EUROPEAN COMMISSION
Community Research

SEVENTH FRAMEWORK PROGRAMME

Specific Targeted REsearch Project

PRISM

D2.1.2: Scenarios and System Requirements

Project acronym: PRISM
Project full title: Privacy-Aware secure Monitoring
Contract No.: 215350
Project Document Number: IST-2007-215350-WP3.2-D3.2.1-R1
Project Document Date: 19/09/2008
Workpackage Contributing to the Project Document: WP2.1
Deliverable Type and Security: Public (Confidential)

Author(s): Brian Trammell, Elisa Boschi, *Hitachi Europe*
Esa Hyytiä, Ivan Gojmerac, *ftw.*
Carsten Schmoll, *Fraunhofer FOKUS*
Andrea Di Pietro, Saverio Proto, Simone Teofili, *CNIT*
Enrico Stinco, *Nettare s.r.l*
Georgios V. Loudakis, Anna Antonakapoulou, Fotios Gogoulos,
Dimitra I. Kaklamani, Iakovos S. Venieris, *ICCS*
Felix Strohmeier, *Salzburg Research*
Francesca Gaudino, *Baker & McKenzie*

Abstract:

This deliverable defines a requirements classification, specifies the functional and technical requirements for the PRISM architecture, and presents scenarios of hypothetical deployments of the PRISM system to solve example existing network traffic measurement problems.

Keyword list: PRISM, IST-2007-215350, Requirements, Scenarios

History

Version	Date	Description, Author(s), Reviser(s)
0.1	7.3.2008	Document creation, Elisa Boschi
0.2	25.7.2008	Reorganization and incorporation of new material, Brian Trammell
0.3	8.8.2008	Incorporation of FE/BE requirements and contributions to date, Brian Trammell ed., et. al.
0.4	10.9.2008	Completion of scenarios and draft completion of requirements, Brian Trammell ed., et. al.
0.5	17.9.2008	Integration of Vienna plenary meeting changes, Brian Trammell ed., et. al.
1.0	19.9.2008	Final version, Brian Trammell ed., et. al.

Table of Contents

Abbreviations

IPFIX	Internet Protocol Flow Information eXp ort
PRISM	PRivacy-aware Secure Monitoring
WISP	Wireless Internet Service Provider

List of Figures

List of Tables

1 Introduction

PRISM is a general-purpose network traffic monitoring and measurement system that provides strong protection for personal data in monitoring and measure ment applications. It does so through a two-stage architecture, which separates measurement tasks between a front end, which observes traffic in the network, and a back-end, which stores and processes the results of the measurements. The architecture effec tively separates trust between these stages, applying new cryptosystems and data protection techniques to captured data as early in the measurement process as possible, and using semantic access control to finely control access to protected information bas ed upon the wider context of the information requested, the entity making the request, and the purpose for which the request was made. The core system provides the basic support for the development and deployment of privacy -aware monitoring applications; the project will develop a pilot implementation of this core system and, in addition, adapt selected monitoring applications to operate in concert with this core.

This document specifies the requirements for the PRISM system. Section 2, Requirements and Requirements Classification, defines a requirements classification and enumerates functional and technical requirements within the context of that classification. As the system is intended to be a framework in which privacy-aware monitoring applications are d eveloped, the measurement functionality requirements are designed to be as flexible as possible. The requirements of the system cannot be described in isolation from the legal and regulatory environment into which it will be deployed. We have already provi ded a survey of this legal and regulatory environment within Europe and selected jurisdictions in a previous deliverable [ref 2.1.1]; the Essential Privacy and Security and the Legal and Regulatory requirements within this document concretely apply this en vironment to the architecture's requirements. Finally, the technical requirements specify constraints on how the system must provide the functionality it does.

Section 3, Scenarios, examines a variety of real world network monitoring and measurement scenarios drawn from the experience and research of the project partners. These scenarios first briefly cover the current state in each situation, and then select a future potential measurement application to be enabled by PRISM, then present a hypothetical use case for the given application.

2 Requirements and Requirements Classification

This section specifies the requirements for the PRISM architecture, and defines a classification for these requirements (See Figure 1)

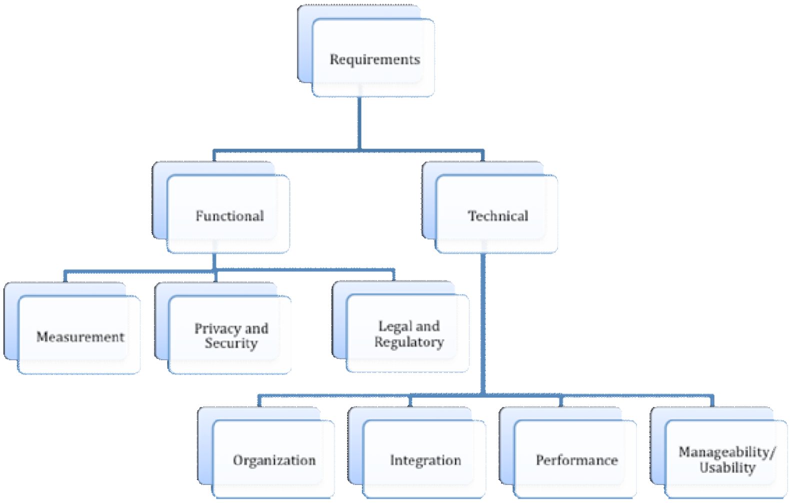

Figure 1: Requirements for PRISM architecture

Requirements are broadly divided into two categories. Section 2.1 specifies functional requirements, which define what the system must do. Section 2.2 specifies technical requirements, which define constraints upon how the system must do what it does. These are then divided into more specific classes, as noted in each section below. Each requirement herein is a concise, easily tested statement of functionality or constraint that should be provided, enabled, or followed by the architecture.

A pilot implementation of the architecture as produced by the project must not represent architectural or technical decisions that preclude a complete implementation of the requirements.

Section 2.3 contains a requirements summary table summarizing each requirement, and the subsection(s) of section 2.1 and section 2.2 from which they are drawn, and is intended as a non-normative quick reference to the requirements for further development of the architecture and implementation of the system.

Note that there are two different levels of requirements specified herein. Requirements defined using the words “must” or “shall” must be met by the architecture without variance, and represent the minimum functionality and technical aspects of the system. Requirements defined using the world “should”, or words such as “to the extent possible” may be modified as research into the problem space addressed by PRISM continues, or to meet technical requirements as necessary.

2.1 Functional Requirements

Functional requirements, as noted, define what the system must do. The functional requirements here are divided into three classes. Measurement Functionality requirements define the space of problems the PRISM architecture must be applicable to in terms of their constituent operations. These are drawn in part from an examination of the Scenarios, which appear later in the document. Essential Privacy and Security requirements define those guarantees that the system must make to the operators and users of the measured networ k with respect to fundamental rights to privacy. Finally, Legal and Regulatory Compliance requirements enumerate those requirements drawn from laws or regulations that must be followed and are not reducible to first principles regarding the right to privac y; these are largely drawn from D2.1.1.

2.1.1 Measurement Functionality

The system shall provide basic measurement application services based upon full packet, partial packet, flow, and flow summary monitoring. This section defines the basic measurement operations the system must be able to perform. Note that certain of these functions have been assigned to specific components as outlined in the Organization section below.

2.1.1.1 Front-end packet capture component

The system must have a component that captures and accep ts raw packet data from a network under measurement. This component must:

- Capture the traffic flowing over a gigabit link (including the headers and the whole payload of each packet) with a precise timestamp.
- Classify each packet based on a classification rule set defined based on a specified flow key and set of flow or packet properties.
- Distribute classified packets or portions thereof to a specific processing component based on a demultiplexing rule set.

2.1.1.2 Front-end processing, protection, and coordinatio n components

The system must have a set of components that generate measurement data from packets collected by the packet capture component from a network under measurement. These components must process information from the front -end packet capture compon ents, and perform the following operations:

- Extract information from packet header and payload sections.
- Summarize packets into flows or aggregates of flows.
- Classify and label packets based on header data, packet content, and flow properties.
- Recognize patterns within packet headers and content, and across multiple packets in a flow, and generate pattern recognition event data.
- Protect extracted packet, flow, aggregate, and summary data by encryption or anonymisation.
- Forward generated and protected data t o back-end components.

Note that the extent of the separation between these components and the packet capture components, and the protocols used among them, are implementation details to be decided in harmony with the Integration and Performance technical requirements below.

2.1.1.3 Back-end storage, analysis, and access control components

The back-end components of the system must store, protect, and provide access to the data received from the front-end. There are three ways in which the back -end components may provide access to the stored data:

1. Direct access to protected or unprotected packet or flow data. Here, the back -end returns “raw” data for input to a monitoring application external to the PRISM back - end components; this access method is used for integrati on with existing monitoring applications.

2. Access to data generated by a specific processing component on the front -end. This allows pre-processing (e.g. pattern detection, summarization) to run on the front -end.
3. Integrated access to data for PRISM-aware applications. In this case, the monitoring application interacts directly with the PRISM back-end via a monitoring application programming interface, that allows for direct operation on protected data.

Access to data via all three of these methods must be c ontrolled with respect to the user, the data accessed, and the intended use of the data.

2.1.1.4 Supported network layers and addressing

The system must support the capture of traffic from IPv4 networks, and the processing of traffic data with IPv4 addressing info rmation.

To the extend possible, the system should support the capture of traffic from IPv6 networks as well, and the processing of data with both IPv4 and IPv6 address information.

2.1.1.5 Support for specific monitoring applications

Most existing monitoring app lications operate on a relatively limited set of data formats and interfaces. Applications that require packet payload data (e.g. signature -based network intrusion detection) generally either directly access the monitored network interface or use a common packet trace file format (e.g. pcap dumpfile) for offline analysis. Applications requiring more aggregated data such as flow or aggregated flow data generally support a de facto flow data standard such as NetFlow V5, or a defined standard format such as IP FIX.

The system must support, at minimum, the capture of packet data from raw network interfaces, and provide raw packet information for analysis. It must be able to provide full packets or partial packets (e.g. headers). It must also provide flow data for analysis supporting the minimum set of flow fields defined by NetFlow V5, via IPFIX, extended with fields required to support IPv6 addressing; note that through flexible definition of the flow key provided by IPFIX, this implies support for flow and aggre gate flow input. It should support, to the extent possible, the acceptance of files containing packet data in pcap dumpfile format, and the acceptance of files containing flow data in IPFIX File format.

Each component of the system should support the addit ion of modules to support new formats in order to support future monitoring needs.

2.1.2 Essential Privacy and Security

This section defines requirements for privacy protection and system security, without reference to specific requirements of applicable laws a nd regulations. These requirements are, however, derived from the basic principles commonly shared among European Union countries that represent the foundation of European Union legislation on data protection and security.

2.1.2.1 Protection of network end users f rom identification

The personal data of the end users of a network shall be processed in a form that allows identification of the end users only when said identification is necessary and functional to achieve the specific network monitoring function that i s sought. To the extent possible, the identification of the end users should be minimized or eliminated, in order to render as much of the processed data as possible “non-personal”. It follows that data of the end users in identifiable form may be processe d only by those components of the system that absolutely need such information to perform their function; e.g., a front end which must necessarily observe packets in the network. When the specific monitoring function does not need to process data in an identifiable form, the system should be able to deploy alternative solutions (for example, use of key-coded data, adoption of anonymisation solutions) that allow protecting the identity of the end users.

2.1.2.2 Protection of unauthorized disclosure of personal data

The personal data of end users of a network shall be protected against unauthorized disclosure. The system must allow the access to personal data of end users only to the components of the system for which the access to the personal data is a prerequisite for the monitoring function that they perform. The system must also allow access to the personal data of end users only to those operators of the system that have a legitimate ground to know such information. Furthermore, the system must protect the personal data of the end users against unauthorized malicious access by external third parties. Lastly, the system must safeguard the personal data of the end users against accidental loss and disclosure. The aforementioned goals should be addressed by the system by adoption of security measures and technical features derived from recognized best practices in system and network security. The system must allow the interception and surveillance of communications and related traffic data only in accordance with and under the limits imposed by applicable laws and regulations.

2.1.2.3 Minimal access to information

Each component of the system must have access only to the information that is strictly required to perform its specific function. This restriction of availability of information within the system also reduces the complexity of the system architecture in terms of the protection of sensitive or legislatively protected information. Indeed, limiting the extent of information to be protected also implies limiting the scope to be taken into consideration by the components of the system devoted to guarantee security and soundness of the information processed.

2.1.2.4 Minimal processing of information

The system should adopt solutions targeted at enabling the processing of the personal data of the end users only when said processing is essential for the monitoring function that is to be performed. Moreover, when the data processing activity is backed up by legitimate grounds, it is also necessary for the system to be capable of offering flexible features that can be fine-tuned to the different requirements to be met, said requirements depending upon the specific processing conditions that occur. For example, it may be the case that personal data are gathered in relation to a given monitoring function and that within said function there is more than one specific purpose that is sought. In this case, the amount and type of information that is essential for the different purposes should be distinguished in two sub set of personal data, and the system should be capable of adapting to these requirements.

2.1.2.5 Anonymisation of personal data

The system shall incorporate mechanisms for the robust anonymisation of data. The anonymisation features should be adopted by the system in order to protect the identity of the end users (see also section 2.1.2.1, above). Moreover, the personal data of the end users must be irreversibly anonymised before any disclosure to an external monitoring application or third party, or when the personal data in identifiable form are no longer necessary to the specific monitoring function that is performed.

2.1.2.6 Privacy-aware semantic access control

Back-end components shall enforce an access control model dependent on the privacy context of each request for information in order to ensure that the system does not disclose any information to any unauthorized entity. This privacy context must enable access to be predicated on any information relevant to the compliance with the laws and regulations governing the protection of personal data, including but not limited to information about the semantic type of the data requested, the intended usage of each access request, and the role of each entity involved in each access request. Additional semantics may be supported by the system as necessary to meet essential privacy and regulatory requirements.

2.1.3 Legal and Regulatory Requirements

The right to personal data protection is acknowledged as a fundamental right of the individuals by the European Union legislation (see for example the European Convention for the Protection of Human Rights and Fundamental Freedoms [1]), and it is also recognized as a constitutional right in certain member states' legislations (for example, Germany, Greece and Italy).

The major regulatory text of the European Union remains the Directive 95/46/EC [2] "*on the protection of individuals with regard to the processing of personal data and on the free movement of such data*". The Directive 95/46 EC is further particularized and complemented with reference to the electronic communication sector by the Directive 2002/58/EC[3], which imposes explicit obligations and set forth specific limits on the processing of users' personal data by network and service providers in order to protect the privacy of the users of communications services and networks. Lastly, the Directive 2006/24/EC[4] is addressed to the providers of publicly available electronic communications services or of public communications networks and imposes a set of obligations with regard to the retention of certain data in order to ensure the availability of said data for purposes of the investigation, detection and prosecution of certain serious crime, as defined by each member state's national legislation.

These three European Directives constitute the basis for the following summary of the main regulatory data protection and security requirements to be taken into account for the specification of the PRISM architecture and operational characteristics. A detailed analysis of the European legal and regulatory framework, also with specific insight into some selected jurisdictions, related to the PRISM project has already been provided in D2.1.1.

It is noted that for purposes of these requirements, personal data are defined as follows in the Directive 95/46/EC, Article 2 (a): *"Personal data shall mean any information relating to an identified or identifiable natural person ("data subject"); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity"*.

Furthermore, the requirements considered for the architecture and operational features of the PRISM project also take into account the specification of the definition of 'personal data' as provided by Art. 29 Data Protection Working Party [5] in its Opinion issued on June 20. 2007 on the concept of personal data.

2.1.3.1 Lawfulness of data processing

The PRISM system shall be able to evaluate the lawfulness of each request for personal data with applicable laws and regulations.

[1] The European Convention of Human Rights and Fundamental Freedoms (7 December 2000, on O.J. 200/C - 364-01).

[2] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data; O.J. L 281, 23 November 1995.

[3] Directive 2002/58/EC of 12 July 2002, concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communication), O.J. L 201/37, 31 July 2002.

[4] Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending Directive 2002/58/EC; O.J. L 105/54, 13 April 2006.

[5] Article 29 Working Party on the Protection of Individuals with regard to the Processing of Personal Data is a Working Party set up by Article 29 of the Directive 95/46/CE; for further information on Article 29 Working Party, please refer to the following web address: http://europa.eu.int/comm/justice_home/fsj/privacy/workinggroup/index_en.htm.

In practice, the system should be structured in a way that it is capable of assessing the legitimacy of a request of access to data submitted by the different components of the system. The lawfulness of a given data processing activity is to be evaluated against the type of data collected and the purposes for which it was collected, taking account not only of the legislation ruling on privacy and data security, but more generally of all applicable laws and regulations. It follows that the system should be configurable with a set of types and purposes deemed to be lawful, and for these specified and pre -identified types and purposes, the system should allow the processing of the personal d ata of the end users.

All such configuration of the system with respect to the lawfulness of stated processing purposes must be done by persons who are competent both in the means used to configure the system and in the applicable legal context.

Any request that is not specifically determined to be lawful according to the set of lawful purposes must be denied.

In practice, if the request is concerned with a certain kind of network monitoring, on the basis of the specific purpose of said monitoring activit y the system should be able to apply the other mandatory legal requirements. For example, the use of data in anonymous or identifiable form would be permitted or not permitted depending upon the specific monitoring function to be carried out.

Above all, the system should guarantee a high degree of flexibility and the capability of applying specific and pre-written rules derived from the applicable regulatory framework.

2.1.3.2 Purposes for which data are processed

The PRISM system shall provide the means for identi fying the purpose of each request in order to comply with the so named "purpose principle ." In practice, the system should function so that it allows the collection and processing of personal data only when said activities are carried out for specified, ex plicit and legitimate purposes. In addition, the system should prohibit that personal data that are collected and processed for some specific and legitimate purposes be used for other purposes that result to be incompatible with these for which the personal data have been originally collected and processed.

The purpose principle also implies that the data controller (notably the entity primarily in charge of the data processing) should act transparently. This implies that the data controller should specify and make explicit the reasons why it is using personal data to the data subjects (that are the entities, as natural persons and in some cases also legal entities, whose data are processed; in our case the data subjects are the end users). To this purpose, the system should allow a certain kind of communication with end users in order to make them explicit the purposes for which their personal data are being gathered and processed, or alternatively, the system should provide technical features that allow a k ind of negotiation with the party submitting the request of processing of personal data, and during said negotiation process the system should be able to verify that the requesting party has complied with the aforementioned requirement towards the data sub jects.

2.1.3.3 Necessity, adequacy, and proportionality of the data processed

The PRISM system shall operate according to the so named "proportionality principle," which requires that the personal data of the end users may be gathered and processed only to the extent that they are adequate, relevant and not excessive if compared with the monitoring function for which said data are collected and processed by the system.

The system in practice should be able to determine what is the amount of personal data that may be processed within a specific monitoring function, and also what type of data may be processed within the same. For example, if the monitoring is aimed at producing statistical figures, the data may be processed in anonymous form, and there is no need of using information that may identify the end users.

Processing activities may be performed only on data that are functional and necessary to the specific purpose that it is sought by the monitoring function. The system should automatically delete or anonymise data that are redundant or no longer needed for a specific monitoring function.

2.1.3.4 Quality of the data processed

The PRISM system shall ensure that the data processed are correct, exact and updated. Moreover, the system should be able to perform correcti ve actions in order to delete or correct inaccurate data, and to delete or update outdated or redundant data.
In addition to these corrective remedies, t he system should also allow periodic audits on the personal data that it stores, so as to verify the le gitimacy of said data.

2.1.3.5 Minimal use of personal identification data

The PRISM system shall minimize to the extent possible the use of identification and personal data only when this is a prerequisite to the specific monitoring function that is to be performed.
When a given monitoring result may be achieved without personal identification data, the system should be able to use anonymous data or alternatively to allow the identification of the data subject only under specific circumstances, for example in ca se of mandatory data retention obligations under Directive 2006/24/EC (see also section 2.1.3.6, below) .

2.1.3.6 Storage of personal data

The PRISM system shall keep personal data in an identifiable form only for the time that it is strictly necessary to the speci fic monitoring function that is carried out. Personal data that are redundant or no longer needed should be deleted or anonymised. As noted above, periodic audits on the data stored by the system should be performed, together with functions that perform automated deletion or anonymisation of redundant or unneeded data.

2.1.3.7 Data retention

The PRISM system shall comply with the requirements set forth by applicable data retention regulations. This implies that the system should store the specific data that are sub ject to the data retention regulation for the time periods specified under the applicable regulatory framework. Moreover, the system should disclose the data only to the law enforcement authorities that are specifically designated and authorized under appl icable legislation.
It should be recalled that compliance with data retention law requirements also implies that the system should fulfil specific and mandatory security requirements to be applied for the storage of the data and relevant access, so for exa mple the data stored for data retention purposes should be kept logically separated from the other data stored by the system.

2.1.3.8 Access limitation

The PRISM system shall authenticate all users of the system, shall provide different levels of access to the stored data, and shall provide for the logging of all access to the stored data in order to detect attempted or successful unauthorized access. These levels of access shall be granted based on the authentication of individual users, the need to know associate d with each individual user's role, and the data to be accessed. For example it may be the case that a specific user profile allows the access and consultation of the data, but does not allow the modification or deletion of the data.

2.1.3.9 Information to and rights of the data subject

The PRISM system shall be capable of informing the data subject that his/her personal data are processed according to applicable data protection legislation.

The means to fulfil this requirement may be either a direct contact with t he data subjects to the extent possible, or a negotiation procedure between the system and the entity asking access to the data in order to make sure that the data subjects have been properly informed. The subset of mandatory information that the data subj ect should receive varies form one member state to another, so it is important that the system allows a high degree of flexibility. In general terms, the data subjects should be informed about the following issues: the purposes and the methods of the data processing; the extent of data communication and/or data diffusion; the mandatory or optional nature of providing his/her personal data and the consequences that he/she may undergo in case of refusal to provide personal data; the contact details of the ent ities in charge of the data processing acting as data controller and data processor.

Furthermore, the system should allow the exercise of the intervention rights that are acknowledged to the data subject by applicable privacy legislation. The data subject should be provided for example with the possibility to access his/her personal data; to ask for specific information about the processing of his/her personal data; to ask for his/her personal data to be integrated, updated, rectified, deleted, transformed in an anonymous form. The data subject should also be enabled to block the processing of his/her personal data in case of breach of applicable laws, and also to object the processing of his/her personal data for legitimate reasons.

2.1.3.10 Consent of the data subject

The system shall guarantee that, when required by applicable data protection legislation, the data subject's consent to the data processing is requested and obtained, and that the data processing is further performed according to the preferences expres sed by the data subject.

The system should also allow the data subject to revoke at any time the consent previously granted (even temporarily in case of location and traffic data processed for the performance of value added communications services).

With regard to the possibility of the data subject to change his/her preferences through the consent, the system should also be capable of properly handling circumstances such as the withdrawal of the data subject's consent or the objection by the data subject to the processing of his/her personal data.

Moreover, it is also important that the consent bears the features as described under applicable data protection legislation, notably the consent of the data subject should be free (in the sense that it should be given by the data subject without the same being forced to do so); express (that is, there should be some kind of material evidence that the data subject provided the consent); written (this usually applies to the processing of sensitive data, and it als o depends on the specific circumstance and on the applicable privacy legislation); specific (notably the consent should be provided by the data subject with regard to a specifically identified data processing activity); and informed (which implies that the data subject prior to giving his/her consent has been provided with the mandatory set of information on the applicable data processing as requested under relevant regulatory framework).

2.1.3.11 Data security measures

The system shall adopt appropriate technical a nd organizational measures with the purpose of protecting the personal data that are collected and processed by the system against the risks of accidental or unlawful destruction, accidental loss, alteration, unauthorized disclosure or access, in particular where the processing involves the transmission of data over a network, and against any other unlawful possible data processing operation or set of operations.

Taking into account the technical state of the art and the economic efforts in terms of implementation, the security measures that are applied by the system should be able to ensure an adequate level of security. The adequacy would be assessed having regard to the risks represented by the nature of the personal data to be protected, and the process ing operations to be performed.

Under some data protection national legislations there may be specific lists of mandatory security measures to be implemented; any deployment of the PRISM system subject to these laws must implement these measures.

With specific focus on the area of telecommunications services, it should be added that the security provisions are addressed not only to the service providers, but also to the network providers. In case security concerns occur in the network or for the performance of a given service, the data subject must be duly informed about said concerns.

2.1.3.12 Special data categories

The PRISM system shall guarantee that the processing of special categories of data (for example, but not limited to, traffic or other location data, sensitive and judicial data) is performed in compliance with the specific requirements that the applicable data protection legislation sets forth for said categories of data.

For example, sensitive data usually require the written consent of the data subject, and in Italy for said data the authentication credentials (e.g. the passwords to access and process sensitive data) should be replaced at least every three months, while for general personal data the expiry date of the passwords is six months. Specific provisions rule judicial data and their use is allowed only when specific circumstances occur, due to the fact that such use poses serious risks to the dignity and freedom of the individuals concerned.

In addition, retention and processing of traffic data allow the determination of behaviors, preferences, activities, and movements of the individual, and may result in invasive surveillance and profiling of the individuals.

For the processing of traffic data and location data the Directive 2002/58/EC requires that the data subject should be provided with some information that supplements the usual set of mandatory information to be given to the data subject when his/her personal data are collected. Indeed, for the processing of location and traffic data the data subject should be specifically informed with regard to the type of location and traffic data that are to be processed, the purposes of the processing (which should be very detailed and clear), the intended duration of the data processing, and (for location data) whether the data are to be transmitted to a third party for the purpose of providing the service requested by the data subject. Moreover, for the processing of traffic and location data the consent of the data subject is requested, even in case the processing is functional to performance of services required by the data subject, while in contrast the circumstance that the processing is necessary to offer to the data subject a service that the same has requested represents a general exemption from the need to obtain the data subject’s consent prior to starting the data processing activities.

The Directive 2002/58/EC also imposes specific security requirements for the processing of traffic and location data, so for example the access to said data and their processing should be restricted to persons acting under the authority of the provider of the public communications network or publicly available communications service or of the third party providing the value added service, and must be restricted to what is necessary for the purposes of providing the value added service.

Lastly, there are also limitations applying to the purposes for which said special categories of personal data may be processed. For example sensitive data usually cannot be used for activities such as profiling and building of pattern behaviors and individuals’ profiles.

It follows from the foregoing that the system must adequately implement the legal requirements above, in the sense of verification that the legal due preconditions do exist (for example the verification of the consent of the data subject).

Furthermore, the system should also implement the tighter security measures and limitations set forth by applicable data protection legislation, in terms of application of the requested security measures and compliance with the limitations imposed for the processing of the special categories of personal data (for example with regard to the limitations imposed on the purposes for which said data may be collected and processed).

2.1.3.13 Coordination with competent national Data Protection Authority

The PRISM system shall monitor compliance with the notification requirement and with the provisions on the authorizations of competent national Data Protection Authorities as ruled under applicable national data protection legislation. Moreover, the PRISM system shall allow communications between the system and the competent national Data Protection Authorities in order to validate and verify that the notification and/or authorization requirements h ave been duly complied with.

This kind of interaction with competent national Data Protection Authorities may result in a kind of alert that the system submits to the referenced Authorities in order to notify them that a certain data processing activity, which is subject to notification and/or authorization requirements, is being performed. Verification of compliance with notification and/or authorization requirements may also be considered within the negotiation process between the system and the entities asking access to the personal data stored within the system. Then it would be up to the competent national Data Protection Authority to verify accomplishments of the due legal conditions.

2.1.3.14 Supervision and sanctions

The PRISM system shall provide the competent national Data Protection Authorities with the means for supervising and controlling all actions of personal data collection and processing. This function is very important, as it often happens that the competent national Data Protection Authorities encounter difficulties in auditing the processing of personal data carried out through technical means and over the Internet, due to the peculiar nature of the technical means deployed, that allow the hiding of the data processing activities performed.

The PRISM system would not act as an enforcement authority, since it would lack the necessary competence; instead it should provide information to the competent national Data Protection Authorities, so that they can perform the necessary verifications and impose the sanctions in cases of breaches of the applicable data protection legislation.

This activity of providing of information should be structured as a communication channel, specified by an accepted technical standard or by agreement, between components of the PRISM system and the competent national Data Protection Authorities, so that the system provides the aforementioned Authorities with a log of data processing activities performed.

2.1.3.15 Communications confidentiality and lawful interception

The system shall be structured consistent with the protection of the confidentiality of communications over the monitored networks. Indeed, the European Union legislation prohibits the listening, tapping, storage or other kinds of interception or surveillance of communications and the related traffic data, unless the user has given consent and such surveillance is technically necessary to provide the data subject with the requested communication service.

The PRISM system should therefore guarantee confidentiality in the communications, but should also be able of complying with the lawful interception requests coming from the competent national public authorities. The system should support the strict legal requirements posed as preconditions for the interception. Interception is allowed only when it is necessary, appropriate and proportionate to safeguard public interests such as national security, defense, public security, and the prevention, investigation, detection and prosecution of criminal offences or of unauthorized use of electronic communications systems. Applicable national member state legislation that specifies the extent, the scope, the authorized entities, the limits and the features of the lawful interception must also be taken into account.

The PRISM system should therefore provide the competent public authorities with the means to perform interception in accordance with the applicable requirements and under the defined conditions. The necessary "hooks" for the lawful interception should under no circumstance become available to other not authorized third parties. Moreover, according to applicable legal framework, the system should allow the transmission of the relevant personal data in a robustly secure way and as requested by the legitimate addresses of the d ata communications. The personal data should usually be immediately and definitively deleted after they are communicated to the competent authorities. There may be an agreement between the system and the competent national public authorities as to the mean s of retention and communication of the personal data representing the subject matter of the interception.

2.1.3.16 Flexibility and adaptability of legal compliance provisions

Given the complexity of the legal environment in which the PRISM system operates, the different legal requirements from member state to member state, and the nature of the law to change from time to time, the PRISM system's design should to the extent possible be flexible and adaptable with respect to the provisions of section 2.1.3. Specifica lly, the system should encode as much of these provisions in policy as practicable, and not in the mechanism of the design.

2.2 Technical Requirements

Technical requirements define how the system must do what it does. These are divided into four classes. Integration requirements are in a way the most fundamental of these, as they define how the various components of the system must integrate with each other, and how those components and the system as a whole must interact with the outside world. This includes initial assumptions about the architectural arrangement of the system. Physical and logical interfaces, and the standard protocols that define them, are included here as well. Performance requirements quantitatively set the minimum performance parameters ex pected of the system. Usability requirements describe and constrain the user interface for the measurement system's end users, whether operational analysts or researchers, while Deployment and Management requirements describe and constrain the user, config uration, deployment and management interface for the measurement system's operators, including interactions with existing network management infrastructure.

2.2.1 Organization

This section defines how the system shall be organized into components, and how the functions of the system shall be divided among these components.

2.2.1.1 Separation of components

The system should be divided into three general functions: the front -end, which accepts data from the measured network; the back-end, which stores the measurement data and provides interfaces for its analysis and retrieval; and the privacy -preserving controller (or PPC), which handles any necessary cryptographic key escrow, key management, identity registry or management facilities. These three functions may be flexibly distributed into multiple components as required by the measurement application and deployment environment.

2.2.1.2 Separation of trust

The functional requirements above must be met without requiring the front -end to trust the back-end. Specifically, the system mu st be able to provide the required measurement functionality without the back-end having access to unprotected measurement data. This implies that the channel between the front-end and back-end must be encrypted or anonymised as required by the specific ap plication (see Situational Requirements, below), and that the back-end must be capable of storing and operating on encrypted or anonymised measurement data.

2.2.1.3 Front-end functions and component organization

The front-end shall be organized into three primary sets of components.

- A packet capture component in charge of capturing the packets from the monitored link and conveying them to the proper processing components.
- A set of processing components in charge of processing and protecting the captured data and of conveying it to the back-end in a standard format.
- Control-plane components to coordinate the actions of the packet capture and processing components and their interaction with the back -end and PPC components.

2.2.1.4 Back-end functions and component organizatio n

The back-end tier must be organized into a variety of functional components, which include:

- Input components, devised for receiving traffic data from the fro nt-end. These components must accept raw packet data, and flow data in IPFIX format, as protected (i.e. encrypted or anonymised) by the front -end, and recognize and suitably treat the protection enforced on the data.
- Output components, in charge of providing the monitoring applications with the data that are necessary for their operation. These compon ents must export raw packet data, and flow data in IPFIX format.
- Control-plane components for exchanging control information with both the front -end and the Privacy-Preserving Controller.
- Components for the enforcement of the policies that result on an ad hoc basis, after the evaluation of the "privacy context", the semantic model of the personal data protection regulations (as specified in the PRISM ontology) and possibly of the user-defined privacy preferences, when available and applicable.
- Components in charge of managing the storage of data, including but not limited to a distributed database system.
- Processing modules, for functions including but not limited to : i) reversing –when certain conditions meet– the data protection set forth to some data by the front-end; ii) protection (e.g. anonymisation) of data prior to disclosure them to an external monitoring application; iii) internal execution of privacy-sensitive functions of monitoring applications, thus eliminating the need of disclosing data for their execution; and iv) execution of tasks imposed by the personal data protection legislation, such as the information of the data subjects regarding some data disclosure or processing event.

2.2.1.5 Execution of complementary actions

The system shall support the triggering of actions based upon access control decisions in order to meet privacy and regulatory requirements, for example but not limited to notifications of users or relevant authorities, explicit consent requests of data subjects, enforcement of data retention periods, and so on.

2.2.1.6 Privacy-Preserving Controller responsibilities

The privacy-preserving controller acts as a central point of trust within the system. It is the authority for any federated identity/certificate system (e.g. X.509), and the po int of escrow for each cryptographic protocol in use for the protection of data or inter -component communications. The privacy-preserving controller may also be used as a central point of trust for other functions as appropriate for a given deployment.

2.2.1.7 Multiplicity of components

Each front-end must be capable of sending data to multiple back-ends simultaneously. Each back-end must be capable of accepting data from multiple front-ends simultaneously. Each privacy-preserving controller must be able to provide services to multiple front-ends and back-ends simultaneously. Multiplicity of front-ends allows distribution of monitoring to multiple observation points and scalability of monitoring to very large networks. Multiplicity of back-ends allows network locality of processing and the potential for load balancing of storage and analysis operations. It is not necessary for a front-end or back-end to support interfacing to more than one privacy-preserving controller.

2.2.1.8 Commodity hardware implementation

To the extent possible, the project should aim to provide software that implements each component without the use of specialized hardware for demonstration purposes and for operation on smaller networks. Such an implementation need not meet the performance requirements in this document.

2.2.2 Integration

This section defines how the system shall interact with external components and systems, and how components within the system shall interact with each other in externally interoperable ways.

2.2.2.1 Control plane and data plane separation

The control plane (transporting configuration, command, and diagnostic information among components) shall be separated from the data plane (transporting protected measurement information) to the extent possible.

2.2.2.2 Control Plane security

Control plane communications among components of the system or between external entities and components of the system must be secured by standard cryptographic protocols such as TLS, SSH, or IPsec. Any component that accepts control plane information must be configured out of band to accept such information only from pre-determined entities identified by the chosen cryptographic credentials. It must not be possible to modify a component's control plane trust configuration from within the PRISM system.

2.2.2.3 Data Plane security

Data plane communications among components of the system must be securable by standard cryptographic protocols such as TLS, SSH, or IPsec, as appropriate to meet performance requirements.

2.2.2.4 Unidirectional data plane communication

Data plane communication between components should be unidirectional to the extent feasible. The integrity of unidirectional communications can be realized by using appropriate forward error correction techniques. This improves the total resistance of the system to attack from downstream sources.

2.2.2.5 Standards-based flow export and storage (IPFIX)

Any component of the system that exports flow data must use IPFIX as specified in [RFC5101], and as extended by the IPFIX Working Group to support anonymisation [IPFIX-ANON]. Any component of the system that accepts flow data must support IPFIX collection. This implies that transmission of flow data between components must use IPFIX as well; this meets the Transport Security and Unidirectional Data Transfer requirements above. Storage of flow data in files that may be used outside the system must be stored in IPFIX Files as described in [IPFIX-FILE].

2.2.2.6 Standards-based control plane

Control plane and directory information shall be transmitted using existing standard protocols, where possible.

2.2.2.7 Efficiency of internal protocols

All communication among or within system components that is not subject to potential interoperability with components outside the system should use communications protocols or interfaces selected for appropriateness to the communication with a focus on low overhead, low latency, and high throughput.

2.2.3 Performance

This section defines the performance requirements for the system, and technical requirements related to ensuring adequate performance of each component.

2.2.3.1 Link speed

Front-end components must be able to process data from a single link operating at one gigabit per second with negligible packet loss. Should front-end components be deployed to measure data from multiple links, the aggregate bandwidth under measurement may be less than one gigabit per second.

2.2.3.2 Back-end scalability

Back-end components must be able to process data from at least one front-end running at full capacity. The back-end should able to cope with the possible increased needs of the system in terms of data storage capacity and heavy computational load. Therefore, at the database layer, the back-end should rely on a distributed database system that will provide the necessary level of transparency and abstraction regarding the physical storage of data. On the other hand, at the application layer, the back-end must be able to be deployed at several platforms, including large clusters, thus rendering its own structure and physical distribution transparent to the other PRISM components, especially the front-end. Moreover, the back-end should be able to be easily extended with the seamless addition of computational resources (e.g., with the on-the-fly addition of computer systems to a cluster).

2.2.3.3 Front-end packet capture performance

To ensure that packet capture performance targets are met, the following requirements apply to the packet capture component within the front-end.

- This component must adopt an efficient classification algorithm, involving a small memory footprint (in order to store the associated data structure into a small and fast memory bank). Adoption of inspection-based packet classification is discouraged, because of the high number of memory access involved. Indeed, in order to look in a packet a given string set, a particular data structure (generally implementing a finite state automaton) has to be checked for each byte of the payload, thus involving a large number of memory accesses; in addition, the whole packet payload has to be read from an high capacity memory bank, thus introducing further overhe ad.

- Since a classified flow can be associated to more than processing component, two kinds of issues arise:
  - Depending on the network processor architecture, sending the same captured data to two different destination may involve copying large memory areas; this, in turn, could lead to a considerably increased number of memory accesses which could significantly affect the system performance, even by resulting in a lower achievable bit rate. Such a possibility has to be carefully taken into account while designing these components.
  - Due to the necessity of sending the same portion of captured data to several destinations, the rate of the data flow produced by the packet capture component could even exceed the rate of the captured flow. As a consequence, a sufficiently fast communication channel between the packet capture component and processing components of the front -end has to be deployed.
- Anonymisation and cryptographic protection algorithms must be designed in order to present a limited number of accesses t o the packet payload and a small memory footprint. Their computational requirements have to be accurately accounted for while considering the delay budget.

2.2.3.4 Front-end classification, protection, and coordination performance

As it is obvious, the precise functional requirements of each processing component are strictly dependent from the kind of processing requested by the particular kind of protection technique. However, it is useful to point out that, since also these processing tasks have to be executed in real time, a delay budget has to be calculated based on an estimate of the rate of the traffic to be processed. Based on such a requirement, a particular processing platform can be chosen.

We point out that, since the processing unit is only a functional abstraction, it can be actually implemented by several physical devices working in parallel, each of them processing a portion of the total traffic flow.

Apart from the requirements imposed by the particular processing task, each physical device of a processing unit must be able to receive with no losses the data flow generated by the packet capture component. If the communication is implemented by using a standard network card, severe constraints must be considered in terms of acceptable traffic rate. As a benchmark, we can consider that a standard pc equipped with a network card can drop more than one half of the received packets, if the packet rate exceeds 400000 packets per second. This is mostly due to the high latencies involved with the interrupt bas ed mechanism which is in charge of packet reception on a PC like architecture. Furthermore, such a performance is measured in case a PC is used only for traffic sniffing, while, in case of a PC used as a processing unit in the PRISM front -end, a considerable amount of data should be sent to the back-end component at the same time, thus leading to a further reduction in terms of packet rate.

As for the implementation of the IPFIX protocol, it should not raise any particular issue in terms of performance. However, since such a protocol involves the dynamic definition of record formats, certain degree of programming flexibility is needed for a processing unit. If a special purpose network processing platform is adopted, this can arise several implementation issues, since programming such devices often involves using low level instructions and taking care of several hardware related details; using extensible message formats, as allowed by the IPFIX standard, could considerably increase the complexity of the proje ct implementation.

2.2.3.5 Distributed storage of trace data

The storage of trace data constitutes a fundamental functionality of the back -end tier. In order to cope with performance issues, as well as to logically and physically separate data for improved security and privacy protection, the system's database should be distributed. Additionally, data distribution should be accomplished following privacy -aware mechanisms.

2.2.4 Usability, Deployability, and Manageability

This section defines the requirements for the interface the system presents to its users and administrators.

2.2.4.1 Front-end configuration

The front-end packet capture component must be configurable on -line to change the classification rules. Front-end processing components must be configurable on -line to select processing actions for each class of captured packets; this includes not only demultiplexing configuration but configuration of processing parameters as appropriate to each component. Configuration of the front-end is subject to prior authentication of the configuring user, and verification that the user possesses the necessary authorization for the given configuration action.

2.2.4.2 Back-end configuration

The back-end tier controls access to collected data; therefore, the fundamental back-end configuration, therefore, consists of the access control rules in force. Access control rules enforced by the back-end component must be configurable to grant access to specific types of data to specified roles and processing purposes; the back -end must not allow the configuration of rules which violate the essential privacy, legal, or regulatory requirements enumerated in this document. Configuration of the back-end is subject to prior authentication of the configuring user, and verification that the user possesses the necessary authorization for the given configuration action.
These rules will be originating by the personal data protection legislation and will be expressed by means of a semantic model. This semantic model will specify not only the access level of applications and other entities to the data, but also the data processing/analysis tasks executed by the back-end.

2.2.4.3 Distribution of configuration

All components of the system must be configurable to accept control plane information only from specified entities. This arrangement of the control plane must be done out of band in order to ensure the security of the system. Note this requirement implies that the system cannot be rearranged dynamically.

2.2.4.4 Monitoring application programming interfaces

Each component of the system must provide an application programming interface for the development of new monitoring applications distributed between the front -end and back-end. This includes the development of packet processing components, back -end measurement logic, and configuration logic for on-line configuration of the same.

2.2.4.5 Adaptation of existing monitoring applications

The system shall, to the extent possible, provide a means for adapting existing monitoring applications to operate "on top of the system", to use the syste m as a source of trace or flow data where appropriate.

2.2.4.6 Self-monitoring

Each system component shall, as necessary, include self-monitoring functions to protect its own operational stability against overload. Continuous monitoring of network interfaces' traffic, CPU load and disk space can make the system aware of available resources and in effect disallow new monitoring and analysis tasks in order to not disrupt already active tasks, especially online analysis tasks.

2.2.4.7 Diagnostic logging

Each component of the system must provide a detailed logging facility for the diagnosis of problems existing within each component and within the communication links between components.

2.2.4.8 Audit logging

The system must provide a detailed logging facility to store timestamps, user identity, and action performed for all data analysis actions, retrievals of data from the back -end, and control plane communications, in order to provide non -repudiation of identify attempted or actual compromises of the privacy protections the system pro vides.

2.2.4.9 Data subject access and preferences

As the data subjects (i.e., the users of the network under measurement) have the right to be informed regarding the collection or processing of personal data, to be asked about their explicit consent, and to access their data, the system should provide an interface for data subjects to exercise these rights. Additionally, data subjects they should be able to specify their privacy preferences with respect to the data collected.

2.2.4.10 Data Protection Authority access

As competent Personal Data Protection Authorities have certain rights and responsibilities with respect to the collection and retention of identifiable data, including the notification of the Authority of certain events, the supervision of data collection and p rocessing procedures, and the means for performing lawful i nterception. The system should provide an interface for the competent Authorities to perform the required tasks .

2.3 Requirements Summary

This table summarizes each requirement, and the subsection(s) o f section 2.1 and section 2.2 from which they are drawn. It is intended as a non -normative quick reference for further development of the architecture and implementation.

Table 1: Requirements for PRISM architecture development and implementation

Requirement	*Section(s)*	*Cpt.*	*Prio.*
Support monitoring applications consuming packet, flow, and aggregate flow data.	2.1.1.5		req
Provide a separation of trust between system components, allowing each component to trust others only as much a s necessary.	2.2.1.2		req
Capture packets at 1Gbps and classify them based on flow key and flow or packet properties; extract information from captured packets; protect extracted information through anonymisation and encryption.	2.1.1.1, 2.1.1.2, 2.1.2.5, 2.2.3.1, 2.2.3.3, 2.2.3.4	FE	req

Provide separate packet capture and packet processing components; allow flexible configuration of the connections between these components and the composition of processing components to implement monitoring processing tasks.	2.2.1.3, 2.2.4.1	FE	req
Collect information extracted from packets as quickly as it is received, protecting collected information via privacy-aware semantic access control.	2.1.1.3, 2.1.2.6, 2.1.3.8, 2.2.3.2, 2.2.3.5	BE	req
Provide separate input, output, control plane, policy enforcement, storage management, and processing components in the back-end; allow the composition of processing components predicated on access control decisions to implement monitoring processing tasks.	2.2.1.4, 2.2.1.5, 2.2.4.2	BE	req
Provide a component to act as a center of trust for cryptographic certificate management and escrow.	2.2.1.6	PPC	req
Provide for multiplicity of connections between front-ends and back-ends for flexibility and scalability.	2.2.1.7		req
Support measurement of IPv4 networks.	2.1.1.4		req
Support measurement of IPv6 networks.	2.1.1.4		opt

Requirement	*Section(s)*	*Cpt.*	*Prio.*
Ensure network end users are identifiable only when necessary, protecting their personal data from accidental or unauthorized intentional disclosure. Enforce this assurance within each component of the system, by limiting each component's access to information.	2.1.2.1–2.1.2.4, 2.1.3.5, 2.1.3.6,		req
Adopt best security practices in design and deployment.	2.1.3.11		req
Provide robust anonymisation mechanisms for the protection of personal data within the system and for publication purposes.	2.1.2.5		req
Ensure all monitoring activities performed are lawful and in compliance with the principles of purpose and proportionality.	2.1.3.1–2.1.3.3, 2.1.3.12		req
Ensure the quality and accuracy of the data used to perform monitoring activities.	2.1.3.4		req
Retain personal information only as long as is necessary to perform legitimate monitoring activities, or as required by data retention regulations.	2.1.3.6, 2.1.3.7		req
Maintain logs of all access to data to defend against unauthorized access to personal data.	2.1.3.8, 2.2.4.8		req
Provide interfaces to allow data subjects to exercise their rights with respect to their personal data, to grant or withdraw consent for the use of personal data, and to inform data subjects of the personal data used for monitoring activities.	2.1.3.9, 2.1.3.10, 2.2.4.9		req
Provide interfaces to allow access to the system by competent national Data Protection Authorities as required in the exercise of their authority.	2.1.3.13, 2.1.3.14, 2.2.4.10		req
Protect the confidentiality of measured communications,	2.1.3.15		req

but provide support for interception activities as required by the law.			
Allow flexible configuration of all aspects of the system with respect to legal and regulatory compliance and access control, to adapt to changing monitoring problems and regulatory situations.	2.1.3.16, 2.2.4.1, 2.2.4.2		req
Logically separate the control plane among components from the data plane.	2.2.2.1		req
Provide standard, configurable cryptographic security for control plane and data plane communications.	2.2.2.2, 2.2.2.3		req

Requirement	*Section(s)*	*Cpt.*	*Prio.*
Transfer flow data among system components using IPFIX.	2.2.2.4, 2.2.2.5		req
Utilize existing standards for control-plane communications.	2.2.2.6		opt
Require out-of-band configuration of trust for control-plane communications.	2.2.4.3		req
Provide application programming interfaces for the development of monitoring applications as collections of PRISM system components	2.2.4.4		req
Provide interfaces for the export of data from the system in order to integrate with existing monitoring applications	2.2.5.5		opt
Provide facilities for self-monitoring and diagnostic logging on each component.	2.2.4.6, 2.2.4.7		req
Provide a software-only implementation of the system on commodity hardware for demonstration purposes, or deployments on small networks.	2.2.1.8	FE	opt

3 Scenarios

This section examines a variety of real world network monitoring and measurement scenarios drawn from the experience and research of the project partners. These scenarios first briefly cover the current state in each situation, and then select a future potential measurement application to be enabled by PRISM, then present a hypothetical use case for the given application.

3.1 Regional Wireless Internet Service Provider

In this scenario we consider the case of a small regional wireless internet service provider (WISP) that provides flat-rate, bandwidth-guaranteed Internet connectivity and in addition standard ISP end-user services such as email, web hosting, and network storage; and experimental voice-over-IP (VoIP) services. End users are generally connected to the network via wireless subscriber units (SUs) or via WiFi hotspots. In all cases, these last-mile links are backhauled to the WISP's core network via either point-to-point radio links or externally provided optical fibre links. This core network, in turn, is connected to the Internet by means of an upstream transit provider. Management information is sent from the managed devices back to the WISP's data centre in-band, over the core network.

This WISP presently deploys a variety of network monitoring applications including an SNMP infrastructure for basic operational network monitoring (e.g., device status, service uptime, and so on). Its flat rate structure means it does not need to deploy per-customer data volume measurement for billing purposes. In this scenario, we extend this largely diagnostic system to include network intrusion detection while preserving the privacy of the end users.

3.1.1 SNMP Measurement Applications

The WISP's present monitoring infrastructure is largely based upon SNMP. There are five basic classes of infrastructure devices within the network: switches, routers, base stations, subscriber units, and data centre services. Of these, only subscriber units are not generally monitored, as older installed subscribed units do not support SNMP management. All others send measurement information back to a central monitoring platform based on OpenNMS located in the data centre. Access to measurement data at the OpenNMS server is protected by group privilege levels and user authentication; therefore, each network technician has access only to measurement data necessary for his or her job function. The WISP collects per-customer statistics (connection time, connection duration, and inbound and outbound octet values), and status information for all links and all services provided in the data centre for monitoring purposes. Note that user location information can be deduced from these records, because each access point or subscriber unit has a known location, and this information is available in the per-customer statistics.

3.1.2 Packet Capture for Debugging

In addition to the SNMP infrastructure, the WISP deploys diagnostic equipment for on-demand packet capture from time to time to diagnose and correct issues within the network. Due to the potential privacy impact of such activities, policies ensure that these are only deployed with the full knowledge of the company's management.

3.1.3 Lawful Intercept of VoIP calls

The WISP provides VoIP services to its customers, with a connection to the public switched telephone network. The VoIP platform features a remotely controllable Interception Server capable of intercepting any call handled by the VoIP service, in order to comply with lawful interception requirements. This is an existing system separated from the rest of the WISP's measurement infrastructure, provided by an external vendor and designed to “drop in” to a VoIP deployment.

3.1.4 Operational Monitoring Scenario

A hypothetical PRISM deployment within this WISP would extend the “debugging” on - demand packet capture facility with an online packet capture, filtering, aggregation, and analysis facility in order to continually monitor the traffic stream at the border. Such a facility has a wide variety of uses; here, we will consider a single example, generating five -minute time series of internal and external unique host and network address counts. Such an analysis can be useful as an indicator of scanning or denial of service activity, an input to a variety of live intrusion detection algorithms, or detection of applications with capacity planning implications, such as peer-to-peer file sharing.

The PRISM front-end, deployed at the network border, would generate packet, flow, or pseudoflow data from the packet stream, protecting IP address information using a reversible, prefix-preserving way. Key material used for reversal of this information is kept on the PRISM a privacy-preserving controller, located within the WISP’s infrastructure but under the strict control of a limited set of administrators. Packet header fields not required for the specified purpose (e.g., port numbers for the address count analysis) are either protected or simply not exported from the front-end, depending on configuration (i.e., whether the front-end might be used for other applications).

The PRISM back-end, deployed in the core network where the present OpenNMS system is located, would then collect information from the front-end and perform unique address counting. A network engineer wishing to retrieve a five minute time -series of host counts would then log into the PRISM system, which would verify that the engineer had sufficient privileges to view that particular data set, derived from that front-end, for the specified purpose. If the front-end had not been configured to collect this data continuously (e.g., for retrospective analysis purposes), then the privilege check would also verify that the engineer had sufficient privileges to configure and start collection on the front-end. Assuming the engineer had sufficient privileges, the back-end would provide the requested data via a graphical user interface or some specified output format to an external analysis tool. Note in this example that not only would IP address information not be presented to the engineer, but the back-end itself would not need the actual IP addresses, and would not have access to such information without the cooperation of the privacy-preserving controller.

Note that in this scenario, PRISM could be deployed alongside the existing VoIP Interception Server without interfering with its operation.

3.2 Mobile Telephony Operator

Third generation (3G) mobile telecommunications operators provide cellular wireless network access to their end users over a wide geographic area, typically covering the populated areas of an entire region or nation. Compared to the standard fixed-line Internet, the wireless links and the inherent mobility of end-user devices necessitate the support of inter-cell handovers, which makes mobile 3G networks far more complex. In mobile networks, additional overhead in planning and operations is also added by the fact that the end-user equipment is often battery powered, making the power consumption an important factor. Furthermore, due to the shared medium character of the radio access bandwidth is also scarcer than in fixed IP networks, which is a constant issue for operators with the growing popularity of mobile Internet services.

Note that while cellular networks normally also support voice communication, here we only consider the mobile data network part, which consists of the Radio Access Network (RAN) and the mobile packet core network. Mobile data networks are best described by looking at the different stages of traffic aggregation from the mobile terminal up to the open Internet egress point. Due to the already mentioned high complexity of 3G data networks, the traffic aggregation tree in 3G has many more stages than the fixed-line Internet, as the traffic needs to cross a much larger number of technologically divergent network stages: the traffic concentration starts in the Radio Access Network (RAN) which is comprised of base stations connected to radio network controllers (RNCs), followed by the mobile packet (fixed -line) core network which carries the traffic from the different RNCs via 3G specific enhanced router nodes up to the GGSN (General GPRS Support Node) as the topmost point of aggregation. Finally, the GGSN provides for the interconnection with the open Internet. It is important to note that while the Radio Access Network (RAN) infrastructure is necessari ly widely distributed due to the need for physical signal coverage of a large geographical area, in most network operator implementations the packet core network elements are located in a small number of physical locations, as with fixed -line ISPs.

3.2.1 Monitoring in 3G Mobile Networks

Traffic monitoring plays an essential role in mobile 3G networks. For example, operators clearly have the need to protect their networks and customers from various kinds of external threats; for example, port scanning attacks which can consume large amounts of scarce bandwidth as well as compromise end-user devices and rapidly deplete their batteries. Additionally, there is a lot of 3G network-specific information to be monitored, such as PDP (Packet Data Protocol) context establishments and traffic generated due to inter-cell handovers of mobile users. Monitoring such events can reveal emerging capacity bottlenecks, incorrectly configured devices or services, and other such situations requiring immediate actions from the operator.

End-user billing is becoming an increasingly important area of application for traffic measurements in mobile networks. With most European operators, the billing of data services is more complex than in fixed Internet access networks. First, due to scarcity of bandwidth, the tariffs are often based upon traffic volume per user instead of a flat rate fee. Second, there are exceptions to volume-based billing, such as for messages to end-users from the operators themselves, which are normally free of charge. This requires mobile operators to collect per-user connection and traffic volume statistics, and potentially other information, which makes a very strong case for the close coupling of billing and network measurement platforms.

3.2.2 Billing Non-Repudiation Scenario

A concrete example in which network measurement systems can offer unique value to the operators is the non-repudiation of data volumes attributable to each individual customer, as with the availability of the recorded data streams the operator can easily demonstrate the exact volumes of traffic generated by each user. Technically this can be performed based upon the International Mobile Subscriber Identity (IMSI) of the users, which can be extracted from the stored traffic traces. However, such identifiers are extremely sensitive and should be considered confidential information which must not be exposed unless mandated by the situation.

The PRISM architecture can be applied to this problem, by collecting traffic information and protecting sensitive identifiers such as the IMSI at the front-end. As before, the front-end passes connection records to the back-end with all identifying information protected by a key kept with the privacy preserving controller, and the back-end does not store or have access to the IMSI information for any subscriber. In the event of a repudiation request, the IMSI of the customer in question, which the analyst at the operator will have explicit permission to use, is projected into the protected information space and compared against the information stored in the back-end to search for that customer’s records alone. Indeed, such an arrangement could even allow for the safe separation of non-repudiation services into a separate administrative domain.

3.3 Community Network

Community networks (e.g., Germany's freifunk.net[6], Funkfeuer[7] in Austria), in contrast to networks run by internet service providers or mobile telephony operators, are characterized by the absence of a single entity that owns and manages the entire network infrastru cture. Instead, every user (or member) of the community network autonomously owns its own infrastructure (e.g., the wireless router) and is responsible for maintaining its connection to the network. Volunteers perform network-wide maintenance and managemen t functions. Community networks are generally built using wireless links in areas of high population density, or in developing countries with limited existing network infrastructure (e.g., Pretoria Wireless Project[8], Fantsuam Foundation[9])

The primary techn ical difference between community wireless networks and WISPs is topological; WISP topologies are generally hierarchical, while community networks are generally flat, with few nodes aggregating traffic for other nodes unless necessary for the scale of the network. This gives interesting opportunities for monitoring as well as rise to privacy concerns at the same time.

In Europe, the primary service community networks offer to their membership is basic Internet gateway connectivity; however, some networks ar e beginning to offer other ICT services such as email, web hosting, and so on. Since community networks are maturing as a technology and service model, they often serve two distinct constituencies: end users, who use the network for Internet connectivity a nd ICT services; and researchers, who often use the community network as a testbed for wireless mesh routing algorithms and other techniques in community network building.

As community networks mature, the services offered evolve as well, and this has impl ications for monitoring in these networks. For example, the emerging availability of inexpensive VoIP hardware is leading to its deployment on community networks, with an eventual aim toward replacement of public switched telephone network services. This i ncreases the usefulness of the network but leads to new challenges, as VoIP applications require more bandwidth and less latency than e.g. web browsing and e -mail. This consequently requires improved traffic engineering that can only be achieved through pr oper monitoring.

In this scenario we examine the current limited state of network monitoring in a given community network, and apply a hypothetical PRISM deployment to the problems both of peer-to-peer application detection and publication of trace inform ation for research purposes.

3.3.1 Measurement Applications in Use

Current monitoring of community networks is limited, consisting mainly of monitoring aggregated bandwidth consumption, node uptime, and link reliability. Bandwidth is monitored both for abuse pre vention and measuring resource consumption in order to cope with long-term traffic growth trends. Node uptime and link reliability are monitored for both operational and research purposes. As community networks are often used as testbeds for wireless mesh network routing algorithms, specific measurement tools have been developed for this purpose.

Since community network members are a mix of end-users and researchers, the measurement needs of community networks can be divided into two distinct activities: op erational network management, which is often focused on understanding the network's behavior and resource usage at a single node in order to best tune that node to participate in the network; and research, which is focused on collecting data from multiple nodes in order to build a coherent picture of the state of the entire community network.

[6] http://start.freifunk.net/

[7] http://www.funkfeuer.at/

[8] http://www.pwp.za.net/

[9] http://www.fantusam.org/ictprojects.html

3.3.2 Privacy Policies and Practices

Given their decentralized nature, specific privacy policies and practices are generally not enacted in the monitoring of community network infrastructure. Each node has full access only to the packet data crossing that node; those nodes that double as access points to the public Internet would in theory have a more complete view of the traffic crossing the border, but this view would not be comprehensive as such networks generally have multiple Internet access points. The limited nature of monitoring in community networks generally doesn't raise many privacy concerns, limited as it is to a high-level view of total data volume and topography changes, and by the fact that the inexpensive, small-scale nature of the hardware used to build the network is not capable of processing and storing the large amount of data generated by detailed monitoring activities. However, community network users are often more technically inclined and as such tend to pay more attention to privacy issues.

3.3.3 Application Detection Scenario

As bandwidth resources are scarce in community wireless networks, high-bandwidth applications such as peer-to-peer (P2P) filesharing are one of the main causes of poor performance in community networks, and their usage is often limited by consensual agreement by community network members. It is therefore desirable to detect P2P activity in order to verify compliance with this agreement, without impacting the privacy of the general user community.

In this scenario, relatively small PRISM front-ends are deployed at a variety of nodes to generate flow data, again reversibly anonymising any identifying information in the flows, as well as the identification of the node at which the traffic was collected. The collected flow traffic is then sent to several relatively small back-ends, and analysed to search for the behavioral fingerprint of the unwanted P2P activity. Detections of P2P applications could then be published to the entire community, allowing only the owner of a given sensor or node to verify whether this activity is occurring on his node (e.g., in the case of a P2P system being installed by malware).

Note that the ability for the front-end to send traffic information to the back-end without necessarily trusting it, by encrypting the traffic, is key to the adoption of the system within a community network. Also important is the many-to-many relationship among front-ends and back-ends, as it allows decentralized ownership and management of the measurement infrastructure, and the composition of the measurement infrastructure from inexpensive components, mirroring the decentralized ownership and management of the network itself.

3.3.4 Trace Publication Scenario

Community networks, as has been mentioned, are often used as testbeds for research in wireless mesh routing protocols. It is useful to be able to take trace data from a given network over a period of time to observe the effect of changes in topology and routing protocol parameters on the flows of traffic in the network. It can further be useful to publish these traces so other researchers can use them in comparative simulations of new developments in mesh routing protocols.

In this case, the data can be much more heavily anonymised than would be possible for detection purposes, and anonymised in a non-reversible way. For example, address spaces can be flattened, layer 4 port information can be removed, timing information can be shifted randomly, and noise can be added to flow volumes, and so on. To anonymise data for publication, a researcher would query the relevant data from a PRISM back-end, with the back-end verifying that the given operator had the necessary privileges to anonymise data for publication.

In this case, the PRISM back-end would keep a log of which data was anonymised and how, as multiple different anonymisations of the same data set can be used to regenerate the original data. This logging would be used to warn the researche r of the danger of deanonymisation, or even to deny publication in case of such danger.

4 Conclusions

This deliverable specified the requirements for the PRISM system architecture and a pilot implementation thereof, organized into a defined classification of requirements and informed by an examination of hypothetical scenarios of the deployment of this system. The specified system is a general-purpose network traffic monitoring and measurement system that provides strong protection for personal data in a wide variety of monitoring and measurement applications.

The architecture that emerges from these requirements has several features unique in network traffic measurement systems. First is separation of trust between network observing and traffic analysis components, allowing the storage and analysis of data without allowing the storage and analysis components full access to the data; this is achieved through novel cryptographic data protection algorithms. Second is semantic access control, which provides access to information at each step of an analysis based upon the wider "privacy context" each request for information occurs in. Third is the treatment of the applicable legal and regulatory environment for monitoring, not just during the design of the system bu t at runtime as well, by predicating access control decisions in part on the provisions of the law. While the core system is in effect a framework, providing a set of services from which privacy -aware monitoring applications can be built, the project will also adapt selected monitoring applications to operate in concert with this core.

The specified requirements also describe a set of minimum targets for any current network measurement system intended for deployment at the scale targeted by the PRISM projec t, including an ability to measure traffic at currently common network border data rates (1 Gbps), and compliance with the latest standards relevant to specific technical requirements (e.g. IPFIX).

From here, the project moves on in several aligned work pa ckages toward its goals. The pilot implementation will require the detailed specification of an architecture in line with these requirements, the design and implementation of the components of the system, and the selection of monitoring application areas f or the pilot implementation. In terms of research the project will devise and implement the novel cryptosystems and data protection algorithms required for separation of trust, advance the state of the art in semantic access control, and apply new technolo gies in packet capture and processing to distribute as much of the work of data protection and analysis as close to the measurement edge as possible.

We envision the final product of the effort following from these requirements to be a core system and a se t of measurement tools built upon it, applicable in a wide variety of measurement scenarios, and a pilot implementation deployed and in daily use at an SME in support of their network monitoring needs as well as the privacy rights of their network's users.

References

[RFC5101] B. Claise, S. Bryant, S. Leinen, T. Dietz, B.Trammell. *Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information.* IETF Request for Comments 5101, January 2008.

[IPFIX-FILE] B. Trammell, E. Boschi, L. Mark, T. Zseby, A. Wagner. *An IPFIX-Based File Format.* IETF IPFIX Working Group work in progress, draft-ietf-ipfix-file-02.txt, July 2008.

[IPFIX-ANON] E. Boschi, B. Trammell. *IP Flow Anonymisation Support*. IETF IPFIX individual submission work in progress, draft-boschi-ipfix-anon-01.txt, July 2008.