Privacy-Preserving Network Monitoring: Challenges and Solutions

Giuseppe BIANCHI[1], Elisa BOSCHI[2], Francesca GAUDINO[3], Lefteris KOUTSOLOUKAS[4], George LIOUDAKIS[4], Sathya RAO[5], Fabio RICCIATO[6], Carsten SCHMOLL[7], Felix STROHMEIER[8]

[1]*CNIT Research Unit Roma Tor Vergata, Via del Politecnico 1, Roma, 00133, ITALY*
Tel: +39.06.7259.7450, Fax: +39.06.7259.7435, Email: giuseppe.bianchi@uniroma2.it

[2]*Hitachi Europe Sophia Antipolis Lab, Route des Dolines 1503, Valbonne, 06560, FRANCE*
Tel: +33.489874100, Fax: +33. 489874199, Email: elisa.boschi@hitachi-eu.com

[3]*Baker & McKenzie, Piazza Meda 3, Milan, 20121, ITALY*
Tel: +39.02.76.231.1, Fax: +39.02.76.231.501, Email: francesca.gaudino@bakernet.com

[4]*National Technical University of Athens, Heroon Polytechniou 9, Athens, 15773, GREECE*
Tel: +30 210 7722423, Fax:+ 30 210 7721092, Email: {gelioud,lefterisk}@icbnet.ntua.gr

[5]*Forschungszentrum Telekommunikation Wien, Donau City Strasse 1, Wien, 1220, AUSTRIA*
Tel:+43.1.5052830.0, fax:+43.1.5052830.99, Email: ricciato@ftw.at

[6]*Telscom AG, Sandrainstr. 17, 3007 Bern, SWITZERLAND*
Tel, 41.31.3762033, Fax: +41.31.3762031, Email: rao@telscom.ch

[7]*Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, GERMANY*
Tel: +49.30.3463.7136, Fax: +49.30.3463.8136, Email: carsten.schmoll@fokus.fraunhofer.de

[8]*Salzburg Research Forschungsgesellschaft m.b.h, Jakob Haringer Strasse, Salzburg, 5020, AUSTRIA*
Tel: +43.662.2288.441, Fax: +43.662.2288.2, Email: felix.strohmeier@salzburgresearch.at

Abstract: Traffic monitoring is a necessary activity for the operation, maintenance and control of communication networks. On the other hand, traffic monitoring has important implications on the user privacy. This paper discusses a novel approach to privacy-preserving traffic monitoring and the related research challenges. This study is based on the analysis of the technical implications and regulatory provisions in the areas of data protection and security. We propose a two-tier privacy-preserving monitoring architecture characterized by i) the adaptation and operation of monitoring applications on protected data, and ii) a back-end middleware system devised to control and orchestrate the access to and processing of the collected data. The feasibility of the proposed architecture faces many challenges. However, by carefully designing data protection mechanisms, by adapting monitoring applications, and by deploying a semantic-rich privacy-aware access control framework, it is possible to concurrently meet strong privacy requirements, achieve efficient solutions for traffic monitoring, and be compliant with the recent regulatory provisions such as data retention.

Keywords: privacy, network monitoring, privacy legislation, access control, semantic model.

1. Introduction

Traffic monitoring is required to support the operation and management of the network infrastructure, to spot network anomalies, to enable troubleshooting, and to defend the network and its users from security threats like attacks, intrusions, and infections. Traffic monitoring is also required to provide legal authorities with a reliable data/call log for investigations and for trace-back purposes on criminal activities.

The other side of the coin is that network monitoring poses obvious concerns in terms of user privacy. Even when data capture is restricted to the header part of the packets or limited to the flow/call signalling information, a large amount of personal information may still be gathered from there (e.g. who is connecting to whom or to which servers, which applications are used, when the user is connecting and at which frequency). Even worse, modern traffic analysis and classification techniques are extremely powerful in extracting potentially sensitive information from as little as basic flow statistics such as packet sizes and inter-arrival times correlation (see e.g. [1,2,3]), characteristics which are not obfuscated by per-packet encryption mechanisms at all.

Countermeasures such as packet anonymization techniques [4,5,6,7] do not provide a comprehensive answer. If they are "too strong", they may cancel/obfuscate information that is essential for the operation of legitimate monitoring applications (e.g. network troubleshooting). On the contrary, if they are "too light" the user privacy remains at the stake of statistical analysis approaches and crypto-attacks.

This paper attempts to break the dichotomy between the acknowledged need for network monitoring on one hand and the right for user privacy and data protection on the other. We propose a single solution for both effective network monitoring and privacy preservation. This can be accomplished through the specification of a two-tier monitoring infrastructure which involves i) data protection mechanisms directly performed on traffic flows during packet capture; ii) adaptation of monitoring applications to operate on protected (encrypted) data, and iii) development of a comprehensive semantic middleware which regulates the access and processing of the data according to context information (data and user types, monitoring application purposes, etc.) in compliance with regulatory provisions. The arguments and approach presented in this paper are being currently carried out in the frame of the European funded FP7 IST project PRISM (contract number 0215350). This project envisions a challenging research agenda, whose ultimate goal is to produce a network monitoring system which retains the effectiveness of monitoring applications without infringing the privacy requirements of the network users.

2. Regulatory issues affecting network monitoring

The design of a comprehensive monitoring system cannot be considered as a purely technical activity, as it affects the society as a whole (in terms of both its implications on privacy and in terms on its consequences on public security). Moreover, European policies and regulatory requirements do affect to a significant extent the technical requirements of a network monitoring system. Goal of this section is to understand and illustrate the tight relation between regulatory issues and technical design implications.

1.1 –The relation between data protection and network monitoring

The data protection right is acknowledged by European legislation as a fundamental right of the individual [8, 9, 10]. In order to understand whether (and to what extent) traffic monitoring activities are subject to data protection legislation, it is essential to clarify what 'personal data' are. Directive 95/46/EC [9] under Article 2 defines personal data as "*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*".

This definition stresses on the explicit reference to indirect identification data. It implies any information that may lead to the identification of the data subject through association with other available information (thus indirectly). The important definition of personal data has been further elaborated in subsequent documents from the Article 29 Data Protection

Working Party, and particularly in WP136[1]. Here, example 15, specifically dedicated to discuss dynamic IP addresses, concludes that "… *unless the Internet Service Provider is in a position to distinguish with absolute certainty that the data correspond to users that cannot be identified, it will have to treat all IP information as personal data to be on the safe side*". Therefore, data gathered through passive monitoring may be considered as personal data, subject to the data protection legislation.

The e-privacy Directive 2002/58/EC [10] has granted a specific and high degree of protection to traffic and location data, imposing strict limits and requirements to their processing due to their peculiar nature. Indeed, traffic data allow knowing user's activities and behaviour and defining the user's personality; traffic data allow the user's localisation and tracking of his movements. Combined traffic and location data enable to build user's profile enriched with geographical information, thus resulting in a significant encroaching into the individual's personal life and in invasive surveillance. Moreover, the amount of data that may be gathered through communications network traffic analysis is potentially indefinite; from a privacy law perspective, the amount of data processed raises privacy concerns, since the application of data mining algorithms and specific elaboration techniques gives the possibility to build precise users' profiles [11].

We can therefore conclude that communications network monitoring i) involves the gathering of information which is 'personal data' under the meaning of the Directive 95/46/EC [9]; ii) represents an activity that poses serious risks to the individual's right to data protection and freedoms, and that iii) data processing activities performed while monitoring traffic should be carried out in line with the set of rules and limitations provided by data protection legislation.

1.2 –The relation between security/safety provisions and network monitoring

The design of a network monitoring system is complicated by the need to additionally satisfy public security and safety constraints. Traffic logging was fostered by governmental entities as a means to protect the citizens' safety by facilitating the trace-back of malicious or criminal network users. Recently, data retention obligations were translated into concrete regulations both at a European level with the Directive 2006/24/EC [12] and at a national level with specific laws' enactment in some specific European Union countries[2].

This poses a number of issues in network monitoring besides the technical and cost challenges concerning storage needs raised by Internet service provides (ISP). First, the need to log information that may be later used for the above mentioned malicious users trace-back purposes restricts the applicability of one-way anonymization techniques. Indeed, the capability to revert the data protection measures set forth (preferably restricted to data strictly necessary for the specific purpose), should always be guaranteed. Second, the presence of such logs for a potentially long period of time (depending on the specific provisions, from a minimum of 6 months to several years), in conjunction with the data protection provisions previously discussed, raises security concerns involving the protection of the stored data from unauthorized users: ISPs must in fact provide means that do not allow the gathered data to be used for any unintended purposes.

The complexity of such supplementary requirements on the monitoring storage infrastructure is best addressed through the brief illustration of the contents of a recent[3]

---

1 WP 136 - Opinion 4/2007 on the concept of personal data, adopted on June 20, 2007; available at http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf

2 For example in Italy under Legislative Decree July 27, 2005 n. 144 and following amendments and integrations.

3 General regulation of the Italian data Protection Authority issued on January 17, 2008 and published on Bulletin n. 91/January 2008; Official Gazette n. 30 of February 5, 2008 - This provision will apply starting on October 31 2008 for a wide number of provider classes.

Italian provision issued by the Italian Privacy authority on the technical and organizational measures to protect and handle retained electronic data. Among these measures, more closely related to our purposes are: i) severe restrictions on the access to the data, to be eventually implemented via biometric approaches; these restrictions further include the system administrators themselves; ii) access restrictions to the locations at which storage systems are deployed; iii) the need to deploy sophisticated authorization systems, where a rigid separation must be exerted between the personnel and/or entities which assigns authorization credentials, and the personnel which shall access the data; iv) audit logs must be deployed to guarantee traceability of the accesses made by the operating personnel; v) data retained for investigation purposes and fraud detection/repression must be stored separately from that used by ordinary operation and management (billing, marketing, statistics, etc.); vi) cryptographic protection of retained data becomes mandatory, in order to avoid access to these data by system administrators, database administrators, and hardware- and software maintainers).

3. An approach towards privacy-preserving network monitoring

We believe that the key engineering guidelines for privacy-preservation are:

- Protect the data as soon as they are captured, i.e. on the on-line monitoring probes
- Adapt monitoring applications to operate on protected data
- Decouple the entity in charge of enforcing data protection (e.g. legal authority) from that one in charge of running, e.g. monitoring applications (i.e., the network operator)
- Provide a comprehensive framework for the control of the access and processing of the stored data traces.

These ideas are reflected in the multi-component, two-tiered system architecture sketched in Figure 1. Unlike traditional architectures, that are typically monolithic from the functional point of view, the envisioned system is comprised of three separate sub-systems that are also administratively independent: the Front-End tier, the Back-End tier and the Privacy-Preserving Controller (PPC).

The **Front-End tier** is deployed on the on-line traffic probe. Its goal is to capture the packets on the network link(s), transform them according to suitably designed data protection mechanisms (e.g. encryption, hashing) using algorithms and keys provided by the PPC, and deliver the data to the Back-End tier. The fundamental difference with respect to traditional approaches is the protection of data at an early stage of the packet capture process, before the delivery to the back-end processing system. This operation prevents the data controller (typically the network operator) from accessing raw data traces and thus provides very high privacy guarantees. It is obvious that in order for the monitoring infrastructure to retain its effectiveness, the data protection mechanisms on the front-end tier need to be carefully designed. Also, monitoring applications need to be adapted to cope with the protected data. These issues will be discussed below in section 4.

The **Privacy-Preserving Controller** (PPC) is an entity developed and administered independently from the rest of the system. We note here that this approach is in line with the spirit of the recent regulatory trends discussed in section 2.2 – in principle the PPC might be operated directly by an external authority, although in practice it is expected to be a separate team inside the operator organization, e.g. the internal legal department. The PPC provides and maintains the cryptographic secrets which are used by the data protection mechanisms enforced on the front-end, and orchestrates the reversion/escrow procedures mandated by the regulatory provisions. Note that the PPC is not meant to be able to access the actual captured data but only to control their protection. As such, a rigid separation between access to the actual data and its protection is enforced.

The **Back-End system** coincides with the traditional data controller role (typically the operator technical staff). As discussed later in section 5 the proposed implementation

involves a semantic-rich policy-based access control middleware. It relies on a semantic model, implemented by means of an OWL ontology [14], which allows integrating into the middleware operation while fulfilling the specific requirements imposed by regulatory provisions. The back-end system is in charge of collecting, storing and mediating access to the stored data traces. It also supports off-line data processing and filtering functionalities. This approach is extremely flexible and provides an efficient support for outsourced monitoring applications (run by external third party entities which need to access privacy-sensitive data). Additionally, the back-end system allows sanitised export of measurement information and/or derived statistics. This is accomplished by applying strong (not reversible) anonymization mechanisms, as in more traditional monitoring infrastructures. In addition to controlling the access to the data – by internal monitoring applications, system users, or external third party entities – the back-end can interact with the PPC. Specifically, when deemed necessary by the operator (e.g., following detection of abuses and/or severe anomalies which require prompt reactions) and/or mandated by regulatory provisions (e.g., for inspection of retained data by public authorities), the back-end interacts with the PPC in order to revert the data protection set forth at the front-end. This operation is performed selectively on the minimal subset of data necessary to perform the considered operation (e.g. subtrace for an individual user) as formally described through the ontology.

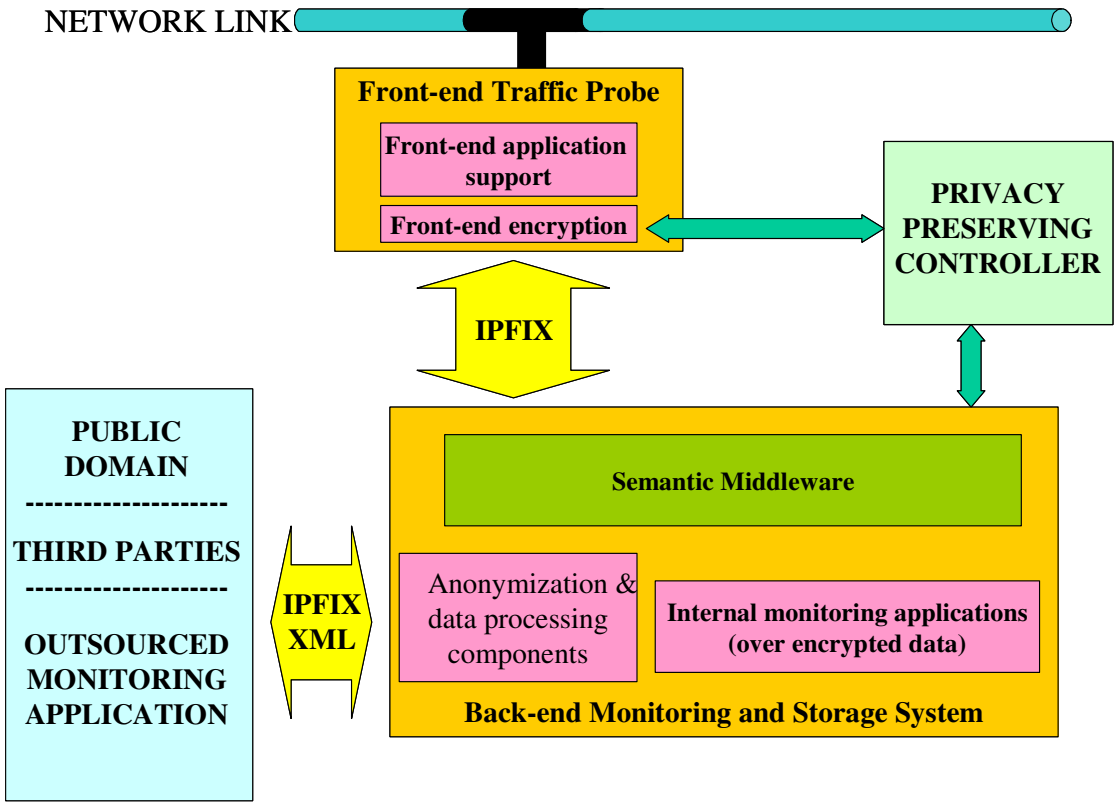

Figure 1 – System Architecture (sketch)

The envisioned architecture also specifies the protocol which allows the architecture components to exchange monitored data and measurement information. Specifically, data export from the front-end tier to the back-end tier and from the back-end tier to other domains, repositories or applications will be based on the upcoming standard for flow information export: **the IPFIX protocol** [13]. The IPFIX protocol has been developed and standardised by the IETF for the purpose of exporting IP flow information from IP devices such as routers or measurement probes to mediation, accounting, and network management systems. Beyond being an upcoming standard and therefore supporting interoperability, IPFIX was chosen for a number of other reasons. First of all, the protocol offers a high flexibility in terms of either choice of the flow key as well as selection of exported fields. The flow keys define the properties used to select flows (or packets) and can be defined according to application needs. This is a significant improvement with respect to other solutions, such as the widely used NetFlow, which selects flows based on a fixed attribute tuple. Information is exported using template records and data records. Templates contain {type, length} pairs specifying which {value} fields are present in data records conforming to the template, giving therefore great flexibility as to what data is transmitted. In addition, IPFIX is easily extensible, to meet the needs of present and future applications; new {value} fields can be easily added by vendors. Finally, and specifically relevant for our scenario, explicit support for anonymized data export in IPFIX has been already foreseen when discussing the protocol requirements (RFC 3917). Ongoing standardization work in IETF is in progress, with the goal of including support for anonymization in a next version of the IPFIX standard draft.

To conclude, it has to be noted that the proposed tiers of our proposed system are complementary and can be separately deployed. In other words, it is possible to deploy only the front-end tier with the companion PPC and rely on a traditional back-end system (at the price of lower privacy guarantees) or, alternatively, to implement only the back-end middleware.

4. Adaptation of Monitoring Applications to Front-End Encryption

An ambitious challenge underlying the envisioned architecture remains the possibility to operate monitoring applications in presence of data protection mechanisms, with null or minimal loss of their effectiveness. We distinguish two important classes of monitoring applications. The first one includes tools for network diagnosis, planning and quality assurance. These tools typically process passively collected traces and extract aggregated data that are used for a wide range of operational tasks, for example to infer network problems like congestion (see, e.g. [15]), to measure the actual performance experienced by the users [16], to quantify the level of resource consumption and feed the planning process, and to verify service level agreements (SLA). These applications rely heavily on protocol information that is available in the packet headers and in the initial part of the payload. In other words, they do not require complete access to the packet payload (which in principle can be removed) and the privacy problem reduces to the protection of the identity associations (anonymization). The second class of security-oriented applications and traffic classification is using deep packet inspection and relies strongly on the possibility of performing fine-grain (bit-level) analysis of the whole payload. These are mainly signature-based Intrusion Detection Systems (IDS), where the key processing module is based on pattern-matching engine (e.g., the SNORT open tool [17]), code emulation [18], or accurate application classification [19]. Resolving the privacy problem is extremely tough in this case as full payload access is an intrinsic requirement.

For both application classes, the solution strategy is to adapt the application in order to be integrated into the two-tier architecture presented above. For most monitoring applications, it is possible to divide the analysis functions into two separate processing

stages, where the first stage is maintained computationally light. As such, it can be performed directly on-the-fly in the front-end tier ("Front-end application support" module in Figure 1) where it can access raw (unencrypted plaintext or binary) data. The first stage produces meta-information useful for the specific application and is delivered in protected form to the back-end that stores it for later analysis. Recall that the following "Front-end encryption module" applies cryptographic transformations to protect the various data components. Different transformations can be applied to each data field, including hashing (e.g., on the IP addresses), encryption (of the payload), and even complete removal of certain parts of the data. Therefore, the second processing stage must be modified in order to work with such protected data and the associated meta-data. The two stages approach allows the complete division of functionalities into separate legal entities and hence the possibility to outsource very specific monitoring applications to specialised third party companies, e.g. for SLA monitoring. Furthermore, it will also allow operators to release privacy-protected data to researchers, enabling them to develop new algorithms for IDSes or traffic classifiers, etc.

Besides processing the sensitive data in different stages, a complementary approach to solve the processing/protection dilemma relies on the so-called "blind" processing techniques. The idea is to perform certain analysis tasks directly in the encrypted data space without decrypting the data to plaintext during the process. This problem has been widely addressed by the cryptography community, especially for its potential application to encrypted databases [20, 21]. A possible choice is the use of "homomorphic encryption" techniques [22], which preserve the possibility to perform selected operations over the encrypted data. Many asymmetric encryption techniques (e.g. the topmost known El Gamal [23] and RSA [24] schemes) indeed are natively homomorphic. The problem is that "slow" asymmetric encryption might be unsuited in terms of performance requirements to current data rates. Faster asymmetric homomorphic encryption approaches or even symmetric methods (hence extremely fast) are indeed available [22], but in most cases at the cost of security weaknesses.

Exploring the viability of blind processing techniques in a network monitoring scenario, and in particular under the severe performance constraints imposed by our proposed two-tier architecture (ultra gigabit per second data rates and implementation of the encryption scheme over the front-end traffic probe), is a highly innovative and widely open issue. As such it is a major challenge for our future research directions. Indeed, to the best of our knowledge, there is only one prior work, namely [25], dealing with blind processing of data packets (and specifically for Intrusion Detection Systems). An advantage of the proposed architecture consists in the fact that the access to encrypted data is regulated by the back-end operation (see section 5). This partially relax the security requirements of the encryption mechanisms employed over the front-end. In other words, the adoption of relatively weak encryption schemes, which would be clearly unsuited if the encrypted data were exposed to public access, might become acceptable when complemented with controlled access restrictions to the data.

5. A semantic model for the Back-End middleware

As discussed in section 2.2, the engineering of a comprehensive back-end operation may be not only recommended, but even mandated by emerging regulations. The back-end tier constitutes a semantic middleware providing support for (at least) the following operations:

i) Management, through interaction with the PPC, of the procedures to revert the data protection means exerted by the frond-end tier. This occurs either because the retained data must be selectively inspected by external authorities or because the monitoring application's operation must trigger reactions in the network environment against potential threats.

ii) Enforcement of a privacy-aware access control mechanism, specifically based on the regulatory provisions. Extending the role-based models, the back-end tier incorporates in the authorization procedures for accessing and processing stored data additional "contextual" information, including the reason why an action is requested, the semantically defined type of data, etc.

iii) Management of the distributed and encrypted storage system accordingly;

iv) Implementation of the appropriate anonymization mechanisms for exporting the gathered data (or related statistics) to third party applications or even to the public community, and whenever appropriate dynamically activate data processing functions whose evaluation result is provided to the accessing party (thus preventing its direct access to critical or weakly protected information).

Privacy-aware access control constitutes a very important feature of the back-end tier. It provides a generalization of the notion of "access" to stored data, where the data access requests are evaluated against a dynamic "context" state characterized by a multiplicity of parameters, including i) the type and identification of the user or requesting entity, ii) the type of the data to be accessed, iii) the processing intended, iv) a stated purpose, v) the environment on which the data will be used, vi) the applicable regulatory provisions [9, 10, 12] for the requested data and requesting entity types, vii) other provisions such as a "consulting" verdict of a Bayesian filter about the access request, etc.

In that respect, the back-end system is empowered with a policy-based decision engine that takes into account the privacy requirements. This policy engine reacts to access requests based on a dynamic "privacy context" that incorporates all parameters into a single evaluation block. Matching the privacy context against a model of defined privacy-related access policies enables the back-end system to take a decision and apply the appropriate protective measure before releasing the data. The model of defined policies is implemented by means of an ontology. It provides a formal and machine-readable representation of policy rules that originate from a semantic description of the monitoring applications operation and that explicitly includes privacy legislation provisions. The ontology constitutes the core of data protection at the back-end tier. It provides the configuration of the decision engine and defines how the back-end tier will treat the data before exposing them to the requesting entity.

It should be noted here that the use of an ontology is preferred for the implementation of the model instead of a legacy policy language, such as OASIS XACML [26] or IBM EPAL [27] due to the fact that the latter two are limited by the low expressive power of formalisms used to describe resources and entities requesting them, compared to ontology-based models. An ontology enables the specification of complex concepts, such as rules governing the combinatorial export of data types which alone are not privacy-sensitive but raise privacy issues when disclosed together. Additionally, an ontology-driven approach provides significant advantages in terms of rules consistency evaluation, reasoning capabilities, as well as integration with other semantic models, e.g., domain ontologies describing data resources, roles or monitoring applications.

Based on the decision engine outcome, the back-end tier mediates access to the stored data by the execution of a processing function (such as a specific anonymization mechanism, a statistic elaboration and consequent production of meta-data, execution of a monitoring application module) or a combination of multiple modules. In that respect, the system is complemented by efficient execution operations that manage the cryptographic procedures and keys to enable access and processing of the encrypted stored data. Additionally, the system stores in a database and log-files all information about published data, including the recipient and the purpose; this way, reactive auditability is enabled.

An immediate consequence of this approach is the ability to support third-party "outsourced" monitoring applications. In principle, monitoring applications developed by

third-parties might require access to the internal captured data in order to perform advanced operations, which a robust anonymization mechanism devised for data public export would impede. By allowing the controlled execution of data processing procedures inside the controller domain, the back-end tier enables outsourced applications to operate on privacy-sensitive data. With this approach only the application output is accessed by the external party, not the input data. In this way it might be possible to define collaboration models for data processing between the network operators and external parties such as research groups in different universities. This would help to revive Internet traffic research, an important area that is starting to suffer seriously from the obstacles caused by various privacy regulations (see [28]).

6. Conclusions

This paper addresses the challenges and issues emerging in the design of privacy-aware network monitoring systems. The solution is seen in a two-tier architecture system. A front-end tier of data protection mechanisms is directly enforced at the traffic probe device, guaranteeing that the data delivered to the back-end storage will be already privacy-protected, while the back-end tier enforces privacy-aware access control to the collected data, orchestrating at the same time the operation of reversing the data protection mechanisms set forth by the front-end tier. The back-end system retains the ability to reverse the data protection measures by cooperating with an external element called "Privacy-Preserving Controller".

The approach proposed in this paper aims at satisfying three conflicting aspects: i) enable legal investigations by law enforcement authorities, ii) safeguard the right of the individual to communication privacy, and iii) allow the network operators to perform traffic monitoring as part of the operation process of their infrastructures.

The technical approach outlined in this paper holds high potential in terms of social impact. By providing guarantees of user data protection, it can shape the way citizens view network monitoring activities, thus enhancing the level of trust in future network infrastructures.

In view of the emerging regulatory trends which aim at favouring the citizens´ privacy, operators and ISPs will be required to upgrade their monitoring infrastructures to meet the new privacy-related provisions. The availability of advanced and regulatory-compliant approaches such as the one proposed in this paper can be of significant help especially for the small and medium scale operators which have limited internal resources to design and deploy their own privacy-preserving access control solutions.

7. Acknowledgements

The authors gratefully acknowledge the contributions of Dr. Esa Hyytiä, Dr. Ivan Gojmerac, and Mr. Peter Dorfinger in the development and revision of this paper. The authors also acknowledge all the researchers working in the PRISM project for the extensive in-depth discussions on the topics tackled in the paper.

8. References

[1] A Hintz, "Fingerprinting Websites Using Traffic Analysis Privacy Enhancing Technologies", in Proc. of 2nd Workshop on Privacy Enhancing Technologies, San Francisco, CA, USA, April 2002.

[2] G. D. Bissias, M. Liberatore, D. Jensen, B. N. Levine, "Privacy Vulnerabilities in Encrypted HTTP Streams", in Proc. of 5th Workshop on Privacy Enhancing Technologies, Cavtat, Croatia, May 2005.

[3] M. Crotti, F. Gringoli, P. Pelosato, L. Salgarelli, "A statistical approach to IP-level classification of network traffic", in Proc. the 2006 IEEE International Conference on Communications, June 2006.

[4] M. Peuhkuri. A method to compress and anonymize packet traces. Internet Measurement Workshop (San Francisco, California, USA: 2001), pages 257–261, 2001.

[5] D. Plonka. ip2anonip – http://dave.plonka.us/ip2anonip

[6] J. Xu, J. Fan, M. Ammar and S. B. Moon, “Prefix preserving IP address anonymization: Measurement-based security evaluation and a new cryptography-based scheme”, in Proc. 10th IEEE Intl. Conference on Network Protocols (ICNP 2002), Paris, France, November 2002.
[7] R. Pang, M. Allman, V. Paxson and J. Lee, “The Devil and Packet Trace Anonymization”, ACM Computer Communication Review, Vol. 36, No. 1, January 2006.
[8] Charter of Fundamental Rights of the European Union, O.J. C 364/1, 18.12.2000.
[9] 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.
[10] Directive 2002/58/EC of the European Parliament and of the Council of 12 July 2002 concerning the processing of personal data and the protection of privacy in the electronic communications sector, O. J. L. 201, 31.07.2002.
[11] C. Clifton and D. Marks, “Security and privacy implications of data mining”, ACM SIGMOD Workshop on Research Issues on Data Mining and Knowledge Discovery, pages 15–19, May 1996.
[12] Directive 2006/24/EC f 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, 13.04.2006.
[13] B. Claise, S. Briant, G. Sadasivan, S. Leinen and T. Dietz, “Specification of the IPFIX Protocol for the Exchange of IP Traffic Flow Information”, Internet-Draft, work in progress (currently in RFC queue).
[14] The World Wide Web Consortium (W3C), “Web Ontology Language (OWL)”, online: http://www.w3.org/2004/OWL/.
[15] F. Ricciato, “Traffic Monitoring and Analysis for the Optimization of a 3G network”, IEEE Wireless Communications, Vol. 13, No. 6, December 2006.
[16] R. Birke et al. "Understanding VoIP from Backbone Measurements", INFOCOM 2007.
[17] The SNORT open tool, home page: www.snort.org.
[18] M. Polychronakis, K. G. Anagnostakis, and E. P. Markatos, "Emulation-based Detection of Non-self-contained Polymorphic Shellcode", Proc. of 10th International Symposium on Recent Advances in Intrusion Detection (RAID). September 2007, Queensland, Australia.
[19] D. Antoniades, M. Polychronakis, S. Antonatos,Evangelos P. Markatos, Sven Ubik, and Arne Øslebø, “Appmon: An Application for Accurate per Application Network Traffic Characterization”, BroadBand Europe 2006, Geneva, Switzerland, December 11-14, 2006.
[20] Dawn Xiaodong Song, David Wagner, and Adrian Perrig, Practical techniques for searches on encrypted data, In IEEE Symposium on Security and Privacy, 2000.
[21] Zhiqiang Yang, Sheng Zhong, and Rebecca N. Wright, Privacy-Preserving Queries on Encrypted Data, Proceedings of the 11th European Symposium On Research In Computer Security (Esorics), 2006.
[22] C. Fontaine, F. Galand, “A survey of homomorphic encryption for nonspecialists”, Int. Journal on Information systems, Vol. 2007, Issue 1, January 2007.
[23] T. El Gamal, “A Public-Key Cryptosystem and a Signature Scheme Based on Discrete Logarithms”, IEEE Transactions on Information Theory, v. IT-31, n. 4, 1985, pp. 469–472.
[24] R. Rivest, A. Shamir, L. Adleman, “A method for obtaining digital signatures and public-key cryptosystems”, Communications of the ACM, vol . 21 no. 2, pp. 120–126, Feb. 1978.
[25] Lukas Kencl, Jose Zamora, Martin Loebl, "Packet Content Anonymization by Hiding Words", Demo at IEEE INFOCOM, Barcelona, Spain, April 2006.
[26] Organization for the Advancement of Structured Information Standards (OASIS), “eXtensible Access Control Markup Language TC”, 2004.
[27] P. Ashley, S. Hada, G. Karjoth, C. Powers, M. Schunter, “The Enterprise Privacy Authorization Language (EPAL), EPAL 1.2 Specification”, *IBM Research Report*, 2003.
[28] P. Ohm, D. Sicker and Dirk Grunwald, "Legal Issues Surrounding Monitoring During Network Research" Internet Measurement Conference, October 24-26, 2007, San Diego, California, USA.