



HAL
open science

ONTIC D1.2: Planning for the Next Period #2

Bruno ; Ordozgoiti, Alberto Mozo, Bo Zhu, Sandra Gomez, Stanislav Vakaruv,
Enrique Sánchez, Philippe Owezarski, Juliette Dromard, Alejandro
Bascañana, Miguel Angel Monjas, et al.

► **To cite this version:**

Bruno ; Ordozgoiti, Alberto Mozo, Bo Zhu, Sandra Gomez, Stanislav Vakaruv, et al.. ONTIC D1.2: Planning for the Next Period #2. Universidad politécnica de Madrid; CNRS-LAAS; Ericsson Spain; SATEC; Politecnico di Torino; ADAPTIT. 2016. hal-01965709

HAL Id: hal-01965709

<https://laas.hal.science/hal-01965709>

Submitted on 26 Dec 2018

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

Deliverable Planning for the Next Period #2

ONTIC Project
(GA number 619633)

Deliverable D1.2
Dissemination Level: RESTRICTED

Authors

Bruno Ordozgoiti, Alberto Mozo, Bo Zhu, Sandra Gómez, Stanislav
Vakaruv, Enrique Sanchez
UPM

Philippe Owezarski, Juliette Dromard
CNRS

Alejandro Bascuñana, Miguel Ángel Monjas
Ericsson

Miguel Angel López
SATEC

Daniele Apiletti, Elena Baralis
POLITO

Stefanos Anastasiou
ADAPTIT

Version

ONTIC_D1.2.2016.01.29.1.0

Version History

| Previous version | Modification date | Modified by | Summary |
|------------------|-------------------|---|---|
| 0.1 | 2015.12.10 | UPM | 1st draft version |
| 0.2 - 0.7 | 2016.01.15 | Ericsson, POLITO, CNRS, SATEC, EMC, Adaptit | Added first contributions by other partners |
| 0.8 | 2016.01.20 | UPM | First version ready for review |
| 0.91 | 2016.01.25 | EMC | Review |
| 0.92 | 2016.01.29 | POLITO | Review |
| 1.0 | 2016.01.29 | UPM | Quality assurance |

Quality Assurance:

| Name | |
|----------------------------------|----------------------------------|
| Quality Assurance Manager | Alberto Mozo (UPM) |
| Reviewer #1 | Fernando Arias (EMC) |
| Reviewer #2 | Daniele Apiletti (POLITO) |



Table of Contents

Contents

| | |
|---|----|
| 1. PURPOSE OF THE DOCUMENT | 5 |
| 2. SCOPE | 6 |
| 3. INTENDED AUDIENCE | 7 |
| 4. SUGGESTED PREVIOUS READINGS | 8 |
| 5. EXECUTIVE SUMMARY | 9 |
| 6. PROJECT STATUS | 10 |
| 6.1 Work Package WP1 Management | 10 |
| 6.1.1 Task T1.1 Administrative Project Management | 10 |
| 6.1.2 Task T1.2 Content and Quality Management | 11 |
| 6.2 Work Package WP2 Big Data Network Traffic Summary Dataset | 11 |
| 6.2.1 Task T2.1 Requirements and Functionalities of a Big data Architecture for Network Traffic Summary dataset | 11 |
| 6.2.2 Task T2.2 Design of Traffic Summary Dataset Provisioning Subsystem | 12 |
| 6.2.3 Task T2.3 Deployment, Configuration and Provisioning of Big Data Network Traffic Summary Dataset | 13 |
| 6.3 Work Package WP3 Scalable Offline Network Traffic Characterization System | 13 |
| 6.3.1 Task T3.1 Scalable algorithms for unsupervised learning | 13 |
| 6.3.2 Task T3.2 Scalable algorithms for supervised classification | 14 |
| 6.3.3 Task T3.3 Scalable algorithms for correlation analysis | 15 |
| 6.3.4 Task T3.4 Scalability evaluation on enterprise-grade cloud environment | 15 |
| 6.4 Work Package WP4 Scalable Online Network Traffic Characterization System | 16 |
| 6.4.1 Task T4.1 Architecture Design of online classification systems for ISP traffic flows | 16 |
| 6.4.2 Task T4.2 Traffic Pattern evolution Subsystem | 17 |
| 6.4.3 Task T4.3 Network Intrusion Detection Subsystem | 17 |
| 6.4.4 T4.4 Subsystems Integration and Testing | 18 |
| 6.5 Work Package WP5 Implementation, Integration and Verification | 19 |
| 6.5.1 Task T5.1 Integration of Offline and Online Network Traffic Characterization Systems | 19 |
| 6.5.2 Task T5.2 Verification and Field Trials | 19 |
| 6.5.3 Task T5.3 Standardization and Regulatory Activities | 20 |
| 6.6 Work Package WP6 Exploitation, Dissemination and Standardization | 21 |
| 6.6.1 Task T6.1 Dissemination, Training and Public Awareness | 21 |
| 6.6.2 Task T6.2 Exploitation and IPR management | 22 |
| 6.6.3 Task T6.3 Impact Management | 23 |

| | |
|--|-----------|
| 7. KEY ACTIVITIES DECLARED BY EACH PARTNER | 25 |
| <hr/> | |
| 7.1 Ericsson..... | 25 |
| 7.1.1 Activities Completed | 25 |
| 7.1.2 Activities in Progress | 26 |
| 7.1.3 Plans and Actions for the Next Reporting Period..... | 26 |
| 7.2 UPM | 27 |
| 7.2.1 Activities Completed | 27 |
| 7.2.2 Activities in progress | 28 |
| 7.2.3 Plans and Actions for the Next Reporting period..... | 28 |
| 7.3 Polito | 29 |
| 7.3.1 Activities Completed | 29 |
| 7.3.2 Activities in Progress | 30 |
| 7.3.3 Plans and Actions for the Next Reporting period..... | 31 |
| 7.4 CNRS | 31 |
| 7.4.1 Activities Completed | 31 |
| 7.4.2 Activities in progress | 32 |
| 7.4.3 Plans and Actions for the Next Reporting Period..... | 32 |
| 7.5 SATEC | 32 |
| 7.5.1 Activities Completed | 32 |
| 7.5.2 Activities in progress | 33 |
| 7.5.3 Plans and Actions for the Next Reporting Period..... | 34 |
| 7.6 EMC Spain | 34 |
| 7.6.1 Activities Completed | 34 |
| 7.6.2 Activities in Progress | 35 |
| 7.6.3 Plans and Actions for the Next Reporting Period..... | 35 |
| 7.7 Adaptit | 35 |
| 7.7.1 Activities Completed | 35 |
| 7.7.2 Activities in Progress | 36 |
| 7.7.3 Plans and Actions for the Next Reporting Period..... | 37 |
| 8. COMPLIANCE WITH PLANNING MILESTONES | 38 |
| <hr/> | |
| 8.1 WP2 | 38 |
| 8.2 WP3 | 38 |
| 8.3 WP4 | 38 |
| 8.4 WP5 | 38 |
| 9. CURRENT AND POTENTIAL ISSUES/PROBLEMS/RISKS | 40 |
| <hr/> | |
| 9.1 WP2 | 40 |
| 9.2 WP3 | 40 |
| 9.3 WP4 | 40 |
| 9.4 WP5 | 40 |
| 9.5 WP6 | 40 |



1. Purpose of the Document

The purpose of this deliverable is to provide an overview of the activities carried out so far in order to assess the overall progress of the project. Based on the observed progress and evolution, the planning for the next period might be subject to slight adjustments in order to ensure that the project maintains its high potential impact.

As stated in the DoW, this document will include, specifically:

- Project status
- Compliance with planning milestones
- Current and potential project issues, problems and risks
- Plans and actions for the next reporting period



2. Scope

This document contains a high-level overview of the activities carried out in the second year of ONTIC project . It is not intended to provide a detailed description of how these activities have been accomplished. Such a description can be found in the corresponding deliverables.



3. Intended Audience

The intended audience includes every partner of the ONTIC project.



4. Suggested Previous Readings

The ONTIC DoW should provide enough background to understand the goals of the project and the activities through which these objectives are meant to be attained.



5. Executive Summary

This deliverable provides a detailed description of the progress made so far by each partner during the second year. A comprehensive enumeration of both completed and ongoing tasks is presented. In addition, this document presents the issues encountered by each partner as well as those that could surface during the second year.

Section 6 presents the project status at the task level. The situation of the work plan activities and the expected outcomes of each task described in the DoW are reported.

Section 7 contains a detailed description of the activities that each partner has already completed, those that are currently being carried out and those that are planned for the next reporting period.

Section 8 examines the level of compliance of the project with the second milestone, set for M24. The results obtained so far on each work package are analysed in terms of their conformity with this milestone.

Section 9 describes issues found so far, as well as the solutions proposed. Additionally, this section presents potential issues identified by the partners of the project.

The contents of this document will serve as a basis to refine the planning for the subsequent periods.



6. Project Status

In this section, we show the ONTIC project status by the month M24 classified by work-package and task. The actual status of each activity and expected outcomes is presented. Activities in progress are accompanied by an estimated completion percentage as perceived by participating partners. Note that these figures are meant to be only illustrative.

6.1 Work Package WP1 Management

6.1.1 Task T1.1 Administrative Project Management

| Work Plan | |
|---|--|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Cover administrative aspects of the project management. | IP (66%) |
| Preparation of a Consortium Agreement: The Consortium Agreement will be drawn up and signed by partners to address intellectual property, exploitation rights, regulations and safety before the project start. | C |
| Administrate the budget in order to match the plans approved by the General Assembly. | IP (66%) |
| One Kick-off meeting at the beginning of the project. | C |
| General Assembly Meetings - every six months. | IP (66%) |
| Biweekly Project Audio Conferences | IP (66%) |

| Expected Outcomes | | |
|---|--|---------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverables |
| Ensured efficiency and optimized outcomes from the involved partners | IP | D1.1, D1.2 |
| Efficient project administrative and financial monitoring | IP | D1.1, D1.2 |
| Documents connected with the project from the consortium to the EC, including reporting documents, cost statements etc. | IP | D1.1, D1.2 |



6.1.2 Task T1.2 Content and Quality Management

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Draft the project quality plan | C |
| Supervise the development and exchange of working documents as well as control the quality of the material. | IP (66%) |
| Supervise the dissemination of the results in conferences, journals and workshops. | IP (66%) |
| Re-evaluate the status and schedule of the project periodically. | IP (66%) |
| Risks identified will be reported to the coordinator for management and mitigation. | IP (66%) |
| Quality checks will be performed upon approval of deliverables. | IP (66%) |

| Expected Outcomes | | |
|---------------------------------|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Work Efficiency | IP | D1.3 |
| Visibility of the Project | IP | D1.3 |
| Scientific Value of the Project | IP | D1.3 |
| Reporting documents | IP | D1.3 |

6.2 Work Package WP2 Big Data Network Traffic Summary Dataset

6.2.1 Task T2.1 Requirements and Functionalities of a Big data Architecture for Network Traffic Summary dataset

| Work Plan | |
|--|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Definition of Big Data solution capabilities for ONTIC analytics and ONTS dataset. | C |
| Mapping between supported ONTIC analytics and ONTS dataset to data requirements. | C |
| Design of ONTIC Big Data Architecture. | C |



| Expected Outcomes | | |
|--|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| ONTIC Big Data Architecture requirements | C | D2.1, D2.2, D2.4 |

6.2.2 Task T2.2 Design of Traffic Summary Dataset Provisioning Subsystem

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Survey of capabilities of different measurement technologies including NetFlow, port replication and optical splitting among others. | C |
| Topology, routing, and traffic analysis of network to identify candidate locations for measurement instrumentation. | C |
| Techno-economic study for each candidate location to decide the appropriate measurement technology. | C |
| Design and implementation of a software component to generate summary records by extracting statistical information from network packets. | C |
| Design and implementation of a software component to move and store summary records from sampling points to the Big Data platform. | C |
| Integration and final testing of Provisioning Subsystem. | C |

| Expected Outcomes | | |
|--|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Design of the Architecture of a Provisioning System for Traffic Summary dataset. | C | D2.4, D2.5 |
| Implementation and Testing of Provisioning System for Traffic Summary dataset. | C | D2.4, D2.5 |



6.2.3 Task T2.3 Deployment, Configuration and Provisioning of Big Data Network Traffic Summary Dataset

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Deployment and configuration of local instances of a Big Data platform based on T2.1 Big Data Architecture. | IP (90%) |
| Deployment and configuration of local instances of a Provisioning Subsystem based on T2.2. | C |
| Test and validation of joint operations of Provisioning Subsystem and Big Data platform local instances. | IP (25%) |
| Implementation and validation of security mechanisms for making the ONTS dataset publicly available. | IP (90%) |
| Provisioning of the ONTS dataset by deploying a Big Data platform and a Provisioning Subsystem in SATEC core network. | IP (66%) |
| Progressive migration of subsets of ONTS dataset records from the Provisioning Subsystem at SATEC to enterprise-grade clouds (Google cloud) as required by experimentation. | IP (25%) |

| Expected Outcomes | | |
|---|---|---|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Big Data platform and Provision Subsystem deployed, integrated and fully operative in SATEC core network. | IP | D2.1, D2.2, D2.4, D2.3, D2.5, D2.6 (pending) |
| ONTS dataset totally provisioned with trillions of network traffic summary records obtained from the core network of SATEC. These records will be initially stored in SATEC Big Data platform | IP | D2.1, D2.2, D2.4, D2.3, D2.5, D2.6 and D2.7 (pending) |

6.3 Work Package WP3 Scalable Offline Network Traffic Characterization System

6.3.1 Task T3.1 Scalable algorithms for unsupervised learning

| Work Plan | |
|-----------------------------|---|
| Activity Description | Status (Completed/In Progress/Not Started) |



| | |
|--|-----------------|
| | Started) |
| Unsupervised learning algorithms identification and analysis. | IP (95%) |
| Definition of improvements and novel unsupervised learning techniques. | IP (70%) |
| Design and development of unsupervised learning algorithms. | IP (70%) |
| Unsupervised learning algorithms assessment and experimental evaluation. | IP (20%) |

| Expected Outcomes | | |
|---|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Definition of the limitations of existing algorithms and identification of novel techniques | C | D3.1 |
| Massively scalable clustering algorithms for offline network traffic characterizations | C | D3.2 |
| Experimental evaluation of the algorithms | IP | D3.3 |

6.3.2 Task T3.2 Scalable algorithms for supervised classification

| Work Plan | |
|--|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Available classification algorithms identification and analysis. | IP (95%) |
| Definition of improvements and novel classification techniques. | IP (70%) |
| Design and development supervised classification algorithms. | IP (70%) |
| Classification algorithm assessment and experimental evaluation. | IP (20%) |

| Expected Outcomes | | |
|---|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Definition of the limitations of existing algorithms and identification of novel techniques | C | D3.1 |
| Massively scalable classification algorithms for offline network traffic characterizations | C | D3.2 |
| Experimental evaluation of the algorithms | IP | D3.3 |
| Define standard assessment metrics to evaluate algorithm performance. | IP | D3.3 |

6.3.3 Task T3.3 Scalable algorithms for correlation analysis

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Available correlation-discovery algorithms identification and analysis. | IP (95%) |
| Definition of improvements and novel correlation analysis techniques. | IP (70%) |
| Design and development of correlation analysis algorithms. | IP (70%) |
| Correlation analysis algorithms assessment and experimental evaluation. | IP (20%) |

| Expected Outcomes | | |
|---|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Definition of the limitations of existing algorithms and identification of novel techniques | C | D3.1 |
| Massively scalable algorithms for offline network traffic correlation analysis | C | D3.2 |
| Experimental evaluation of the algorithms | IP | D3.3 |

6.3.4 Task T3.4 Scalability evaluation on enterprise-grade cloud environment

| Work Plan | |
|--|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Scalability evaluation definition in terms of performance metrics. | NS |
| Selection of the cloud provider. | NS |
| Dataset loading to the cloud storage of the selected provider. | NS |
| Distributed analysis execution. | NS |
| Experimental evaluation. | NS |



| Expected Outcomes | | |
|---|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Definition of the performance metrics to evaluate the scalability of the algorithms | IP | D3.3 (pending) |
| Experimental scalability evaluation results | NS | D3.3 (pending) |

6.4 Work Package WP4 Scalable Online Network Traffic Characterization System

6.4.1 Task T4.1 Architecture Design of online classification systems for ISP traffic flows

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Analyse data structure as provided by the network traffic dataset for designing the intelligent scheduler. | IP (90%) |
| Analyse the parallelization capabilities of algorithms designed in tasks 3.1, 3.2, 3.3, 4.2 and 4.3, and design the related parallelization/synchronization mechanisms. | IP (60%) |
| Definition of the assessment metrics of the system performance. | IP (90%) |

| Expected Outcomes | | |
|---|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| The architecture for running in an integrated and parallel way all the functions for the online classification system with real-time constraints (in D4.1). | C | D4.1 |
| Experimental parameters for evaluating the entire online classification system (in D4.2) | IP (60%) | D4.2 |



6.4.2 Task T4.2 Traffic Pattern evolution Subsystem

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Set up the system connected to the flowing data. | IP (90%) |
| Define the algorithmic processing that has to be applied on the data. | IP (50%) |
| Check if the system is able to identify the patterns (clusters). | IP (50%) |
| Tune the system if required. | IP (50%) |
| Add additional machines to the cluster and check the scalability. | NS |

| Expected Outcomes | | |
|--|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Scalable architecture able to process the traffic data as it is flowing (in D4.1) | C | D4.1 |
| Visualization tool | IP (25%) | D4.3 (pending) |
| Parameter optional (Number of clusters, radius, min density threshold, window time, etc. | IP (50%) | D4.2 |

6.4.3 Task T4.3 Network Intrusion Detection Subsystem

| Work Plan | |
|--|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Available systems identification and analysis. | C |
| Design of a robust clustering algorithm for anomalies and intrusions detection coping with the curse of high dimensionality. | IP (60%) |
| Design of an autonomous method for anomalies abnormality ranking. | IP (40%) |
| Design and development. | IP (50%) |

| Expected Outcomes | | |
|--|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Definition of the limitations of existing anomalies and intrusion detection systems and issuing novel techniques. | IP (80%) | D4.2 and D4.3 |
| Massively scalable, efficient, and dimensionality insensitive clustering algorithm for anomalies and intrusion detection. | IP (60%) | D4.2 and D4.3 |
| Autonomous ranking for anomalies abnormality, and automatic defense filtering rules generation for security device online configuration. | IP (40%) | D4.2 and D4.3 |
| Experimental evaluation of the subsystem on scalability, robustness, accuracy and temporal efficiency. | IP (30%) | D4.3 |

6.4.4 T4.4 Subsystems Integration and Testing

| Work Plan | |
|--|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Integrate the algorithms developed in tasks 3.1, 3.2, 3.3, 4.2 and 4.3, and the architecture issued in task 4.1. | IP (10%) |
| Develop the visualization subsystem. | IP (10%) |
| Test the full system and evaluate its global performance level. | NS |
| Issue some demo on some specific use cases. | IP (30%) |

| Expected Outcomes | | |
|--|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| An integrated system for real evaluation | IP (20%) | D4.3 |
| Demos on some specific use cases | IP (30%) | D4.3 |



6.5 Work Package WP5 Implementation, Integration and Verification

6.5.1 Task T5.1 Integration of Offline and Online Network Traffic Characterization Systems

| Work Plan | |
|--|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Integrate ONTIC online and offline mechanisms. | IP (20%) |
| Parameterize and characterize mechanisms, techniques and algorithms to be used by data mining developers in other application domains. | Not Addressed (reviewers recommendation) |
| Set up the ONTIC framework for developing autonomous, proactive and scalable characterization systems. | IP (25%) |

| Expected Outcomes | | |
|--|---|---------------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| An integrated online and offline network traffic characterization system | IP (20%) | D5.3 (pending) |
| An open source framework for developing highly scalable autonomous characterization systems. | IP (20%) | D5.10 (pending) |

6.5.2 Task T5.2 Verification and Field Trials

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Define scenarios and roadmap of global verification and field trials. | IP (90%) |
| Select final scenarios for implementation in experimental driven approach. | IP (66%) |
| Use cases definition. Prototypes design, implementation and set-up. Global verification and fields trials to demonstrate the operation of online and offline ONTIC mechanisms and techniques. | IP (66%) |
| Evaluate the concepts and identify the most significant problems to indicate further research. | IP (75%) |



| Expected Outcomes | | |
|---|---|---|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Integration of key project results in the test-bed environment. | IP (25%) | D5.1, D5.2, D5.3, D5.4, D5.5, D5.6, D5.10 (pending) |
| A set of scenarios for the validation of key ONTIC concepts. | IP (90%) | D5.1, D5.2, D5.4, D5.5, D5.6 (pending) |
| Evaluation of the concepts and identification of topics for further research. | IP (66%) | D5.1, D5.2, D5.3, D5.4, D5.5, D5.6 (pending) |

6.5.3 Task T5.3 Standardization and Regulatory Activities

| Work Plan | |
|--|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Definition of ONTIC standardization strategy. Analysis of existing standardization Work Items that is relevant to the context of the project solution design and aim of the project. | IP (100%) |
| Identify which standardization bodies are most relevant to ONTIC objectives. Define the field of contribution for each standard. | IP (100%) |
| Actively participate in the standardization work in the relevant standardization bodies | IP (66%) |
| Using the outcomes of the research and experimentation of ONTIC for shaping standardization efforts. | IP (25%) |
| Analysis of the International and European regulatory framework. | IP (90%) |
| Identification and analysis of legal problems linked to the implementation of the system. | IP (66%) |
| Evaluation of regulatory impacts of the project architecture and solution. | IP (50%) |
| Build and disseminate recommendations for future regulation activities. | NS |

| Expected Outcomes | | |
|---|---|---|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Document describing ONTIC standardization strategy. | IP | D5.7 D5.3, D5.8, D5.9, D5.10 (pending) |
| Detailed report for summarizing ONTIC standardization effort. | IP | D5.7 D5.8, D5.9 (pending) |
| Document describing ONTIC regulatory | IP | D5.7 |



| | | |
|---|--|----------------------|
| aspects with recommendations to relevant regulatory institutions. | | D5.8, D5.9 (pending) |
|---|--|----------------------|

6.6 Work Package WP6 Exploitation, Dissemination and Standardization

6.6.1 Task T6.1 Dissemination, Training and Public Awareness

| Work Plan | |
|---|---|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Create the project branding (templates, colours and fonts to be used), produce a project logo of professional quality and use it consistently on its web site and its communications. | C |
| Prepare the initial set of communication materials (flyers, posters, pop-up stand and banner). | C |
| Create the project website, under the .eu domain. | C |
| Prepare and maintain a dissemination plan including conferences, publication possibilities and clustering activities. | C |
| Install publication repository including tutorial and training material. | IP (75%) |
| Coordinate and support the joined publication of project partners and external user group participants. | IP (50%) |
| Presentation of project results at booths at conferences including proofs of concept of the context aware service platforms developed in the project. | IP (66%) |
| The ONTIC consortium commits to issuing press releases within one month of the start of the project, as well as to issuing at least two press releases per year on all important milestones and events. The project should identify which specific partners will undertake this action. Such press releases will target the local or national press of the partner entrusted with this task. The press release will describe the goals of the project in simple, jargon free language and whenever possible highlight the benefits to the region/country and the importance of the local partner being part of a EU consortium. A PDF scan of the collection of published press releases will be included as annex to deliverables D6.4, D6.5 and D6.6. | IP (66%) |
| EC may set in motion Support Actions addressing e.g. best practices and standards work, technology transfer and cross-disciplinary training, research roadmaps and technology foresight, in areas especially relevant for the objective of this project. Therefore, ONTIC partners shall undertake - when invited - to contribute to and participate in focused concertation actions, themed seminars or special interest groups, for example the European Data Forum. | IP (66%) |
| All presentations, contributions and publications (also online publications and web sites) even partially funded by the project shall include the project logo and prominently acknowledge the grant under which the project is funded, using the standard text in Article II.30.4 of the Grant Agreement. | IP (66%) |

| | |
|---|----------|
| Details of all publications even partially funded by the project shall be uploaded to some specific and agreed Bibliographic social networks such as http://www.citeulike.org/ , http://www.mendeley.com or http://www.bibsonomy.org/ . Every such publication must be tagged with the project identification “ONTIC 619633”. | IP (66%) |
| Whenever such Bibliographic social networks allow for a catalogue of publications to be retrieved by tag or published as an RSS feed, the project's web site should expose such a catalogue. | IP (66%) |
| All presentation materials for which this is appropriate shall be published on the project's web site under a Creative Commons licence (http://creativecommons.org/) or another appropriate license. | IP (66%) |
| All open source software produced by the consortium shall be published on publicly available software repositories such as http://sourceforge.net/ , http://gitlab.com/ or http://osor.eu . | IP (66%) |
| All data sets for which this is appropriate shall be published on the project's web site under a Creative Commons licence (http://creativecommons.org/) or another appropriate license. | IP (75%) |

| Expected Outcomes | | |
|--|--|--|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Public project presentation | C | D6.1 |
| Project Website | C | D6.2 |
| Dissemination plan including conferences and journals. | IP | D6.4 (C) D6.5 (C) D6.6 (pending) |
| Publication repository for conference papers, presentations, journal papers, standard contributions and patents. | IP | D6.8 (C) D6.9 (C) D6.10 (pending) |
| Networking with stakeholders | IP | D6.3 (C) D6.4 (C) D6.5 (C) D6.6 (pending) |
| Project Showcase | NS | D6.7 (pending) |

6.6.2 Task T6.2 Exploitation and IPR management

| Work Plan | |
|--|--|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Awareness-raising and sensitizing activities | IP (66%) |



| | |
|---|----------|
| IP identification, protection and mediation | IP (66%) |
| Market and application studies | IP (66%) |
| Exploitation roadmap | IP (66%) |
| Business development | IP (66%) |

| Expected Outcomes | | |
|--|--|--|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Clear IPR situation for each technology | IP | D6.3 (C) D6.4 (C) D6.5 (C) D6.6 (pending) |
| Exploitation roadmaps and clear paths to market entry for the major technologies | IP | D6.3 (C) D6.4 (C) D6.5 (C) D6.6 (pending) |

6.6.3 Task T6.3 Impact Management

| Work Plan | |
|---|--|
| Activity Description | Status (Completed/In Progress/Not Started) |
| Establishment of the Advisory Board (AB). | C |
| Management of the AB. A meeting with advisory board members. | IP (66%) |
| Identify additional stakeholders and players like manufacturers, application developers and research institutions. | IP (66%) |
| Prepare information and support material for the information of the potential stakeholders in the domains of Big Data and Network management. | IP (66%) |
| Monitor and evaluate the impact of the project's activity and give feedback to the relevant work packages and tasks, mainly task 1.2 on quality management. | IP (66%) |

| Expected Outcomes | | |
|--|--|--------------------|
| Description | Status (Completed/In Progress/Not Started) | Target Deliverable |
| Fully and complete answer to stakeholders needs. | IP | D6.4 (C) |



| | | |
|--|----|--|
| | | D6.5 (C) D6.6 (pending) |
| High market and research impact of the project results. | IP | D6.4 (C) D6.5 (C) D6.6 (pending) |
| Integration of current market trends and need into project framework | IP | D6.4 (C) D6.5 (C) D6.6 (pending) |



7. Key Activities Declared by Each Partner

The ONTIC partners participate in all work packages to a certain extent. However, each partner develops key activities in specific work packages, tasks and deliverables. Firstly, each partner summarizes the activities already completed in which they have been involved during the second year. Secondly, the ONTIC partners identify and declare specific key activities that are still in progress or that they plan to be involved in during the third year. These activities should not be considered as definitive and their contents and adequacy to specific ONTIC tasks and work packages will be discussed in the 4th Plenary Meeting to be held in February 2016 in Athens.

7.1 Ericsson

7.1.1 Activities Completed

7.1.1.1 WP1/WP5

Ericsson has been in charge of the coordination of the WP5 and UC #3. Therefore, most of the activities carried out are related to such items:

- WP5 coordination
- D5.2 edition
- Support to other deliverables review
- UC #3 coordination
- Definition of the general use case framework for UC #3
- UC #3 requirement specification update by using the user stories approach
- UC #3 Policy Governance Function (PGF) design refinement. Data model specification, interfaces definition.
- UC #3 development activities:
 - Policy Governance Function business logic
 - PGF data model and data tier implementation
 - PGF interfaces and GUI
 - Simulation and Recommendation tools
- Providing additional datasets, based on internal tests on different video applications (Youtube, Netflix, hulu, shoutcast, yomvi...)
- Standardization monitoring
- 3GPP User Plane Congestion (UPCON) follow-up. Attendance to 3GPP meetings.
- Regulation follow-up
- D5.8 edition and contribution to it

7.1.1.2 WP6

- Internal dissemination and customer engagement activities with key players in the field.
- Cross product unit collaboration involving policy and analytics product units thanks to the concepts and prototypes developed within ONTIC.
- Mock-up and training provided to the Ericsson team at the MWC 2015
- Contribution to the 2nd International Workshop on Big Data Applications and Principles (BigDap 2015) with a paper titled : “Adaptive Quality of Experience: a novel approach to real-time big data analysis in core networks”.
- Continuous patentability analysis in order to identify suitable inventions to file.



7.1.2 Activities in Progress

7.1.2.1 WP5

- WP5 coordination
- UC #3 coordination
- Task 5.1:
 - Providing additional datasets containing payload information so that QoE can be effectively measured (the current ONTIC dataset is not able to provide enough information for handling QoE as it has been stripped of all the payload information).
 - Refining the Quality of Experience concept and the ways to measure it.
- Task 5.2:
 - Development activities related to UC #3: Policy Governance Function business logic, PGF data model and data tier implementation, PGF interfaces and GUI, Mitigation Plan Simulation and Recommendation tools.
- Task 5.3:
 - Current work focus on regulatory issues. This work will be reported at the end of the third year. The team will also follow up potential 3GPP groups where the ONTIC contributions could be relevant.

7.1.2.2 WP6

- Collaboration with product units in order to capture additional requirements for the PGF Mitigation Plan Simulation Tool and, if possible, to add it to the product offering.

7.1.3 Plans and Actions for the Next Reporting Period

- UC #3 coordination
- WP5 coordination
- UC #3 requirement specification updated (if needed)
- Task 5.1:
 - Integration in the UC#3 PoC of the algorithms developed by the research partners that deals with the new dataset structure needed in the UC#3.
 - Related to the PGF:
 - Integration of the Analytics Function developed by other partners into the UC#3 PoC.
 - Connection of the PGF to actuation capabilities.
 - Gathering more complete datasets, so that they can be used to measure QoE and as input to the Analytics function.
- Task 5.2:
 - Giving a mathematical definition for the Quality of Experience concept. Taking into account the parameters extracted by the Tstat tool.
 - Implementation of the actuation part of the system to enable a real actuation on the network side based on the predictions, provided by the AF, and the policy plans, provided by the PGF.
- Task 5.3:



- Standardization and regulation conclusions.

7.1.3.1 WP6

- Continue the Internal dissemination and engagement activities with key players in the field.
- Pushing forward the AQoE concept and adapting it to new architecture specifications
- Inclusion of some of the functionalities developed in ONTIC to the Ericsson product offering.

7.2 UPM

7.2.1 Activities Completed

7.2.1.1 WP1

UPM have contributed actively in the edition and revision of deliverable D1.2. UPM are in charge of the project coordination. The main activities we have carried out during this year in this work package are the overall project management, coordination and follow up to ensure that the project successfully achieves its stated objectives on time, within budget and with the expected level of quality. These activities are included in tasks T1.1 and T1.2 and will be done during the whole life of the project.

7.2.1.2 WP2

UPM have tested the native Turbocap API in order to compare its performance to that of the libpcap library. UPM have also carried out an anonymization process on the ONTS dataset. The ONTS dataset has been made freely available for ONTIC partners and publicly available to the rest of the research community on request. UPM have developed scalable unsupervised feature selection algorithms, which are being used as part of the feature engineering mechanisms for preprocessing the ONTS dataset with the goal of adapting and preparing its contents to the specific data formats that have been used in WP3 and WP4. These activities are included in tasks T2.2 and T2.3.

7.2.1.3 WP3

In year 2, UPM have focused their research work on the design and development of scalable parallel unsupervised learning algorithms on top of the novel Spark platform. Two scalable subspace clustering algorithms and four feature selection algorithms have been proposed. These activities have been carried out in task T3.1. UPM have contributed actively in the edition and reviewing of D3.2 deliverable.

7.2.1.4 WP4

UPM have devoted their efforts mainly to studying the applicability of different forecasting algorithms to the problem of congestion control. A framework of online forecasting algorithms has been designed and implemented to be applied in the proactive network congestion control use case in WP5. The results of this work have materialized as significant contributions to deliverable D4.2.

7.2.1.5 WP5

We have expanded an existing simulation environment based on the discrete event simulator PeerSim for running the congestion control use case experiments. A full set of experiments have been designed to test the newly designed congestion control protocols. We have designed a novel architecture to integrate the forecasting module into the congestion control system.

7.2.1.6 WP6

UPM actively participated in the second BigDap workshop, co-located with the ADBIS international conference 2015, and contributed 3 papers. In addition to the BigDap workshop,



UPM have also carried out several dissemination activities. Specifically, UPM have produced and submitted three additional scientific papers to other high-level international conferences whose topics are highly relevant to the goals of ONTIC. So far two of the three submitted papers have been accepted by the 24th European Symposium on Artificial Neural Networks, Computational Intelligence and Machine Learning (ESANN 2016). Additionally, UPM have maintained contact with a dozen industrial and academic entities which have shown a keen interest in the activities being carried out within the project.

7.2.2 Activities in progress

7.2.2.1 WP1

UPM is coordinating the whole project. The level of compliance of each work package with the corresponding milestone is under careful observation. UPM also coordinates to solve problems that rise during the project and to avoid potential risks.

7.2.2.2 WP2

UPM are applying proposed feature selection algorithms to the massive ONTS datasets to achieve dimensionality reduction. UPM is also in charge of the management of the requests to access the ONTS dataset, as well as the ongoing capture and storage process.

7.2.2.3 WP3

UPM and Polito are actively cooperating to the generation of an experimental evaluation of existing scalable clustering algorithms. In addition, we are leveraging the matrix factorization techniques for an exploratory analysis that might help in anomaly detection, study the compatibility of the employed feature selection framework with nonlinear methods and research parallelized versions of well-known factorization algorithms. With respect to the clustering algorithms, we are planning to leverage approximation techniques to reduce the time and space complexity of the affinity propagation algorithm, and design a parallel version to strengthen its scalability with respect to the size of datasets. In addition, several new subspace clustering algorithms are under design and implementation.

7.2.2.4 WP4

We are designing a parallel version of a new online clustering algorithm based on variational inference. A strengthened version of the forecasting framework with parallel versions of forecasting techniques is under implementation, including a network traffic analysis component.

7.2.2.5 WP5

UPM has been involved mainly in use case 2 (Proactive Congestion Detection and Control). We are carrying out a thorough feature-extraction process at the router link level to leverage different types of machine learning techniques.

7.2.2.6 WP6

UPM is collaborating with Polito for a scientific survey on clustering algorithms based on valuable experimental comparisons. Several novel unsupervised learning algorithms have also been recently proposed and experiments on the ONTS dataset and other synthetic and real datasets are intensively carried out to compare with other alternatives. More scientific papers are under edition and submission. UPM is targeting prestigious international conferences and scientific journals whose topics are highly relevant to the goals of ONTIC. A non-comprehensive list of these is presented in deliverable D6.5.

7.2.3 Plans and Actions for the Next Reporting period

7.2.3.1 WP1

UPM will be in charge of the project coordination. The main activities to be done during next year in this work package are the overall project management and coordination and the follow



up to ensure that the project successfully achieves its stated objectives on time, within budget and with the expected level of quality.

7.2.3.2 WP2

Due to the huge volume of data, experiments of applying proposed feature selection algorithms to the massive ONTS datasets will be carried out on the enterprise grade cloud environment to be made available to the consortium. UPM will also be in charge of the management of the requests to access the ONTS dataset, as well as the ongoing capture and storage process.

7.2.3.3 WP3

A joint journal paper will be generated between UPM and Polito based on the experimental evaluation of existing scalable clustering algorithms. A thorough set of experiments on the ONTS dataset has been planned for the newly designed and implemented clustering algorithms.

7.2.3.4 WP4

We plan to integrate the forecasting module into EERC network protocols to improve their performance in transient state. We will research online versions of the feature selection and subspace clustering algorithms developed in WP2 and WP3.

7.2.3.5 WP5

We plan to integrate the forecasting module developed in WP4 into the congestion control system. Additionally, we also plan to apply the feature selection and subspace clustering algorithms proposed in WP2 and WP3 to the other two use cases.

7.2.3.6 WP6

Research on scalable feature selection, online and offline clustering and forecasting techniques has great potential to materialize in scientific papers, patent applications, contacts with industry and academia, contributions to the Big Data PPP and the rest of dissemination and exploitation activities declared in WP6. UPM will also actively participate in the third BigDAP workshop (2016).

7.3 Polito

7.3.1 Activities Completed

7.3.1.1 WP1

Polito actively contributed to Deliverable D1.2, and served as document reviewer. Furthermore, Polito supported the coordinator in management activities when requested.

7.3.1.2 WP2

Polito operated a local Hadoop+Spark cluster and contributed to the establishment of the Big Data Lab within Polito premises, jointly with external funding. These resources have been exploited to design, test, develop and scale the WP3 algorithms.

7.3.1.3 WP3

Polito was leader of D3.1, completed at M12, and D3.2, completed at M24, where state of the art survey and code of the algorithms have been released. Currently, no task is fully completed, as the state of the art in Big Data is evolving at a fast pace and we are still following the latest trends, even if with a marginal effort (95% completion of the state of the art analysis).

7.3.1.4 WP5

Polito actively contributed to UC#3 definition, analysis, and solutions by exploiting algorithms and results from WP3 to be specifically customized to address the application domain and the use-case challenges, jointly with Ericsson, and supported by EMC, UPM and AdaptIT.



7.3.1.5 WP6

Polito heavily contributed to the dissemination activities by leading the organization of the second BigDAP workshop (September 2015), co-located with the ADBIS international conference, and publishing the highest number of peer-reviewed papers in different international venues among all partners. Furthermore, Polito kept contact with different companies interested in the ONTIC results (e.g., Telecom Italia, Fastweb, etc.), and actively performed joint work with other European projects (mPlane, led by prof. Marco Mellia at Polito, and BigFoot, led by prof. Pietro Michiardi at EURECOM).

7.3.2 Activities in Progress

7.3.2.1 WP1

Polito is supporting the coordinator in management activities when requested, such as organizational matters, coordination and suggestions on the high-level project directions.

7.3.2.2 WP2

Polito is operating a local Hadoop+Spark cluster and is directly involved in the management of the Big Data Lab within Polito premises, jointly with other researchers. These resources are being exploited to develop, test, and scale the WP3 algorithms before the experimental ultra-large scale validation on the Google Cloud Platform.

7.3.2.3 WP3

Polito is continuing the development of the algorithms in WP3.

In particular, WP3 consists of three main design and development tasks (T3.1, T3.2, and T3.3), which started in M03 and will end in M33, and an experimental evaluation task (T3.4), which is not started yet. Each of the design and development tasks addresses a very different class of data mining techniques:

- T3.1 addresses unsupervised clustering
- T3.2 focuses on supervised classification
- T3.3 aims at extracting correlations by means of association-rule mining

Hence, the structure and contribution of each task is similar in structure, but different in content. All tasks contributed to D3.2 (Scalable algorithms for offline network characterization) with different content depending on the specific class of algorithms, and the development will continue until M33.

Furthermore, Polito is collaborating with UPM for a shared experimental evaluation of existing scalable clustering algorithms.

7.3.2.4 WP5

Polito is actively contributing to UC#3 analysis and solution design by exploiting algorithms and results from WP3 to be specifically customized to the application domain and the use-case challenges, jointly with Ericsson, and supported by EMC for the development, UPM and AdaptIT as network domain experts.

7.3.2.5 WP6

Polito is heavily contributing to the dissemination activities by leading the organization of the third BigDAP workshop (2016), co-located with the ADBIS international conference in Prague (proposal submitted, pending approval), and is preparing papers to be published in different venues with the most recent algorithms developed. Furthermore, Polito is keeping contact with different companies interested in the ONTIC results. Finally, Polito is collaborating with UPM for publishing a survey on clustering algorithms with valuable experimental comparisons.



7.3.3 Plans and Actions for the Next Reporting period

During the next reporting period, WP3 activities will shift from mainly focusing on algorithm design and development to their experimental evaluation with the major contribution of the Task T3.4 starting from M30, according to the work plan.

Each of the ongoing tasks will contribute on a very different class of data mining techniques:

- T3.1 addresses unsupervised clustering
- T3.2 focuses on supervised classification
- T3.3 aims at extracting correlations by means of association-rule mining.

Results will be collected into the Deliverable D3.3 at M36.

Furthermore, dissemination activities will continue in WP6 with new publications stemming from the ongoing work on algorithms, which will be also at the core of the WP5 efforts related to UC#3, aimed at providing a specific solution able to exploit data mining techniques in the given context.

7.4 CNRS

7.4.1 Activities Completed

7.4.1.1 WP2

In order to have computing resources for testing the algorithms and use case in which CNRS is involved two computing servers Dell FC430 and their related FX2 bay have been bought and installed locally. The machines are equipped with two processors having 14 cores each (or 28 in hyperthreaded mode).

7.4.1.2 WP4

CNRS is involved in three tasks of WP4, and is leading this workpackage. Task 4.1 related to the global architecture of the online classification system has not been changed much according to what was proposed in D4.1 at month 12. It was validated by the researches and results got during second project year.

Task 4.3 aims at proposing an online anomaly detection algorithm. In task 4.3, CNRS did provide new versions of the UNADA algorithm that was at the basis of the proposal. It was improved in order to cope with scalability and accuracy issues. An incremental version was designed, and tested. Its performances were widely studied especially by testing many sets of different features. These tests exhibited that simplest algorithm configurations can lead to very efficient solutions in terms of detection accuracy, and provides shorter computation times and scalability. These tests were run using the MAWI and KDD'99 public datasets, as CNRS was not granted the SATEC one for legal issues. It took eight months of the second year of the project to lawyers to solve the legal issue. Tests on SATEC dataset are then re-scheduled for project year 3.

7.4.1.3 WP5

CNRS is involved in Use Case #1 on anomaly detection. The work done relates to T5.1. It involves a strong collaboration with SATEC that lead to the design of a global architecture for the use case, integrating results from WP2 and WP4 (online access to collected flowing traffic, classification results, anomaly detection algorithms, and a full dashboard for displaying in real time all results). All interfaces and messages formats have been fixed.

7.4.1.4 WP6

In terms of dissemination the two main results of CNRS during the second year of the project deals with:



- Starting a new engineering master on security in which results of ONTIC WP4 are taught (clustering algorithms and their use in anomaly and intrusion detection).
- Starting a new research and technology transfer project with the BORDER6 company for the UNADA algorithms that needs to be adapted to the specific devices designed and commercially distributed by the company.

7.4.2 Activities in progress

7.4.2.1 WP4

The activities in progress are related to continuing the design of the scalable and efficient versions of the anomaly detection algorithms. As CNRS and SATEC lawyers have been negotiating an agreement on the use of SATEC trace dataset, CNRS researchers didn't have access to the SATEC dataset until the last week of October, 2015. CNRS is then now putting a strong effort in re-doing all tests on this new dataset.

7.4.2.2 WP5

The work in progress relates to T5.2 and to assessing the WP4 (Task 4.3) algorithms for Use Case #1.

7.4.2.3 WP6

CNRS is currently working on the writing of papers about the work performed during the second year of the project. It was not possible to publish them before because it was not possible to provide tests on the tangible SATEC dataset.

7.4.3 Plans and Actions for the Next Reporting Period

7.4.3.1 WP4

The algorithm for anomaly detection will be stabilized. An in depth evaluation will be run. A synthetic dataset for assessing the anomaly detection will be created. The anomaly detection algorithm will also be integrated and evaluated with the other algorithms designed and developed in the framework of WP4.

7.4.3.2 WP5

Use Case #1 on anomaly detection will be finalized, deeply evaluated, and presented as a demo for leading to the end both task T5.1 and T5.2.

7.4.3.3 WP6

Actions for third project year will include the writing of scientific and technical papers about ONTIC outcomes, especially WP4 and WP5.

We will also prospect for new technology transfer contracts. A new contact is arising with Airbus for the anomaly detection algorithm.

7.5 SATEC

7.5.1 Activities Completed

7.5.1.1 WP2

During the second period SATEC has been contributing to work package 2 (Big Data Traffic Summary Data Set) in the designing, deploying, managing and provisioning of the capture system which is deployed and running at Interhost. SATEC has been also dealing the issues relating to the public access to the data (data privacy and protection, NDAs, legal basis, etc.). These activities have been carried on the task T2.2 and T2.3 and have been documented in the deliverable D2.5.



7.5.1.2 WP5

During the second period SATEC has been actively working in the task T5.2 (Verification, Use Cases and Field Trials) Work Package 5 (Implementation, Integration and verification). In this period an extension of the requirements specification of the Use Case # 1 has been delivered and the first operating implementation (as a demonstrator) of the analytical control dashboard subsystem, including its internal architecture based in Big Data technology, has been implemented.

At the end of the period a simple integration between Anomaly Detection engine and dashboard subsystem has been completed and a demonstrator of the Use Case #1 which uses real data captured in Interhost, is available.

Results of all the work described above have been documented in the deliverable D5.2 (progress on the Use Cases).

7.5.1.3 WP6

All the tasks of this WP6 are active during the entire life of the project (from month 1 to month 36), but during the second year of the project, as the leader of WP6, we worked on getting all partial results are reflected in the following deliverables:

- **Deliverable D6.5 (Progress on Exploitation and Dissemination Plans -Part I)**
- **Deliverable D6.9 (Big Data Analytics Workshop #2)**

7.5.2 Activities in progress

7.5.2.1 WP2

The WP2 tasks in progress on the next period in which SATEC will be involved as key activity are:

- **T2.3 Deployment, Configuration and Provisioning of Big Data Network Traffic Summary Dataset** [from 9 month to month 36]

7.5.2.2 WP5

The WP5 tasks in progress on the next period in which SATEC will be involved as key activity are:

- **Task 5.2 Verification, Use Cases and Field Trials** [from month 1 to month 36]:

7.5.2.3 WP6

The WP6 has kept open its three tasks during the second year of the project and the three will continue to progress until the end of the project:

- **Task T6.1 Dissemination, Training and Public Awareness** [from month1 to month 36]: which will continue with the dissemination planned activities. This task will contribute in the next period to the development of the deliverables D6.6 (Progress on Exploitation and Dissemination Plans - Part II), D6.7 (Project Showcase) and D6.10 (Big Data Analytics Workshop #3)
- **Task T6.2 Exploitation and IPR management** [from month1 to month 36]: which will continue with the exploitation and IPR activities as described in the deliverable D6.3 (Exploitation Strategy). This task will contribute in the next period to the development of the deliverable D6.6 (Progress on Exploitation and Dissemination Plans - Part II)
- **Task T6.3 Impact Management** [from month1 to month 36]: which will continue with the activities focused in the positioning of ONTIC in different application domains. This task will contribute in the next period to the development of the deliverable D6.6 (Progress on Exploitation and Dissemination Plans - Part II)

7.5.3 Plans and Actions for the Next Reporting Period

7.5.3.1 WP2

On the next period SATEC will contribute to the Task T2.3 (Deployment, Configuration and Provisioning of Big Data Network Traffic Summary Dataset) of Work Package 2 (Big Data Network Traffic Summary Dataset).

SATEC will propose to add the dashboard implemented in the Use case #1 as part of the visualization layer of the Big Data platform.

SATEC's contributions will be documented in the following deliverables:

- **Deliverable D2.6 (Integration, validation, and release of ONTIC Big Data Architecture and Provisioning Subsystem)**
- **Deliverable D2.7 (ONTS (ONTIC Network Traffic))**

7.5.3.2 WP5

On the next period SATEC will be mainly focused in the task T5.2 (Verification, Use Cases and Field Trials) of Work Package 5 (Implementation, Integration and verification). The goal is to finish a full prototype of Use Case #1 so that SATEC will perform the following activities:

- Final requirement specification for Use Case #1.
- Dashboard subsystem final design (functional and Big Data architectural design).
- Final integration between Anomaly Detection Engine and Dashboard subsystem.
- Final Use case #1 prototype implementation.
- Dashboard tests and validation (performance and scalability).

As a result of this work we will contribute to the deliverable D5.4 (Use Case #1 Network Intrusion Detection) with documentation about use case #1 and specifically about the dashboard subsystem.

7.5.3.3 WP6

During the next period SATEC plans to continue working in activities related with dissemination and exploitation specially trying to carry ONTIC results to domains such as: Tele-communications, Smart Grids, IoT (Internet of Things), Smart Cities, M2M (Machine to Machine), e-Health, etc.

The following deliverables will be delivered at the end of the next period:

- **Deliverable D6.6 (Progress on Exploitation and Dissemination Plans - Part II/final)**
- **Deliverable D6.7 (Project Showcase)**
- **Deliverable D6.10 (Big Data Analytics Workshop #3)**

7.6 EMC Spain

7.6.1 Activities Completed

As a partner in the ONTIC consortium, EMC has contributed in supporting activities to all WPs defined in the project. Specifically, these are the main activities completed during the second year:

7.6.1.1 WP2

During this second year, EMC has been exploring and testing the different Big Data Cloud Platforms available in the market with the goal of using one of them as Testing Platform for the Ontic Algorithms developed in WP3 and WP4. Finally, Google Cloud Platform has been selected as the most suitable platform to be used by the consortium for testing purposes.



7.6.1.2 WP3

As some of the work made by EMC during this second year is related with the WP5:UC3 Adaptive Quality of experience (QoE) Control, where WP2 Algorithms will be use, EMC has been collaborating in this WP3 to be as much as possible aware about the algorithms and frameworks used for them, as this algorithms will be used at the end in the UC3. For this, we have collaborated also in the review process of the deliverable D3.2: Scalable Algorithms for offline network characterization.

7.6.1.3 WP5

EMC has been working in collaboration mainly with Ericsson and Polito in the development of the Use Case 3: Adaptive Quality of experience (QoE) Control. In this context, we have built and configure a small lab for video testing, where WP2 algorithms will be tested, and also developed some algorithms using Spark Streaming to make a pre-processing of 'tstat tool' output to be used also in the training process of WP2 algorithms.

7.6.1.4 WP6

As a member of the Program Committee for the second international workshop on Big Data celebrated in Poitiers on September 2015, we helped in reviewing and evaluating some of the papers sent to the committee.

Also, EMC have kept working in the task of giving awareness and visibility of the ONTIC project inside EMC through internal/external channels as reflected in deliverable D6.5: Progress on Exploitation and Dissemination Plans.

7.6.2 Activities in Progress

Current activities by EMC are focus in the following tasks:

- Providing a Cloud Platform based on Google platform to be used by the consortium for testing purposes. This activity is shared with Adaptit.
- Giving support in defining and preparing the testing requirements for WP2 algorithms in collaboration with Ericsson and Polito for the UC3

7.6.3 Plans and Actions for the Next Reporting Period

For the next period , these are the main activities that will be carrying out by EMC:

- Giving support and helping in the testing process of the different algorithms developed by WP3 and WP4 using the Google Cloud Platform.
- In collaboration with Ericsson and Polito, building the mock-up for the UC3, integrating WP2 algorithms within the UC3 scenario.
- Collaborate in the review process for the BigDAP 2016.
- Keep working in giving awareness and visibility of the ONTIC project to our partners and customers, to try to find real applications where ONTIC outputs could be leveraged.

7.7 Adaptit

7.7.1 Activities Completed

7.7.1.1 WP2

- WP2 coordination
- Spark installation and configuration in the Google cloud through the Cloudera's Hadoop open source distribution.
- Hands-on review of devops tools for enabling automated deployment of larger-scale Spark clusters in the Google cloud.



- Sample prototypes for Big Data ingestion in the cloud using Apache Flume and Elasticsearch Logstash.
- Contributions to D2.3 (Progress on ONTIC Big Data Architecture): outline of the ‘Pre-processing’ module; initial description of the deployment and applicability of the ONTIC Analytics System in NSP domains; description of Spark design and configuration for optimising application performance based on working experience; instantiation of the ONTIC architecture on the Google cloud and description of procedures for performing experimentation; overall editing.
- Contributions to D2.4 (Progress on the Provisioning Subsystem) on the design of an application-level procedure for Big Data migration to the cloud; overall editing.
- Attendance of the first review meeting, project plenary meetings and telephone conferences; preparation of presentation material.

7.7.1.2 **WP3**

- Review of the updated version of Deliverable D3.1.

7.7.1.3 **WP4**

- Review of Deliverable D4.2.

7.7.1.4 **WP5**

- Review of the updated version of Deliverable D5.1.
- Detailed comments on the SLBN max-min fair allocation algorithm for rate-based congestion control used in congestion avoidance use case.
- Hands-on review of programming libraries for visualizing large network topologies (Python igraph and network, Javascript D3JS) for building the front-end of the congestion avoidance use case.
- Review of Deliverable D5.2.

7.7.1.5 **WP6**

- Presentation of the ONTIC project in the Mobile World Congress, Barcelona, February 2015.
- Concretization of the business case with First Data Hellas on Big Data usage-based cost and revenue analysis and forecasting; coordination for building a pilot prototype based on actual operations data and demonstrations to senior technical staff and executive management.

7.7.2 Activities in Progress

7.7.2.1 **WP2**

- Administering and maintaining an instance of the ONTIC Big Data architecture in the Google cloud in various cluster configurations as required by algorithm experimentation needs.
- Study and hands-on verification of recommended practices for Spark application performance tuning.

7.7.2.2 **WP3**

- Follow-up development of off-line ML algorithms for traffic classification and assessment of their applicability from the business perspectives of an SME solution provider.



7.7.2.3 WP4

- Follow-up development of on-line ML algorithms for short-term forecasting and anomaly detection and assessment of their applicability from the business perspectives of an SME solution provider.

7.7.2.4 WP5

- Collaboration with UPM on the congestion avoidance use case.
- Follow-up development of the anomaly detection and QoE management use cases.

7.7.2.5 WP6

- Promotion of ONTIC through meetings with customers.
- Exploitation of gained know-how in professional services and pre-sale activities.

7.7.3 Plans and Actions for the Next Reporting Period

7.7.3.1 WP2

- Support of project experimentation activities in the Google cloud; set-up of different cluster configurations, initialization scripts, import of test traffic datasets, reports on algorithm execution.

7.7.3.2 WP3

- Visualization of the developed algorithms for quality of clustering.

7.7.3.3 WP4

No particular activity is envisaged besides the on-going ones.

7.7.3.4 WP5

- Development of a suitable Web-based front-end for demonstrating the effect of the developed ML algorithms in the context of the congestion avoidance use case.
- Creation of synthetic flow-based traffic data based on statistical characteristics of respective distributions (inter-arrival times, number of active flows) drawn from analyzing the ONTS dataset; this is for creating appropriate test scenarios for exhibiting the congestion avoidance use case.

7.7.3.5 WP6

- Further development of the business case with First Data Hellas; enrichment of the provided solution with ML algorithms based on gained know-how from the project.
- Pursue of similar business cases with other customers; there is already one in the pipeline with Cosmote Greece, nowadays a convergent network (fixed and mobile) service provider, part of the Deutsche Telecom Group.



8. Compliance with Planning Milestones

This section discusses the level of compliance of each work package with the second milestone (MS2).

8.1 WP2

WP2 has progressed according to plan in both of its work dimensions: Big Data architecture and the ONTS dataset containing summarized IP traffic records from ISP operational environments.

The ONTIC Big Data architecture has been specified from system capabilities and technology viewpoints along with its deployment and applicability in NSP domains. The proposed architecture combines off-line and on-line analytics and the selected Apache Spark distributed computing technology realizes this integration in a streamlined and inexpensive manner. Local instances of the architecture have been instantiated in partners' premises and in the Google cloud.

The ONTIC Provisioning System for capturing, managing and making available the ONTS dataset has been refined. It has been running uninterruptedly at SATEC premises since October 2014 having collected by the time of writing almost 80TB of data. The system has been enhanced with suitable security mechanisms such as IP address anonymization and access procedures as required for protecting data privacy in order to make the captured network traffic traces publically available.

All above work has been documented in deliverables D2.3 and D2.5 and accomplishes milestone MS2 (stable validated mechanisms for off/on-line analytics). During the last year, WP2 work will focus on: supporting the project experimentation activities; enhancing and further evaluating the specified feature selection algorithms; and, on refining the Big Data architecture and the provisioning system.

8.2 WP3

Deliverable D3.2 delivers the software (code) of the developed algorithms for the offline network traffic characterization system. D3.2 was released according to schedule at M24.

8.3 WP4

Given the delay due to the negotiation between SATEC and CNRS lawyers during which CNRS researchers were not granted the access to SATEC dataset, the validation mechanisms have not been tested much. MS2 related to issuing a stable version of online validation mechanisms is then not fully achieved. Based on our experience with this kind of validation, validation mechanisms have been issued on time, but not enough validated yet. The validation of the mechanisms will then continue on the very beginning of the third year of the project. But there shouldn't be many changes with regard with what has been proposed in D4.2.

8.4 WP5

T5.1

The algorithms developed by UPM will be tested for the validity of their application to the use cases. Specifically, the unsupervised feature selection techniques will be used as a preprocessing step for UC#3 and to help choose the parameters of the time series analysis methods used in UC#2. The subspace clustering algorithms will be integrated into the offline network traffic characterization system. The subspace clustering results can help to generate ground-truth knowledge of the network traffic data, and validate the outcome of the online network traffic characterization system. The consortium will also study the



applicability of these techniques as an alternative to the clustering algorithms used for UC#1.

T5.2

The algorithms developed by UPM have been tested on a moderate cluster, showing good scalability and efficiency. In the third year, they will be tested on a large cluster to further test their scalability and their ability to ingest huge datasets.



9. Current and Potential Issues/Problems/Risks

9.1 WP2

No major issues holding up work progress have been encountered during the reporting period.

The technical challenges pertinent to real-time traffic capturing and the flow aggregation of a huge number of packet traces have been addressed and satisfactory working solutions were put in place. The delicate issue of data privacy underlying public availability of network traffic traces has been resolved according to current legislation and consultation with experts. WP2 will continuously look for changes and/or alternatives for enhancing the capabilities and deployability of the ONTIC Big Data architecture and for yet improving the accuracy and performance of the ONTS capturing, managing and availability processes.

9.2 WP3

Year 2 activities of WP3 focused mainly on algorithm design and development, whose results were successfully included in D3.2 on time. Therefore, WP3 did not present issues or problems during the reporting period.

9.3 WP4

Dealing with WP4, the anomaly detection algorithm design and assessment has been impacted by the eight months duration of negotiation between SATEC and CNRS lawyers, for granting the access to SATEC dataset for CNRS researchers involved in ONTIC. The delivery of the final version of the algorithm and its related evaluation will then be delayed, thus certainly leading to the need of extending task T4.3. It will also have an impact, despite less important, on Use Case #1 of WP5.

9.4 WP5

As described in deliverables D5.2 and D2.5, the main challenge UC#3 has had to deal with is the lack of application payload in the ONTS dataset. It prevents the Key Performance Indicators defined in project's first year from being computed. The risk control strategy has been two-fold: on one hand, the project has defined a VLC client/server test-bed, deployed at EMC2 premises, to obtain application payload in a safe and secure way; on the other, Ericsson has internally sought for datasets that, even smaller, could contain the necessary information for KPI estimation and enable the analytics tasks performed by WP3. This analytics task will be another challenge (considering feature reduction, prediction, etc.) for a smooth integration with the Policy Governance Function.

However, the risk control strategies will help to fix the lack of application payload issue, there is another key challenge that needs to be managed along next year and it is related to the so called "actuation" part. To cope with this challenge the team has deployed router with bandwidth regulation in the VLC simulated environment in order to enable the actuation part in the UC#3 implementation. The integration of all these components into an end-to-end proof of concept is full of challenges that need to be managed in the year three of the project.

9.5 WP6

The activities of task T6.1 have not presented any issues during the reporting period and some of specific outcomes (for example networking with stakeholders) are progressing faster than expected. The activities of tasks T6.2 and T6.3 have not presented any issues during the reporting period. However, we should review progress on IPR and publications, and analyze possible risks in the next year.



Although the long negotiation process between SATEC and CNRS lawyers has slowed down the experiments on the ONTS dataset, affecting as well our ability to produce related publications, this problem has now been finally solved. CNRS expects to retake these experiments and produce new publications during the next period of the project.