computing and virtualization for Big Data
- Big Data frameworks and architectures
- Big Data integrity and privacy
- Big Data search and mining
- Big Data analytics and visualization/Scalable Data Mining and Machine learning techniques
- Research challenges on Big Data management and analytics

The calendar for the event was established by the Programme Committee, as shown in the following table:

Important Dates	
April 15 th , 2016	Deadline for paper submission
May 3 rd , 2016	Deadline for paper submission (EXTENDED)
May 25 th , 2016:	Acceptance Notification
May 31 th , 2016:	Acceptance Notification (EXTENDED)
June 10 th , 2016:	Camera-Ready Submission
June 6 th , 2016:	Early registration
August 25 th , 2016	Regular Registration

Table 2: BIGDAP 2016 Important dates

The program was structured in a single block featuring both academic and industrial contributions and one invited talk. The program was divided into three sessions:

- Keynote speech: "Internet measurements: collect, analyse, extract value The experience from mPlane" was presented by Marco Mellia, mPlane FP7-IP project coordinator.
- Invited Talk: "Reducing Big Data by Means of Context-Aware Tailoring" by Paolo Garza, Elisa Quintarelli, Emanuele Rabosio and Letizia Tanca.
- Accepted papers: 8 papers were submitted, of which only 3 were accepted.

The BigDAP website (http://dbdmg.polito.it/bigdap2016) offered a compilation of the information about the CFP, the accepted papers and the programme of the workshop (Figure 1).







Figure 1: BigDAP16 web page

In the manner of the ONTIC portal, the BigDAP website contains links to the communication channels for diffusion and feedback (linking from/to the project's website and the project's twitter page).

The website can be also accessed through the ONTIC portal through the following URL (Figure 2): <u>http://ict-ontic.eu/</u>





Online Network Search. Trafflc Characterization CONTACTS HOME NEWS DATA ¥ DISSEMINATION V EVENTS ¥ る ト YOU ARE HERE: HOME \\ ABOUT \\ PARTNERS \\ UPM AAA BigDAP 2014 UNIVERSIDAD POLITÉCNICA DE MADRID BigDAP 2015 TWITTER BigDAP 2016 Tweets by @ONTICProject

Figure 2: Access from ONTIC portal to BigDAP16 web

To promote the dissemination of the Workshop, several different channels were used:

- Via email, using international email distribution lists from academic fields related to Computer Science in Europe.
- Via email, sent by the Programme and Organization Committee members to personal contacts in academic and industrial fields.
- Using webpages and other internal or external dissemination means available in each of the consortium partners.
- Via the ONTIC project webpages, through the 'Event' and 'News' sections.
- Via twitter, using the official ONTIC project twitter.

The workshop proceedings have been generated by Springer in both digital and physical formats, and are available at http://www.springer.com/us/book/9783319440651

We estimate that about 30 persons attended the BigDAP Workshop; there is no official data disclosed by the ADBIS organization about the number of attendees at the main conference or at specific workshops.

Conclusively, ONTIC has successfully organized its third consecutive BigDAP Workshop in colocation with ADBIS, 2016. The following summarize the main outcome of the event:

- Presentation and dissemination of the state of the art in different areas of research and development in the field of Big Data, both in the academic and the industrial worlds. The level of participation was satisfactory given that BigDAP is still young.
- Networking: New contacts have been established with people connected to the industrial and academic and/or scientific fields, achieving a useful interexchange of ideas and knowledge.
- Visibility: The ONTIC project work and results were made known to a wider industrial and scientific community.
- Encouragement of undergraduate and postgraduate students, especially PhD students, to get involved in research fields related to Big Data, Big Analytics and Computer Science.
- Produced a proceedings volume published by Springer, both in digital and physical format, with an associated ISBN.





4.3 Public Availability of the ONTS dataset

The ONTIC project had originally planned to make publicly available the ONTS dataset at the end of the project. However, following the project reviewers' recommendation we contacted with the suggested security and privacy expert Dr. Simone Fischer-Hübner in order to double check that all the adequate access constrains according the current national and international legislation would be applied when making publicly available the ONTS dataset. Finally, we announced the public availability of the ONTS dataset on December 2015 via the ONTIC website and twitter. Details of this process can be found in Deliverable 2.5 (Progress on the Provisioning Subsystem) [6]

4.3.1 ONTS dataset dissemination

The following list enumerates companies, researchers and projects that have contacted in order to download and use the ONTS dataset:

- H2020 SHIELD project (<u>https://www.shield-h2020.eu</u>).
- University of Thessaly, Greece, Dept. of Electrical and Computer Engineering, NITLab
- University College London, UK, Dept. of Electronic & Electrical Engineering, Communications and Information Systems Group

4.3.2 SynthONTS ground truth

In the context of the ONTIC project, a ground truth named 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 ONTS dataset.

It contains 14 different types (smurf, fraggle, Syn DDOS, Syn flooding, UDP flood, UDP scan, TCP connect scan, TCP Syn scan, Ping scan, Null scan, IP protocol scan, Scan OS, ChristmasTree scan, brute force attack) of anomalies which can be classified in three types: network recognition anomalies, DOS or DDOS and other type of attacks. For more information, about the ground truth generation see deliverable D4.3 [7] of the ONTIC project.

This ground truth is available 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).

4.4 **ONTIC Datathon**

In the annual review of the second year, reviewers explicitly emphasized the importance of improving the popularity and accessibility of the final fruits of ONTIC project. In order to promote the ONTS dataset, which is one of our most important final contributions, to wide audiences from both research and industrial areas interested in data analysis, we will organize a datathon at the end of the project. This event will hopefully succeed in providing a "big" free choice of publicly accessible network traffic datasets to data analysts who are currently suffering from a lack of such datasets.

A datathon is an event where multiple teams consisting of data scientists, computer engineers and telecom experts compete to provide better solutions to specific task(s) of concern in the telecoms domain based on collected data. The ONTS dataset will be used in the datathon and different State-of-the-Art or novel approaches are expected to be applied.





Careful tuning is however necessary to achieve the best outcome. The team obtaining the best score (according to appropriate evaluation metric) will win the event.

The ONTIC datathon information is available at the following link: http://ictontic.eu/index.php/events/datathon

We will announce the ONTIC datathon by means of email lists and the ONTIC twitted account. In the rest of the section we describe the contents of the datathon web page: First we present a description of the ONTIC datathon, next we describe the tasks to be done and the dataset that will be utilized. Following, we describe the registration submission and evaluation information along with important dates. Finally, we provide the official link to access to this information.

4.4.1 **Description of ONTIC Datathon**

The new generation of communications networks, is expected to cater to the needs of billions of interconnected devices of a heterogeneous nature. Therefore, there is an increasing interest in leveraging recent advances in machine learning and data mining to the analysis and characterization of network traffic. The objective for organizing this challenge is to motivate the application of ML in automated network control and management, since machine learning techniques have shown great success in other related domains. Via this challenge, the ONTIC project aims to promote participants' enthusiasm to build ML-based system for the prediction of network traffic using real network traffic data.

Among the different challenges that network service and infrastructure providers are interested in tackling, demand prediction is one of the most important. The ability to detect patterns in the behaviour of traffic and thus allowing make reliable predictions can be instrumental to efficiently managing virtualized infrastructure, deploying resources when and where they are most needed for optimising service and network performance. The principal task proposed by this challenge is to predict network traffic in short term at different temporal levels, based on real traffic that the project has collected from the network of a Spanish ISP. This huge dataset, the ONTS dataset, will be public available to encourage future advanced research works. A more detailed description is provided in 4.3

4.4.2 Tasks and Dataset

The principal task of the datathon is to predict the number of flows crossing a network link based on a time series of collected values. The dataset to be used has been collected at a medium-sized Spanish ISP and it is a subset of the ONTS dataset collected in the course of the project. It contains a time series per one-second intervals, with each datum representing the number of TCP flows crossing the core network of the ISP during the previous interval. The dataset spans a period of 5 weekdays. Participants will be given the time series corresponding to these five weekdays. . Therefore, the dataset totals 432,000 data points. A dataset from a different period will also be provided for validation purposes.

For the testing phase, a set of sampled time series from different periods will be employed. Participants will be allowed to use up to 1000 data points to predict 1 step, 2 steps, 4 steps and 8 steps ahead in the future. That is, given the time series $t_1, t_2, ..., t_{1000}$, participants will be required to produce predictions for the values at $t_{1001}, t_{1002}, t_{1004}$ and t_{1008} .

The test dataset will consist of 20,000 different periods, provided as a CSV file of 20,000 rows and 1000 columns. Participants are required to produce eight files, each of which must contain 1000+1 values corresponding to the predictions at each of the time steps per input





row. For instance, the first file must contain the 1-second-ahead predictions for each row, the second file will contain the predictions 2 seconds ahead, and so on.

4.4.3 **Registration and Submission**

Interested participants must register and perform following the timeline below. The registration process consists in sending the following documentation:

- A participation form should be properly filled and sent by email. An agreement should be signed by the participant and sent.
- An accreditation document specifying that the participant is a European Union citizen or a valid identification card number for European residents or students: national identity document or national identity document for foreign residents in the European Union.
- Note that this challenge is only addressed to European Union citizens, European Union Organizations (with a European Union Fiscal number) and foreign residents in the European Union with a valid identification card number for the European Union.
- A successful registration, requires that all above documents must be provided in a duly abided way. The organizers only will accept complete and valid registrations.
- The organizers will confirm by email registration to the participants with valid credentials as specified above.

Registered participants are required to submit the following material including the results of their work:

- Forecasting results obtained on the test data, following the same format as the provided validation results, together with the configuration and parameter settings used.
- Source code implementing the proposed algorithm for generating the forecasts. Programming languages are restricted within the scope of Python, Java, Scala, R and Matlab. If a compilable language is used, the submitted source code should be able to generate an executable. Using the executable the organizers should be able to reproduce the results and obtain the same forecasts given the provided configuration settings. Submitted algorithms will be evaluated using the results provided by the organizers in the test dataset.
- If any non-standard distribution library is used in the proposed implementation, the source code of such library should be provided as well.
- In the case that two solutions have sufficiently similar forecasting precisions, the organizers will reward the one considered more novel and interesting.

In order to make sure that all participants will submit their results in a common form, the results should be provided in the same format as the testdata. In addition, the scripts used for measuring the score of the predictions (see next) must also be provided.

4.4.4 Evaluation

The quality of the predictions will be measured with the mean squared error (MSE) of the forecasts with respect to the actual values. The final score of a submission will be the sum of the MSEs over the four different experiments.





4.4.5 Important Dates

Following is the timeline that participants should be aware of:

- March 3, 2017: Deadline for registration.
- March 6, 2017: Communication of acceptance.
- March 9, 2017: The challenge starts. FTP access opens to download training and validation datasets.
- March 16, 2017: Release of test dataset and opening of the results submission page.
- March 23, 2017: Submission of results due
- March 30, 2017: Publication of results

4.4.6 Official Webpage of ONTIC Datathon

http://ict-ontic.eu/index.php/events/datathon

4.5 **Other activities**

4.5.1 **UPM dissemination activities**

In addition to scientific dissemination (section 4.1.1 $\,$), UPM has undertaken the following dissemination activities:

- Teaching:
 - The UPM team is coordinating and teaching a course titled "Massively parallel machine learning" in EIT Digital Master School's Master's Programme in ICT Innovation: Data Science. The course focuses on theory and practice for using machine learning methods at a large scale, specifically on Apache Spark. The Master's programme in ICT Innovation is a two-year international programme run in collaboration with a variety of prestigious European universities, including KTH (Sweden), UNS (France), TUB (Germany), TUE (The Netherlands) and UPM (Spain).
 - http://fi.upm.es/?id=masterdatascience
 - https://www.eitdigital.eu/
 - The UPM team has also taught courses on Spark and Hadoop for UPM's Master's Degree in Software for Distributed and Embedded Systems, as part of a subject titled "Services and protocols for Internet applications".
- Talks:
 - Bruno Ordozgoiti (UPM) gave a talk at the 2016 Workshop for Young Researchers in Applied Mathematics and Engineering, entitled "Discrete optimization for unsupervised dimensionality reduction", discussing some of the feature selection methods developed in ONTIC.

4.5.2 Ericsson dissemination activities

In order to foster the dissemination of the ONTIC results and highlights, Ericsson Spain has contributed them to the Network Analytics Lab, an initiative by the Technology and Innovation unit within Ericsson Spain, in cooperation with Telefónica. A team of fifteen grant-holder students have been involved in hands-on Data Science projects. The lab has been hosted by Madrid R&D from April 1st to Sep 30th. Adaptive Quality of Experience has





been one of the projects agreed with Ericsson RMED OSS&BSS unit. A team of Ericsson engineers and students has adapted the ONTIC AQoE proof-of-concept to a decoupled environment with Apache Kafka as building block. The results were presented to Telefónica in an ad-hoc event and are being currently handled by the customer units within Ericsson Spain. See video (in Spanish) at https://www.youtube.com/watch?v=9SrOBooHjsg

NETWORK ANALYTICS LABS LAUNCHED AT MADRID R&D

On April 1st, we kicked off the Network Analytics Labs, its team and first projects going into execution.

This initiative, conceived together with our colleagues at RMED EPs and ITSS Iberia, aims at accelerating the development and early market validation of innovative Analytics solutions.



With that goal in mind the first set of PoCs to be delivered has been prioritized, their development has already started and dialogs with local customers have been opened to enable the collection of early feedback from the market. This approach and way of working is expected to enable a competitive advantage for our company in this growing business area.

A powerful team combining Ericsson expertise -at R&D, EPs and ITSS- with the Data Science capabilities supplied by URJC Professor Isaac Martín and a team of 10 talented students that he is mentoring is already in place, and executing three projects dealing with analytics application on Security, Automation and Customer Experience.

If you want to know more about this initiative please don't hesitate to contact <u>Manuel Lorenzo</u> (R&D) and <u>David</u> <u>Espadas</u> (RMED EP).

Figure 3: Ericsson Spain Network Analytics Labs launch internal press note

4.5.3 **POLITO dissemination activities**

Daniele Apiletti and Elena Baralis held the third edition of the ONTIC seminar in the undergraduate course of "Database Management Systems" for the Computer Engineering students al Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino, Italy, in November 2016. Recent advances in Big Data technologies related to ONTIC activities were presented to an audience of approximately 200 students enrolled for the course.

Furthermore, prof. Elena Baralis disseminated ONTIC activities into her PhD course at the Dipartimento di Automatica e Informatica, Politecnico di Torino, held from mid-September to mid-October 2016 (10 lessons overall). ONTIC concepts and research activities were spread across different lessons covering clustering techniques, classification models and correlation analysis by means of association rule mining. The audience included approximately 20 PhD students, mostly coming from the Computer and Control Department.

Finally, Paolo Garza holds a new Master-degree course in Big Data technologies and techniques within the Computer Engineering degree. Contents and lessons have been greatly influenced by ONTIC research activities.

Polito plans to iterate on the above activities in the coming years, updating the contents and improving the dissemination based on feedback from the audience.





4.5.4 CNRS dissemination activities

Philippe Owezarski has been involved in the design and set-up of a new joint specialty in the engineering schools and university of the city of Toulouse (started in 2015). This new joint specialty, called TLS-SEC (Toulouse Security) is related to the teaching of all aspects of security aspects in networking and software engineering. It includes a lecture given by Philippe Owezarski on network anomaly detection in high bandwidth networks. It includes the data mining and machine learning algorithms that have been designed in the framework of the ONTIC project, the documentation provided to students including ONTIC related papers. In 2016, the quoted lecture did continue with several updates related to the advances of the ONTIC

CNRS has also established a collaborative work with the H2020 ENDEAVOUR project. ENDEAVOUR aims at designing new network architectures, mechanisms and protocols leveraging the SDN concept for IXPs. ENDEAVOUR, in its use case for IXP members, includes anomalies and intrusion detection and mitigation. The ORUNADA algorithm has then been ported on the ENDEAVOUR architecture. This architecture especially takes advantage of NetFPGA, which ORUNADA has been ported on.

Among the ENDEAVOUR Partners, IBM (Zürich) exhibited a very strong interest with ORUNADA. But they have a much stronger requirement than the one expressed in ONTIC in terms of targeted throughput. For that purpose, a new version of ORUNADA leveraging Apache SPARK has been issued as detailed in D4.3 [4]. This version aims at being much faster than the previous ORUNADA version. The performance evaluation on IBM traffic is still ongoing at the time of writing this deliverable.

4.5.5 SATEC dissemination activities

In order to present the ONTIC project results and possible commercial uses, SATEC has undertaken the following activities and actions:

 Paper "Fast data en la era del Big Data" published by Miguel Angel López in the digital edition of Expansion newspaper (Spanish business newspaper). October 2016.



Figure 4: Paper in a Spanish business newspaper





 Presentation "Fast Data" in Semana del Impulso TIC (Event organized by the Official Associations of Computer Engineers (COIIPA) and Technical Engineers in Informatics (CITIPA) is the reference event for Spanish/Asturian professionals in the ICT sector). November 2016. Niemeyer Center, Avilés (Spain).

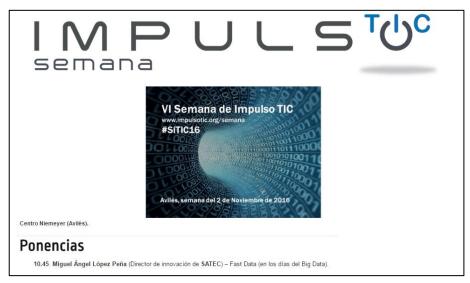


Figure 5: Paper and presentation in Spanish event.

- Presentation in the SATEC's Operation Committee "ONTIC project results". September 2016. SATEC Group. Madrid (Spain).
- Participation in the SATEC Big Data and Analytics internal task force 2016. Moving the ONTIC results (mainly Fast Data architecture) to proposals to SATEC's customers.
- Big Data internal workshop 2016 for Production, Business Development and sales departments of SATEC. "Trends on Big Data and Scalability".
- Some enters in SATEC Blog "Talento y Tecnología" (Talent and Technology). <u>https://satecblog.wordpress.com/</u>.

4.5.6 ADAPTIT dissemination activities

ADAPTIT has participated in the Mobile World Congress, Barcelona, March 2016, and presented the ONTIC project to visitors in its stand.

Continuing its dissemination activities from second year, the vision and work activities of ONTIC and the particular involvement of the company have been presented to a number of customers, including service, network providers and academia. These presentations happened during business meetings for showing the commitment and know-how of the company in the area of Big Data infrastructure solutions and analytics.

4.5.7 EMC dissemination activities

During the year 2016, we have continued with the activities started the previous year to create awareness of the ONTIC project inside EMC, both the sales area as the primary focus and the delivery teams:

 Quarterly sessions organized by Professional Services - EMC Spain, to explain sales force/delivery teams the kind of projects the delivery team is working on. The ONTIC project has been included as one of the Big Data references to tell our customers, to identify possible candidates interested in ONTIC goals and progress.





During the third year we have emphasized on Use Case 3, Quality of Experience Control, explaining in detail the prototype we have been working on.

In addition to the internal quarterly sessions, this year we have leveraged the event DELL EMC Forum, organized in Madrid the 29th of November, and in Barcelona the 1st of December, to show our customers and partners the ONTIC project work and results; EMC (now DELL EMC after the acquisition of EMC by the DELL company) organized an event for customer and partners, called DELL EMC Forum, to present and share with them the latest news of products and solutions of the company. This year, Professional Services had a stand in both events, and we took the opportunity to present the ONTIC project and the Use Case prototype to all partners and customers who came to the stand.



Figure 6: Dell-EMC Forum





5. Exploitation Activities During Third Period

5.1 UPM exploitation activities during the Third period

The main activities made by UPM related to project exploitation are:

- Nokia Bell labs Israel: The collaboration of UPM with Nokia Bell Labs Israel has resulted in the development of a ML-based method to detect noisy neighbours on cloud infrastructure; this work has been resulted to a paper accepted in ESANN 2017 (Core B rated conference). Nokia product unit has expressed interest in transforming this method into a product. In addition, UPM is preparing a contribution to the NFVRG (NFV Research Group) IRTF standardization body based on this work.
- Telefónica: The ongoing collaboration between UPM and Telefónica has resulted to need of setup of a laboratory for generating realistic network traffic traces and reproducible and controlled experiments. Based on these traces we intend to develop methods for classification and anomaly detection, whose effectiveness in real environments will be validated by Telefónica engineers. In addition we held a meeting with ElevenPaths (www.elevenpaths.com) a subsidiary of Telefonica that delivers disruptive innovation in cybersecurity. Their interest was to know about our results in the application of machine learning in traffic classification.
- UPM has recently started to explore a collaboration with Orange Labs France. The goal of this joint work would be to develop forecast methods to aid in the task of predicting SLA violations. In particular, we plan to explore the applicability of reinforcement learning techniques to this scenario.

5.2 Ericsson exploitation activities during the Third period

Ericsson contribution to ONTIC focuses on the Quality of Experience area. The scenario is about the application of ML-algorithms for increasing the degree of automation in the control and management of telecommunication networks.

The exploitation activities carried out by the ONTIC team within Ericsson Spain have focused on three main areas, sometimes not clearly decoupled from the dissemination activities outlined in section 4.5.2 : first, to increase awareness inside the Ericsson group with regard to ML-based analytics for detecting QoE degradation situations and to the applicability of such closed control loops; secondly, to promote the ONTIC findings and proofs of concepts among the Ericsson product organizations, so that the features outlined by ONTIC can be added to new or existing products; and, finally, to showcase the ONTIC results to Ericsson customers in order to start relevant commercial discussions.

With regard to the internal awareness activities, the Technology and Innovation unit within Ericsson Spain presented a QoE Analytics Function, based on suitable ML algorithms specified by Polito and the test dataset produced by DellEMC, in the Ericsson Machine Learning Summit 2016 (Kista, Sweden, October 4th-6th). This event, sponsored by Ericsson Research, is a yearly event where the Ericsson community interested in data analytics gather and discuss new solutions and business scenarios. The demo "Classification of video sessions using Machine learning (TCP/UDP Classification)", drawn from ONTIC work, was showcased. As a result, the proof-of-concept and associated material has been shared with Ericsson commercial organizations in several regions and discussed with several customers.







Figure 7: Ericsson Machine Learning Summit leaflet

The promotion of the ONTIC findings within product organizations has also been pursued focusing around the Ericsson Analytics Implementation Architecture (AIA). The Analytics Implementation Architecture is the part of the Ericsson Architecture that aims to enable reuse of analytics applications and components in order to increase efficiency in the development of products, services and research. Ericsson AIA promotes the usage of integration and extension points while it promotes the usage of a common SDK in order to combine or extend compliant assets. The Policy Governance Function and the QoE Analytics Function implemented in the project -based on ML algorithms by Polito and the DellEMC dataset- is currently being added to the AIA catalogue. As a consequence, any unit within Ericsson can benefit from ONTIC findings and expertise.

Finally, the Technology and Innovation unit within Ericsson has contributed to the services of the Ericsson RMED OSS&BSS unit portfolio. Meetings with three different customers within Ericsson Iberia have been taken place, where the AQoE solution developed by ONTIC was presented. Talks for launching an innovation project around this worked out concept are currently being carried out.

As an additional point, Ericsson Spain has committed to open source all its deliverables and therefore, all the source code is offered under the Apache License 2.0 (https://gitlab.com/users/ontic-ericsson/contributed). Docker container images are available in Docker Hub (https://hub.docker.com/r/onticericssonspain/). Customization is also possible as dockerfiles are also available.





5.3 **POLITO exploitation activities during the Third period**

POLITO exploitation plans have been focused on technology and competence transfer to companies, by using ONTIC outcomes to increase awareness among industry and professionals. Thanks to the experience on Big Data analytics, enhanced through participation in ONTIC, Polito established new partnerships with many companies, not only for network-related data analysis ones, but also in different domains; all facing the challenge to extract data and create machine-learning models from heterogeneous sources to distil intelligence and actionable metrics. Collaborating companies include leaders in the automotive industry, such as FCA and General Motors, with whom we have already started research contracts, and companies in the Smart City area, specifically on the "green" energy challenge, such IREN, and ENEL.

Polito has exploited ONTIC results as a base showcase for research on novel Big Data algorithms and data mining techniques in subsequent H2020 proposals and national projects (PRIN) while it has let students exploit ONTIC activities for completing their Degree Final Project.

Elena Baralis, WP3 leader, has joined the scientific and steering committee of the Master degree in Big Data Engineering by Politecnico di Torino, where most participants come from international companies. The course content has taken advantage of the Big Data experience gained from ONTIC activities, and allowed the Polito team to establish relationships with companies interested in this area.

Finally, the new Big Data Lab at Polito, whose steering committee includes Elena Baralis, and which is hosting some ONTIC-funded analytics servers, allowed us to attract the interest of many companies which are interested in exploiting our foreground from ONTIC (algorithms and data-driven models) into their domains. Promising collaborations has been seeded with companies such as Fastweb (ISP provider), Juventus FC (football club), and a number of local SMEs.

5.4 CNRS exploitation activities during the Third period

CNRS, as an academic organization, has to bridge the gap between its theoretical and applied research results with educational organizations (universities, engineering schools, etc.) on one side, and companies on the other side.

Dealing with educational schools, CNRS continues contributing to an educational engineering diploma shared between the main engineering schools in Toulouse (INSA, INP, ENAC, and ISAE) and the University of Toulouse. This new diploma aims at educating students about security aspects in the digital world. Philippe Owezarski gives a course about the analysis of network traffic for detecting attacks and anomalies, malicious behaviours, bad configurations, etc. He especially disseminates the results gain in ONTIC on WP4 and Use Case 1 about online traffic analysis and anomaly detection. This new diploma (started in September 2015) is called TLS-SEC (for TouLouSe - SECurity).

Dealing with companies, experience and results gained from ONTIC WP4 and Use Case 1 interested the French company BORDER6. BORDER6 aims at developing network access devices capable of analysing inbound and outbound traffic in real-time in order to optimize routing paths and QoS enforcement. BORDER6 was strongly interested in our iterative version of the UNADA algorithm (IUNADA), proposed by WP4 during the second project period. They expect significant performance improvements with this algorithm, as well as the integration of security capabilities like the ones provided by IUNADA for detecting anomalies and attacks from traffic traces. BORDER6 and CNRS have signed a research contract in May 2015 for a three year period that aims at adapting IUNADA to BORDER6 devices, adding security





functionality, and transferring afterwards the technical skills of CNRS to BORDER6. This, currently alive, contract sponsors a PhD student working on transferring CNRS experience gained in ONTIC to BORDER6.

CNRS has also started a project, funding a PhD student with Airbus. Airbus is working on protecting the information and communication of the extended Airbus Company i.e. airbus and all its suppliers, clients, etc. They were attracted by the ONTIC-elaborated unsupervised learning-based approach for enforcing security. The aim of this project is to leverage the results of ONTIC in WP4 and WP5/UC#1 for the purpose of securing the extended Airbus communications infrastructure.

CNRS has also established relationships with the H2020 ENDEAVOUR European project. ENDEAVOUR aims at designing new network architectures, mechanisms and protocols leveraging the SDN concept for IXPs. The unsupervised anomaly detection algorithms are of interest to ENDEAVOUR especially in its use cases on applying network security enforcement for IXP customers. The ORUNADA algorithm has then be ported on the ENDEAVOUR architecture. This architecture especially takes advantage of NetFPGA, on which ORUNADA has been ported.

Among the ENDEAVOUR Partners, IBM (Zürich) exhibited a very strong interest with ORUNADA. But, they have a stronger requirement than the one considered in ONTIC, in terms of targeted throughput. For that purpose, a new version of ORUNADA leveraging Apache SPARK has been issued as detailed in D4.3 [4]. This version aims at being much faster than the previous ORUNADA version. The performance evaluation on IBM traffic is underway at the time of writing this deliverable.

Last, CNRS is also in the process of exploiting the labelled anomalous traces that it is producing in the project. These traces are synthetic traces thus ensuring the validity/truth of the labels. This work makes the network traces provided by the project an ideal support for assessing the detection accuracy of any intrusion detection system.

CNRS also takes advantage of its strong relationships with MAWIIab for getting a first set of labels on a week of ONTS traces.

5.5 SATEC exploitation activities during the Third period

The main focus of SATEC in the third project period has focused on the development of the final version of the analytic dashboard defined as part of use case #1. This has resulted in a new design and development (integration of several modules, programming libraries and tools) of a very efficient and scalable architecture based in the concept of "Fast Data". Under this new architecture the dashboard prototype (web application) has been implemented as the final proof of concept/prototype.

This work has provided SATEC not only the opportunity to acquire a valuable expertise in Big Data and scalability but also the opportunity to develop an own solution (Fast Data multipurpose architecture) ready to be offered in many different customer proposals.

The specific exploitation activities developed by SATEC during the third period are listed in the following:

- Analytic Dashboard demonstrations: the dash board prototype has been shown to SATEC's customers, including:
 - Gas Natural Fenosa (utilities).
 - Endesa (utilities).
 - Enagas (utilities).





- Repsol (utilities)
- IBM Spain (services and manufacturing).
- Junta de Comunidades de Castilla y León (government).
- Red Eléctrica Española (utilities).
- Proposals to customers: the results obtained in ONTIC are being moved to specific commercial solutions and incorporated into customer proposals:
 - \circ SIEM (Security Information and Event Management) for EJIE society (Basque government).
 - Smart Grid analytics for Endesa Electricity (Málaga, Spain)
 - \circ SLAs Management system for gas road transport external services.
 - Smart City consultancy for Junta de Comunidades de Castilla y León (Spanish regional government).
 - SIEM tool for R-Cable (Galicia, Spain)
- Results of the dashboard architecture (developed in the course of use case #1) are being incorporated in technical solutions offered by SATEC. For example:
 - WIM: fast data component to enrich Netflow data.
 - Angola Telecom: actor based Sw module for CDR processing.

5.6 Adaptit exploitation activities during the Third period

ADAPTIT is interested in the ONTIC project results from the view-point of a solution provider focusing on infrastructure and performance management. In particular, ADAPTIT is interested in the design and deployment of the project Big Data architecture for traffic characterization, its provisioning through cloud services and the development and visualization of analytics for distilling network and application performance insights and deriving actionable metrics. ADAPTIT is committed to capitalizing on the experience gained from its involvement in the project and on the collaboration with other partners of the consortium.

As set from the first year, the exploitation of project results spans two directions:

Strengthen company's position in the market: The participation in ONTIC has deepened and expanded the company's know-how in Big Data deployment, analytics development and visualization and thus has provided a strong competitive advantage.

Enhance product/service portfolio with Big Data solutions and analytics: The involvement of the company in the areas of cloud-based Big Data set-up, analytics for stochastic characterization and generation of realistic network traffic has provided grounds for developing innovative service solutions for performance management. Big Data-as-a-service is opted as a promising sales alternative. The strategy is to pursue FOA (First Office Application) for exhibiting the benefits of the proposed services, with ADAPTIT undertaking the cost of the pilot project.

To the above direction, ADAPTIT has progressed during this year the business case with First Data Hellas (www.firstdata.com/en_gr/home.html/) on Big Data analytics for cost, revenue analysis and forecasting based on usage measurements. Following the developments of the previous year (pilot demonstration, [3]), positive feedback has been received from executive management and the case has reached at a contract stage.





In addition, ADAPTIT has developed during this year a new business case revolved around the RMMS (Remote Monitoring and Management system) product¹. RMMS offers reliable, scalable and managed services for monitoring operational, security and environmental conditions in remote sites. The analysis of measured traces as they continuously come from a big number of diverse sources -sensors per site- becomes of paramount importance and adds significant business value to the product. For example, through suitable on/off-line analytics, anomalies could be spotted indicating malfunctions or environmental changes while seasonal trends and correlations between different remote sites could be asserted. Presently, this case is at prototype development stage. The target is not only to address current and future RMMS customer needs but to enter multi-source, data-demanding market niches such as smart-agriculture and smart-transportation.

Evidently, the cases above are directly aligned with the activities of the company in ONTIC. The analytics for stochastic characterization and generation of network traffic can be used for generating hypothetical usage time-series required for what-if scenarios, also for establishing evolution patterns or spotting out anomalies required for trend analysis.

Conclusively, ONTIC has acted catalytically into moving ADAPTIT to Big Data infrastructure management and analytics, which clearly constitute directions of increased business potential.

5.7 EMC exploitation activities during the Third period

During the third project year we have continued working closer with EMC sales forces to identify possible candidates interested in ONTIC work and results. For showing tangible results to our customer, we have worked hard in the prototype for the Use Case 3 (Video User Experience).

During the celebration of the DELL EMC Forum in Madrid (November 29th) and Barcelona (December 1st) we invited customers in our stand in Professional Services to demonstrate the ONTIC activities and results, in particular the Use case 3 prototype. High interest was noted, and so we are planning to take this opportunity further. Also, leveraging ONTIC experience in the Big Data initiatives of our customers has and is currently being pursued.

¹ RMMS is provided by ADAPTIT's affiliate TELCOSERV: http://www.telcoserv.gr/en/products-telcoserv/remote--monitoring-and-management-system-rmms





6. References

- [1] ONTIC. "Deliverable D6.3. Exploitation Strategy" Internet: <u>http://www.ict-ontic.eu/</u>, May. 2014 [May. 28, 2014]
- [2] ONTIC. "Deliverable D6.4. "Exploitation and Dissemination Plans " Internet: <u>http://www.ict-ontic.eu/</u>, Feb. 2015 [Feb. 1, 2015]
- [3] ONTIC. "Deliverable D6.5. "Progress on Exploitation and Dissemination Plans Part I." Internet: <u>http://www.ict-ontic.eu/</u>, Jan. 2016 [Jan 31, 2016]
- [4] ONTIC. "Deliverable D4.3. Experimental evaluation of algorithms for online network characterization", Internet: <u>http://www.ict-ontic.eu/</u>, Dec. 2016.
- [5] ONTIC. "Deliverable D5.10. ONTIC Framework", Internet: <u>http://www.ict-ontic.eu/</u>, Feb. 2017.
- [6] ONTIC. "Deliverable D2.5. Progress on Provisioning Subsystem." Internet: <u>http://www.ict-ontic.eu/</u>, Feb. 2015.
- [7] ONTIC. "Deliverable D4.3. Experimental Evaluation of Algorithms for Online Network Characterization". Internet: http://www.ict-ontic.eu/, Feb 2017