ETSI GS MOI 002 V1.1.1 (2012-07)

Measurement Ontology for IP traffic (MOI); Requirements for IP traffic measurement ontologies development

Disclaimer

This document has been produced and approved by the Measurement Ontology for IP traffic (MOI) ETSI Industry Specification Group (ISG) and represents the views of those members who participated in this ISG.
It does not necessarily represent the views of the entire ETSI membership.

Reference
DGS/MOI-0002

Keywords
IP, ontology, requirements, traffic

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

Individual copies of the present document can be downloaded from:
http://www.etsi.org

The present document may be made available in more than one electronic version or in print. In any case of existing or perceived difference in contents between such versions, the reference version is the Portable Document Format (PDF). In case of dispute, the reference shall be the printing on ETSI printers of the PDF version kept on a specific network drive within ETSI Secretariat.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at
http://portal.etsi.org/tb/status/status.asp

If you find errors in the present document, please send your comment to one of the following services:
http://portal.etsi.org/chaircor/ETSI_support.asp

Contents

Intellectual Property Rights

IPRs essential or potentially essential to the present document may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (http://ipr.etsi.org).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Foreword

This Group Specification (GS) has been produced by ETSI Industry Specification (ISG) Measurement Ontology for IP traffic (MOI).

Introduction

Defining a complete set of concepts and their relationship to support coherent developments of traffic measurement systems needs not only an extensive mapping of key performance indicators (KPI), key quality indicators (KQI) and currently used parameters to describe IP networks, but a serious pre-definition of the framework and extension of the ontology to be set up. If it is true that ontologies should be complete and internally coherent, one ought to be aware of the final purpose of establishing such an ontology, namely supporting information systems to achieve IP traffic monitoring and quality of service (QoS) applications that can exchange information and back service level agreements (SLA) up and fulfil the expectations of customers by assuring the quality of experience (QoE).

The present document starts setting up the limits to the ontology that needs to be defined for such purpose. Then, after reviewing the parameters that define IP networks, some use cases are used to analyse which internal specifications should be respected in order to give rise to a coherent ontology for IP traffic monitoring useful for practical purposes.

1 Scope

The present document identifies the requirements that should characterise an ontology for the semantic conceptualisation of information related to IP traffic measurements. The requirements are obtained through the analysis of use cases spanning across a variety of related application categories and domains of interest, as well as the consideration of additional qualitative needs, such as the protection of personal data. Additional inputs arise from user experience, as well as the 'GS/MOI-010' Work Item study, entitled "Report on information models for IP traffic measurement" [1]. The general difficulty of setting limits to an ontology, taking concepts from outside is also dealt within the present document that states MOI focus on IP traffic measurement concepts and let's side ontologies dealing with other subjects, an easy way to link. Thus a rather practical approach to define MOI ontology will be laid so that further QoS, traffic monitoring and Internet governance issues can be built on top of it by means of semantic tools.

2 References

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the reference document (including any amendments) applies.

Referenced documents which are not found to be publicly available in the expected location might be found at http://docbox.etsi.org/Reference.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

2.1 Normative references

The following referenced documents are necessary for the application of the present document.

[1] ETSI GS MOI 010: "Measurement Ontology for IP Traffic (MOI); Report on Information Models for IP Traffic Measurement".

2.2 Informative references

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] FP7 ICT project MOMENT (Monitoring and Measurement in the Next Generation Technologies).

NOTE: Available at http://fp7-moment.eu/

[i.2] FP7 ICT project PRISM (PRIvacy-aware Secure Monitoring).

NOTE: Available at http://fp7-prism.eu/

[i.3] FP7 ICT project DEMONS (DEcentralized, cooperative, and privacy-preserving MONitoring for trustworthiness).

NOTE: Available at http://fp7-demons.eu/

[i.4] IETF RFC 2330: "Framework for IP Performance Metrics".

[i.5] ITU-T Recommendation Y.1561: "Performance and Availability Parameters for MPLS Networks".

[i.6] ITU-T Recommendation Y.1540: "Internet protocol data communication service - IP packet transfer and availability performance parameters".

[i.7] IETF RFC 2679: "A One-way delay Metric for IPPM".

[i.8] ITU-T Recommendation Y.1544: "Multicast IP performance parameters".

[i.9] IETF RFC 5644: "IP Performance Metrics (IPPM): Spatial and Multicast".

[i.10] IETF RFC 3393: "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)".

[i.11] ITU-T Recommendation Y.1543: "Measurements in IP networks for inter-domain performance assessment".

[i.12] IETF RFC 2681: "A Round-trip Delay Metric for IPPM".

[i.13] IETF RFC 2680: "A One-way Packet Loss Metric for IPPM".

[i.14] IETF RFC 3357: "One-way Loss Pattern Sample Metrics".

[i.15] IETF RFC 4737: "Packet Reordering Metrics".

[i.16] IETF RFC 5560: "A One-Way Packet Duplication Metric".

[i.17] IETF RFC 3148: "A Framework for Defining Empirical Bulk Transfer Capacity Metrics".

[i.18] IETF RFC 2678: "IPPM Metrics for Measuring Connectivity".

[i.19] ITU-T Recommendation Y.1540 Amendment 1.

[i.20] IBM, "Understanding the Autonomic Manager Concept".

NOTE Available at http://www.ibm.com/developerworks/library/ac-amconcept/

[i.21] D. E. Monnat, A. L. Ethen, "A Primer on the Federal Wiretap Act and Its Fourth Amendment Framework", Journal of the Kansas Trial Lawyers Association, Vol. 28, No. 1, pp. 12-15, 2004.

[i.22] Council of the European Union, "Council Resolution of 17 January 1995 on the lawful interception of telecommunications", Official Journal of the European Communities, No. C 329, pp. 1-6, November 1996.

[i.23] ETSI TS 102 233: "Lawful Interception (LI); Service specific details for E-mail services".

[i.24] ETSI TS 102 234: "Lawful Interception (LI); Service-specific details for internet access services".

[i.25] ETSI TS 102 656: "Lawful Interception (LI); Retained Data; Requirements of Law Enforcement Agencies for handling Retained Data".

[i.26] ETSI TS 102 232-1: "Lawful Interception (LI); Handover Interface and Service-Specific Details (SSD) for IP delivery; Part 1: Handover specification for IP delivery".

[i.27] A. Zugenmaier and J. Claessens, "Privacy in Electronic Communications", in Network Security: Current Status and Future Directions, C. Douligeris & D.N. Serpanos (Eds.), pp. 419 - 440, Wiley-Interscience & IEEE Press, 2007.

[i.28] G. V. Lioudakis, E. A. Koutsoloukas, N. Dellas, N. Tselikas, S. Kapellaki, G. N. Prezerakos, D. I. Kaklamani, I. S. Venieris, "A Middleware Architecture for Privacy Protection", Computer Networks, Vol. 51, No. 16, pp. 4679 - 4696, November 2007.

[i.29] L. F. Cranor, "I Didn"t Buy It for Myself", in Designing Personalized User Experiences in E-Commerce, C.-M. Karat, J.O. Blom, & J. Karat, (Eds.), pp. 57 - 73, Kluwer Academic Publishers, 2004.

[i.30] G. D. Bissias, M. Liberatore, D. Jensen and B. N. Levine, "Privacy Vulnerabilities in Encrypted HTTP Streams", in Proceedings of the 5th Workshop on Privacy Enhancing Technologies (PET 2005), Cavtat, Croatia, May 30 - June 1, 2005, LNCS 3856.

[i.31] M. Crotti, F. Gringoli, P. Pelosato and L. Salgarelli, "A Statistical Approach to IP-Level Classification of Network Traffic", in Proceedings of the IEEE International Conference on Communications (ICC) 2006, Istanbul, Turkey, June 11 - 15, 2006.

[i.32] A. Hintz, "Fingerprinting Websites Using Traffic Analysis", in Proceedings of the 2nd Workshop on Privacy Enhancing Technologies (PET 2002), San Francisco, CA, USA, April 14 -15, 2002, LNCS 2482.

[i.33] Q. Sun, D. R. Simon, Y.-M. Wang, W. Russell, V. N. Padmanabhan and L. Qiu, "Statistical Identification of Encrypted Web Browsing Traffic", in Proceedings of the 2002 IEEE Symposium on Security and Privacy (SP" 02), Marseille, France, May 12 - 15, 2002.

[i.34] S. Bellovin, "A Technique for Counting NATted Hosts", in Proceedings of the 2nd ACM SIGCOMM Workshop on Internet Measurement (IMW" 02), Berkeley, CA, USA, November 6-8 2002.

[i.35] European Parliament and Council, "Directive 2002/58/EC of the European Parliament and of the Council concerning the processing of personal data and the protection of privacy in the electronic communications sector (Directive on privacy and electronic communications)", Official Journal of the European Communities, No. L 201, pp. 37 - 47, July 2002.

[i.36] European Parliament and Council, "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", Official Journal of the European Communities, No. L 105, pp. 54 - 63, April 2006.

[i.37] "United States Code 18, § 2701: Unlawful access to stored communications".

[i.38] European Parliament and Council, "Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data", Official Journal of the European Communities, No. L 281, pp. 31-50, November 1995.

[i.39] G. V. Lioudakis, F. Gaudino, E. Boschi, G. Bianchi, D. I. Kaklamani, I. S. Venieris, "Legislation-Aware Privacy Protection in Passive Network Monitoring", in Information Communication Technology Law, Protection and Access Rights: Global Approaches and Issues, I. M. Portela, M. M. Cruz-Cunha (Eds), IGI Global, 2010.

[i.40] The World Wide Web Consortium (W3C), "XML Schema Part 2: Datatypes Second Edition", W3C Recommendation 28 October 2004.

NOTE: Available at http://www.w3.org/TR/xmlschema-2/

[i.41] Morfeo project, Measurement Units Ontology.

NOTE: Available at http://forge.morfeo-project.org/wiki_en/index.php/Units_of_measurement_ontology

[i.42] NASA, Semantic Web for Earth and Environmental Terminology (SWEET) v. 1.1, Units Ontology.

NOTE: Available at http://sweet.jpl.nasa.gov/1.1/

[i.43] T. R. Gruber, G. R. Olsen, "An Ontology for Engineering Mathematics", in Proceedings of the 4th International Conference on Principles of Knowledge Representation and Reasoning (KR'94), Bonn, Germany, May 24-27, 1994.

[i.44] FP7 ICT project MOMENT, MOMENT Units Ontology.

NOTE: Available at https://svn.fp7-moment.eu/svn/moment/public/Ontology/MomentUnits.owl

[i.45] H. Stuckenschmidt, F. van Harmelen, "Information Sharing on the Semantic Web", chapter 7: Sharing statistical information, Springer, 2005, ISBN: 978-3-540-20594-4.

[i.46] E. Shaya, "Ontology of Statistics (background for Astronomy Ontology)".

NOTE: Available at http://www.astro.umd.edu/~eshaya/astro-onto/ontologies/statistics.html

[i.47] F. Guala, "An Ontology of Economics?", Extended version of the review of "The Elgar Companion to Economics and Philosophy" appeared in the Economic Journal, 116 (2006), pp. 318-321.

NOTE: Available at http://people.exeter.ac.uk/fguala/OntologyLong.pdf

[i.48] OWLIM Semantic Repository.

NOTE: Available at http://www.ontotext.com/owlim/

3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

CC	Content of Communication
CIM	Common Information Model
CPU	Central Processing Unit
ICMP	Internet Control Management Protocol
IPTV	Internet Protocol TeleVision
IRI	Interception Related Information
ISP	Internet Service Provider
KPI	Key Performance Indicators
KQI	Key Quality Indicator
LEA	Law Enforcement Agency
LI	Lawful Interception
MOS	Mean Opinion Score
MPLS	Multi Protocol Label Switching
OSI	Open System Interconnection
OWL	W3C Web Ontology Language
QoE	Quality of Experience
QoS	Quality of Service
RDF	Resource Description Framework
RDFS	RDF Schema
SLA	Service Level Agreement
SNMP	Simple Network Management Protocol
TMA	Traffic Monitoring and Analysis
TCP	Transmission Control Protocol
UDP	User Datagram Protocol
XSD	XML Schema Definition

4 Framework of the Ontology for IP Measurements

Network monitoring constitutes a key task in current communication networks, with a broad applicability domain. It is crucial for the effective operation and management of communication networks, since it enables the acquisition of essential information and the identification, among other, of performance bottlenecks, while the offline analysis of the resulting traces is very useful for network planning and the accounting and billing of network services. In this context, network monitoring can additionally serve for the validation of Service Level Agreements and generally for the observation and fine-tuning of parameters related to Quality of Service and Quality of Experince. With respect to security and protection of networks, it constitutes the fundamental basis for Intrusion / Anomaly Detection Systems (IDS/ADS) which trigger alarms and set up countermeasures in reaction to events, such as network intrusions, denial-of-service attacks and worm infections. In addition, the data traces that constitute the result of network monitoring are very useful for the research community that investigates the fields mentioned above, as well as other network-related research domains. Last but not least, network monitoring provides the means for the implementation of several obligations mandated by the law; these include the retention of certain data for ensuring their availability if needed for the investigation, detection and prosecution of serious crimes, as well as the performance of Lawful Interception.

Different monitoring tools and platforms have been developed through the years to obtain active and passive measurements about a variety of metrics. The integration of such measurements can be valuable for network operators to obtain network weathermaps or network tomographies. However, this integration in a single view is difficult because each platform uses its own data structures and its own interaction interfaces.

Ontologies can become very useful for the conceptual integration of the different measurement data models. Dealing with the underlying information at a semantic level can enable some degree of inference and automatic reasoning over the retrieved measurement data. At the same time, ontologies let define the information at different abstraction levels, allowing the definition of specific classes of measurements that are derived from generic ones. The present document focuses on this approach, applying the concepts provided by ontologies to address the integration of measurement information from a semantic viewpoint.

The MOI ontology has to describe the set of concepts, properties, relationships and axioms in the domain of Internet Network Measurements. In that respect, it is very important to define the limits of the ontology, and also set connections to other already defined and globally accepted ontologies. Given that the network measurement domain is very broad, it is important to be pragmatic, avoiding very large models that are difficult to apply in a real use case. For this, the ontology will be focused on what can be measured in an Internet Network; Quality of Service (QoS) and Quality of Experience (QoE) will be essential parts of this ontology, whereas special focus will be put on the critical issue of personal data protection as an integral part of the ontology. Other concepts will be taken from other existing ontologies, to a some extent investigated in [1]. Other general concepts are also available in already defined network management information models such as the CIM model, SNMP MIBs or ITU M.3100.

The work presented here is supported by three European research projects. The aim of the FP7 ICT MOMENT [i.1] was precisely to find ways to solve the integration problem. This integration was twofold: definition of a standard interface to access the information, as well as definition of a homogeneous view of the available information. For the latter integration, it was necessary to combine database schemas of the existing network measurement infrastructures, and leverage existing definitions from PerfSONAR, CAIDA, IPPM IETF WG, etc. On the other hand, FP7 ICT PRISM [i.2]focused on the issue of privacy protection in the context of passive network monitoring; a fundamental activity has been the specification of a semantic model. Finally, this work is supported by the FP7 ICT DEMONS [i.3]; it focuses on distributed and cooperative network monitoring for fulfilling a variety of objectives mostly centered around security, while it continues the work of PRISM with respect to personal data protection leveraging semantic technologies.

Next clauses investigate the requirements that should drive the development of the MOI ontology. A few representative use cases where the MOI ontology can be used are provided, while information domains and types that should be specified in the ontology, such as Key Performance Indicators (KPI), are outlined. Moreover, in order to cope with the personal data protection aspects, the underlying principles are also investigated. As a result, the set of requirements of the ontology are defined, specifying both main features and global requirements, as well as the requirements for the ontology structure.

5 KPI Descriptions and Mapping to MOI

Operators and private network managers require tools to manage the network, prevent and detect problems, plan and engineer sections.

The Key Performance Indicators (KPI) serve the purpose of quantifying metrics that reflect factors critical for the correct network behaviour in the sense of operational and/or business objectives. They help in the assessment of IP service quality levels that meet user needs or service agreements. They measure and show the performance of critical network services.

The KPIs will differ depending on the network architecture, business objectives, critical network sections or services, application performance requirements, system reliability, etc. Different network managers will select, configure and monitor different sets of KPIs. The KPIs are selected from parameters monitored by commercial or ad-hoc network performance monitoring tools and they are usually monitored in a dashboard fashion for quick access.

These parameters have been target of study for recommendations from both ITU-T and IETF organizations. Their recommendations make an effort in creating consistent definitions but anyway the metrics differ slightly, creating multiple alternatives for similar objectives.

Without trying to be extensive or detailed in parameter definitions, a short review of defined parameters is presented, trying to highlight (by grouping and specific comments) relations, similarities and discrepancies between parameter definitions from different organizations. These discrepancies and large space of parameters for similar concepts raise the need for a common semantic.

ITU-T specifications are rooted on Recommendation Y.1540 [i.6]which defines parameters that may be used in specifying and assessing the performance of speed, accuracy, dependability and availability of IP packet transfer services. IETF's work is developed on the IP Performance Metrics (IPPM) framework [i.4]. Extensions of ITU-T parameters for technologies like MPLS follow analogous definitions [i.5].

Many of the mentioned metrics can be modified by some statistical operators like: mean, minimum, median, quantile-based limits, interval-based limits, peak, etc. In some recommendations, the operator results in the definition of a new parameter.

5.1 Parameters Related to Delay or Delay Variation

- **IP packet transfer delay (IPTD)**, defined as the time between ingress and egress time of a packet to a network section [i.6].
- **Type-P-One-way-Delay**, as the time between the first bit of the packet in the wire at the source to the last bit of the packet in the wire at the destination [i.7]. This, like any following "Type-P" metric is a set of metrics where the "type-P" section makes reference to the type of packet used in the probe (ICMP ECHO Request, TCP SYN, UDP packet, etc).
- **Type-P-One-way-Delay-Poisson-Stream**, defined as a sequence of packet probes from a Poisson process and the one-way-delay measured by each one [i.7].
- **Global multicast mean one-way delay**, calculated as the sum of one-way delays for all successful IP packet transfers divided by the total successful IP packet transfers at the registered multicast destinations [i.6].
- **Multicast group mean one-way delay**, calculated as the sum of mean IPTD delays for all destinations divided by the number of registered destinations [i.8].
- **One-way mean delay range over multicast group, One-way delay variation range over multicast group** [i.8].
- **Type-P-Spatial-One-way-Delay-Vector**, consisting of a vector of one-way delay measurements from the source node to the destination nodes including nodes in the path [i.9]. Notice that ITU-T group parameters differ from IETF spatial parameters due to the former refer to multicast groups and the latter to nodes in a path.
- **Type-P-Segment-One-way-Delay-Stream** [i.9].
- **Type-P-One-to-group-Delay-Vector**, the set of one-way delays between source and destinations in a multicast group [i.9]. This parameter supports also that the destination is not a multicast group but a single host.
- **2-point packet delay variation (PDV)**, defined as the difference between the IPTD of a packet to a reference packet IPTD. Difference options for the reference definition are available [i.6].
- **Type-P-One-way-ipdv-Poisson-stream**, measuring the delay variation (IPDV) between pairs of packets in a Poisson stream [i.10]. IETF delay variation specifications are based on the difference between two successive delay measurements while ITU-T Recommendation Y.1540 [i.6] PDV is based on observations on the delay distribution. However, IETF specification has sufficient flexibility to produce whether inter-packet delay variation or the delay variation by using a fixed minimum delay reference [i.11].
- **Type-P-One way-ipdv-jitter**, measuring the IPDV between consecutive packets and taking the absolute values [i.10].
- **Type-P-Round-trip-Delay-Poisson-Stream**, as a sequence of packet probes from a Poisson process and the round-trip-delay measured by each one [i.12].
- **Type-P-Spatial-One-way-ipdv-Vector**, consisting of a vector of Type-P-One-way-ipdv measurements to nodes in the path from source to destination.
- **Type-P-Segment-ipdv-prev-Stream** [i.9].
- **Type-P-One-to-group-ipdv-Vector** [i.9].

5.2 Parameters Related to Errors and Losses

- **IP packet error ratio (IPER)**, defined as the ratio of total errored IP packets at the measurement point to the total successful IP packet transfers [i.6].
- **IP packet loss ratio (IPLR)**, defined as the ratio of total lost IP packets at the measurement point to the total transmitted IP packets [i.6].
- **Spurious IP packet rate**, measuring the packets that appear at the egress measurement point that do not have a corresponding ingress packet [i.6].
- **IP packet severe loss block ratio (IPSLBR)** [i.6].
- **Type-P-One-way-Packet-Loss-Poisson-Stream**, defined as a sequence of packet probes from a Poisson process and the result of the probe reaching or not the destination measurement point [i.13]. IETF packet loss definition differs from ITU-T Recommendation Y.1540 [i.6] IPLR in that errored packets are designated lost in IETF's definition but not in ITU-T's one.
- **Type-P-One-Way-Loss-Distance-Stream**, measuring the distance in sequence numbers between two successive losses [i.14].
- **Type-P-One-Way-Loss-Period-Stream**, identifying the losses in the same burst loss [i.14].
- **Type-P-One-Way-Loss-Noticeable-Rate**, a statistical operator on the previous one that considers not relevant losses that are far apart [i.14].
- **Type-P-One-Way-Loss-Period-Total**, measuring the total number of loss periods [i.14].
- **Type-P-One-Way-Loss-Period-Lengths**, measuring the length (in packets) of the loss periods [i.14].
- **Type-P-One-Way-Inter-Loss-Period-Lengths**, measuring the distance between successive loss periods [i.14].
- **Multicast global loss ratio**, measuring the sum of all lost packets divided by the sum of packets transmitted to each destination while a member of the specified multicast group [i.8].
- **Multicast mean group loss ratio**, measuring the sum of all point-to-point IPLR divided by the number of registered destinations that were members of the specified multicast group during the defined period [i.8].
- **Loss ratio range over multicast group, Comparative multicast group delivery ratio** [i.8].
- **Type-P-Spatial-Packet-Loss-Vector**, measuring whether the packet arrived to path members from source to destination [i.9].
- **Type-P-Segment-Packet-Loss-Stream** [i.9].
- **Type-P-One-to-group-Packet-Loss-Vecto**r [i.9]. Notice that IETF the basic results for losses in multicast groups in vector form while ITU-T aggregates these results [i.8].

5.3 Parameters Related to Packet Reordering, Replication and Duplication

- **IP packet reordered ratio (IPRR)**, defined as the ratio of reordered egress packets to the total of successful IP packet transfers [i.6]. Extension to multicast groups is also available [i.8].
- **Type-P-Reordered-Ratio-Stream**, measuring the ratio of reordered packets in a stream [i.15].
- **Type-P-Packet-Reordering-Extent-Stream**, measuring the maximum distance in the reordering [i.15].
- **Type-P-Packet-Late-Time-Stream**, measuring the time distance in the reordering extent [i.15].
- **Type-P-Packet-Byte-Offset-Stream**, measuring the received bytes before a reordered packet [i.15].

- **Type-P-Packet-Reordering-Gap-Stream, Type-P-Packet-Reordering-GapTime-Stream, Type-P-Packet-Reordering-Free-Run-x-numruns-Stream, Type-P-Packet-Reordering-Free-Run-q-squruns-Stream, Type-P-Packet-Reordering-Free-Run-p-numpkts-Stream, Type-P-Packet-Reordering-Free-Run-a-accpkts-Stream, Type-P-Packet-n-Reordering-Stream** [i.15].
- **IP packet duplicate ratio (IPDR)** [i.6] (extensible to multicast groups [i.8]).
- **Replicated IP packet ratio (RIPR)** [i.6].
- **Type-P-one-way-packet-arrival-count**, counting the number of packets arriving for each packet sent [i.16].
- **Type-P-one-way packet-duplication**, indicating the number of additional copies of an individual packet received by the destination in the time interval [i.16].
- **Type-P-one-way-Packet-Duplication-Poisson-Stream, Type-P-one-way-Packet-Duplication-Periodic-Stream** [i.16].

5.4 Parameters Related to Connectivity and Service Availability

- **Percent IP service unavailability (PIU)**, the percentage of total scheduled IP service time that is categorized as unavailable. Availability is declared if the IPLR is smaller than 0.75 [i.6].
- **Percent IP service availability (PIA)** [i.6].
- **Type-P-Instantaneous-Unidirectional-Connectivity**, measuring whether a packet reached its destination or not [i.18].
- **Type-P-Instantaneous-Bidirectional-Connectivity** [i.18].
- **Type-P-Interval-Unidirectional-Connectivity**, if there is availability somewhen in an interval [i.18].
- **Type-P-Interval-Bidirectional-Connectivity, Type-P1-P2-Interval-Temporal-Connectivity** [i.18].
- **Multicast group IP service availability**, defined as the ratio of destinations in the available state and the total destinations [i.8].

5.5 Throughput Related Parameters

- **IP-layer Bits Transferred, IP-layer Link Capacity, IP-layer Path Capacity, IP-layer Used Link Capacity, IP-layer Link Utilization, IP-layer Available Link Capacity, IP-layer Available Path Capacity, IP-layer Tight Link Capacity [**i.19**]**.
- **Bulk Transport Capacity**, measuring the expected long term average data rate of a single ideal TCP implementation over the path in question [i.17].
- **Point-to-point IP packet rate (IPPR)**, measuring the total number of IP packet transfers per service-second [i.8].
- **Point-to-point octet-based IP packet rate (IPOR)**, the number of octets in the IP packets resulting in IP packet transfers per service-second [i.8].
- **Multicast group mean packet rate**, as the sum of IPPR for all destinations divided by the number of registered destinations [i.8].
- **Multicast group mean octet-based IP packet rate, One-way packet rate range over multicast group** [i.8].

5.6 Operational Real-world KPIs

As explained before, KPIs depend on what a business or network administrator considers "Key" for their particular objectives. This situation creates KPIs in the Information Technologies world that extends from the network parameters presented to service availability or performance, device management issues, post-processed results, etc. As an example, some KPIs present in use in medium-large networks are:

Availability parameters:

- Availability of key network nodes and links.
- Availability of the network core measured as the percentage availability of a set of critical links.
- Availability of key servers (percentage of time): directory servers, email servers, calendar servers, web servers, application servers, etc. Availability can be measured from the internal network, from the exterior, from key network locations, etc.
- Availability of Internet access: measured by contacting key selected Internet servers, by checking connectivity with adjacent ISP router, by checking reach ability to a set of destinations, etc.

Performance parameters:

- Average times: time to login, time to deliver emails between exchange servers.
- Percentage of requests to key servers that are resolved in less than a maximum time.
- Average (or maximum) utilization on key network links.
- Average (or maximum) utilization of key servers.

6 Information Models Requirements for IP Traffic Monitoring Applications

This Work Item benefits from the activities of the previous Work Item on the analysis of ontologies for IP traffic measurements and the associated document [1], while it leverages realistic use cases for extracting practical requirements for the ontology of IP traffic measurements. In order to complete the conceptual model, Quality of Service, as perceived by the users, elements and complex parameters to quantify business issues related to service level agreements are also included so that future developments, based on objective TMA (traffic monitoring and analysis) systems, can be used to such purpose.

6.1 Use Case Scenarios

This clause illustrates the application of KPI and IP network descriptors to manage practical situations in order to extract conceptual relationships required to deal with them within a universal and coherent approach. Obvious parameters requirements are skipped so that the focus is put on characteristics that might give rise to lack of interoperability between network characterization mechanisms.

6.1.1 IP Networks Characterisation

Aside from the general concepts of time, address, hops, neighbouring, packet size or the more specific ones related to session parameters, routing, synchronization, etc., a MOI ontology should be open enough to characterize any IP service through simple traffic measurements carried out either by the network elements (hosts and routers) or probes. In the following, a series of use cases are presented to extract concepts and relations that should be taken into account just to describe an IP network:

a) *End to end routes characterization*: The user asks the system which routes can be used to connect two nodes of the network and wishes to know their characteristics. This brings elemental objects and topology parameters like host, router, node (address), neighbour, hop counting, length, traceroute, AS (autonomous system), Topology concepts are essential in IP networks and MOI should deal with them following classical conceptual models: node-link-route, length, routes passing a node. Furthermore, additional concepts like port, socket, ring, tree, source and destination should be added to construct the MOI.

b) *Network topological characterization*: Similar concepts may be required by different questions like:

- Maximum number of hops between two nodes of a given network (domain).
- Number of direct neighbours to a given node.
- Nodes in a tree.
- Distribution of paths per length in a given domain.

This extend the concepts of node, address (with a semantic address notation) to parent-son dependency, net, sub-net and attributes to specific nodes like gateway, server, probe, etc. Furthermore, MOI would include rules for format exchange as far as all these objects and attributes concern in order to increase the ontology usability; for example, IP address can be stored in dot notation, as integer or any other format.

Some details like the required distinction between node (host, gateway, router, etc.) and IP address should be highlighted and need the corresponding ontological differentiation. Besides, IPv6 should also be considered for the MOI set up as far as covering and translating concepts like subnet or the simple address notation concerns.

6.1.2 QoS Measurements in IP Networks

Traffic monitoring aims at supervising network services indeed. Analysing the network topology can be used to understand what may cause connectivity troubles and so help managing the services but their performance has to be checked by direct measurements. Devices to accomplish such task should use the same data to support coherent operation systems. Such interoperability relies on a common MOI compliant with use cases like the following ones:

a) *Bandwidth availability monitoring*: The user asks the system to check the bandwidth availability for a given connectivity service between two hosts, either in the same domain or in different ones. This complements a (static) vision of the network to show not only its topology but its capacity link by link and serves to monitoring (dynamically) its behaviour. The tools to accomplish such task and the accuracy to present data may vary but the conceptual entities are simple: One way and round trip bandwidth available end to end (e2e) in uplink and downlink direction.

   Related to this simple use case, for operational purposes, the user of an IP monitoring system could be interested in:

   - Determining the flows maintained through links of a given path.
   - Knowing the percentage of the capacity e2e that is used by established flows between the two ends or by specific applications. This capacity analysis requires measurements and data mining on objects like flow, and protocols related to upper layers in the OSI stack (session and application, at least) to care about connection oriented services if the MOI is going to provide QoS measuring support.

Thus such objects and their measuring should be included in the MOI namely flow, protocol and attributes to characterize packets as part of a flow or multicast service, whether it is an ACK, ERROR message, signalling packet or if it has been sent to duplicate a previous (failed) one.

b) *Transport quality measuring*: IP networks deal with packets on the "best effort" assumption but network operators should respond for the complete service of supporting applications either oriented to connection or connectionless, in real time or with time elasticity and possible error recovery. Therefore, a MOI has to include concepts to support and give interoperability to devices and repositories in cases like:

- Determining the delays of packets sent from one (source) host to another one (destination). This time stamping issue is really relevant in QoS measurement; its statiscal treatment is also very important and the possibility of setting alarm levels too: All statistical concepts should be part of the MOI as well as thresholds and alarms derived from different entities like mean value and variance. Besides, MOI ought to distinguish between one way delay (OWD) and round trip delay (RTD) for e2e connections.

- Monitoring the packets loss rate as well as the actual throughput for a given socket: These are important entities for QoS in MOI.

- Describing a session by statistical distribution of packets. In other words, analysing a TCP performance for a given e2e connection that can help users to diagnose connection troubles that should be treated in the ends or can be caused by network QoS problems. For this, a distribution of packets size, ACK and retransmissions is essential and so these concepts, derived from the root packet are.

- Additionally, QoS for connection-oriented applications require measuring when sessions are dropped. This involves not only the concept of session but also its set up and drops down. Besides, the special delay for the initial set up should be considered as particular object to respond to the "time to service" indicator that indeed responds to quality of experience or the perceived QoS (see clause 6.2).

Aside from the requirement of MOI to include the object probe to support descriptions of the QoS management system itself, it is worthy to consider that MOI will need to include QoS classes for a complete traffic description.

6.1.3 Traffic Monitoring for Security Applications

A promising application of IP traffic monitoring is enabling network operators and service providers to develop systems to detect attacks for fraud or hacking their clients by detecting anomalies in the traffic pattern or other peculiar user's behaviour. Hence, MOI has to cope with the expected added parameters and notions issued from traffic monitoring systems so that repositories and control systems can exchange information coherently.

Most security applications use encryption and authentication mechanisms that require no more entities than others. In fact, the main concern for MOI about security issues is related to privacy protection since security, integrity and privacy are three topics that need to be worked in equilibrium. However, MOI should be aware of intrinsic difficulties of security and authentication solutions that may create bottlenecks by increasing spurious traffic, flooding, denial of service (DoS) attacks, fraud traffic (either from false address or zombies) and even more "irregular" routers. Network-centric strategy to detect and deal such troubles can be implemented through TMA techniques and so MOI will have to include appropriated concepts to support the expected information exchange that can be illustrated with the following use cases:

a) *Secure information exchange between nodes*: The user wants to supervise whether his connections are hacker-proof once he has provided them with authentication mechanisms: From a TMA approach, this task requires systems that exchange information about the established connections to a given host (server, for instance) or any other data repository; such information (and the tools to extract it) should deal with concepts like:

- Kinds of hosts and their roles: Server, gateway, proxy, router, client, public authority (public keys provider), etc.

- Types of messages as far as security issues concern: Some examples have already been mentioned like ACK and protocol-related messages. For sake of security, encryption keys are relevant enough to deserve a place in the MOI.

b) *Confidentiality of the information exchanged*: The user wishes to certify the mechanisms deployed to preserve some communications between his clients. Normally such opacity to outsiders is achieved by means of tunnels or VPN (virtual private networks): Information is carried out without major encryption/decryption procedures (that are high CPU consumers and slow down sessions) once privacy of the transmission is assured. The TMA approach to check this out relies on anomalies detection which, in turn, implies MOI to deal with concepts of:

- Geolocation: latitude, longitude, city and country or simply Autonomous System (in IPv4).

- VPN, as extension or attribute to specific network sections (not included in the topology class).

6.1.4 Autonomic Network Monitoring and Management

The increasing size, complexity and the dynamic character of future networks make traditional monitoring systems inadequate to be continuously updated and sense the endless changes in the topology and communication conditions. In a dynamic environment with frequent topology changes and link failures, centralized approaches suffer from low service operational quality –e.g. performance is reduced when connectivity is not feasible to the central node-, have limited scalability –e.g. data is stored in a single node and service degradation is expected by increasing the incoming requests-, and resilience –e.g. the central node is a single point of failure.

The need for reduction in the network management complexity and the administrator's operational burden imposes the design and implementation of self-functionalities and the adoption of self-management schemes. Autonomic mechanisms have to be designed to control traffic monitoring within a network based on the distribution of roles and points of decision within the network. Traffic monitoring functionality is usually assigned to predefined nodes in the network that are responsible to collect statistics in the packet or flow level. However, in dynamic environments, this approach is not able to adapt efficiently to network changes. Thus, traffic monitoring functions have to be realized in a decentralized manner taking in account the trade-off between the imposed overhead and the performance of the traffic monitoring mechanisms.

However, the incorporation of autonomicity to communication systems requires the efficient representation of the available information and the extension of existing context models. Representation of context data can be realised through various formats from databases and XML files to more advanced formats such as information models or ontologies. The outcome of each representation scheme in autonomic networking is the extraction of knowledge through the efficient representation of data and support of self-monitoring, self-diagnosing, and self- adapting functionalities. Thus, it is crucial to select the proper representation scheme according to the requirements imposed by the network and the applications scope.

The basic concept that has to be described is that of the control loop that is necessary for the design of any autonomic functionality. The concept of control loop was firstly specified in the IBM-MAPE model [i.20]. A control loop is orchestrated by a Decision-Element that manages a Managed-Entity and can be applied in node and network level. The Decision-Element may trigger some behaviour, enforce a policy or exchange information with the Managed-Entities. By the design of several control loops in the network, autonomic functionalities may then be supported.

Taking in account these concepts, a representation scheme has to be developed that describes the interactions among the network entities for the support of autonomic functionalities. Information flow, management of network entities and conflicts avoidance are key issues that have to be described.

6.1.5 Law Enforcement

Traffic monitoring provides in certain cases the authorities with the means for law enforcement, e.g. for the purpose of public safety. In that respect, there exist specific regulations with respect to the Lawful Interception (LI) of communications. In the U.S.A., the Federal Wiretap Act has been enacted already in 1968 and since then has been adapted in order to include modern communication systems [i.21]. In the European Union, the Resolution on the Lawful Interception of Telecommunications [i.22] has explicitly recognized the requirement for the availability of the lawful interception means to the Law Enforcement Agencies (LEA).

Lawful Interception is the legally authorised process by which a network operator or service provider grants some law enforcement officials with access to communication data (such as, telephone or VoIP calls, e-mail messages, etc.) of private individuals or organisations. Lawful Interception is becoming crucial to preserve national security, to combat terrorism or other serious criminal activities, as well as to investigate these kinds of social mishaps. In the typical case, some LEA orders to a provider the delivery of data that constitute the product of network monitoring; they may refer to:

- Interception of the Content of Communication (CC).
- Collection of Interception Related Information (IRI).
- Collection of data for the enforcement of the data retention regulatory provisions.

IRI depends on the particular case of interception and may refer to a variety of data models. In the general case, the IRI shall contain:

- The identities used by or associated with the target identity, that is, the user subject to interception.

- The identities that have attempted communications with the target identity, successful or not.
- The details of services used and their associated parameters.
- Information relating to status.
- Time stamps.

However, the exact types and nature of the data comprising the IRI differs on a case-by-case basis, based on the type of the service in question. A variety of ETSI Technical Specifications define the different data types comprising the IRI and the associated data models. For instance, in the case of an e-mail send event, the following data constitute the IRI [i.23]:

- Server IP
- Client IP
- Server Port
- Client Port
- E-mail Protocol ID
- E-mail Sender
- E-mail Recipient List
- Total Recipient Count
- Server Octets Sent
- Client Octets Sent
- Message ID
- Status

Other ETSI Technical Specifications include [i.24], [i.25] and [i.26].

Regarding the enforcement of data retention, it should be noted here that there are additional data types that are requested, such as the name and the address of the user, as well as her/his location. Nevertheless, these data types are not directly collected through the network monitoring procedure but require additional "back-office" processing for their generation.

6.2 Requirements Derived from the Quality of Experience Concepts

The QoS, as perceived by the end users in sevices over Internet deserves an independent section despite its links to QoS. In fact, the correlation of QoS to QoE is a thorny though promising subject for business applications. TMA can help advancing in this field by means of smart application of statistics and machine learning systems based on traffic data that ought to be complemented with additional information to take into account:

- Types of application: This, in turn will be given by some kind of correlation with protocols, type of connection, characteristics of the messages, etc.
- KQI (Key Quality Indicator): Equivalent to the KPI to measure the perceived QoS, like MOS (Mean Opinion Score) and other particular parameters for different applications (IPTV, VoIP, web access, file transfer, etc.) like time to service, pixellation, resolution, aliasing, noise, etc. Obviously, the MOS concept is not analytically linked to TMA values and so it should be directly added to the MOI in order to extend it QoE management systems. Other concepts can be related to traffic measurements but should be simply included to stand by their own.

- Statistical service description: Since subjective opinions are dependent on users´ memory, the number and frequency of service failures are also important QoE descriptors to deal with and include in the MOI as extension of simple service availability, namely peak values of service withdraws or mean values of time to service that can support the use case of a service provider that has to supervise its performance versus users demands.

Quality of Service can be defined as the collective effect of service performance that helps to pinpoint the degree of satisfaction of a service user, including any performance issue. However, this is only the quality from the network or service point of view. Nevertheless customers may have a different perception of quality. This new concept of quality is known as Quality of Experience (QoE) or Quality of Perception (QoP, pQoS) and is focused on the subscribers. The evolution of quality management from the customer point of view shifts end-to-end quality services from subscriber perspective. This new outlook will allow identifying network degradation and performance results before affecting the customer, since QoS is just a technical concept that it is measured and understood in terms of networks. So it is a subset inside the QoE scope and it is noted that one of the goals of Quality Assurance will be delivering QoS through the user experience.

This way the Quality of Experience consists of a set of indicators that show the perceived satisfaction of using the service by the end-user. These indicators include a vast variety of parameters from the multimedia encoding domain, transport as well as the terminal on which the media is presented and finally the type of content the customer is using. The QoE ideally looks at the correlation of all these parameters to maximize the experience of the users while minimizing the resources of the provider.

Nevertheless it is not trivial to determine what is the quality perceived by the clients, because it will depend on a great variety of factors. The quality experienced by the customers depends on many factors: the components that set up the service, the business processes related to the service, the resources on which the processes are supported and the performance of the underlying network. With the aim of quantifying the perceived Quality of Service a SP should know the key quality (KQI) and performance (KPI) indicators for networks and services, and fulfil a methodology that interrelates any factor.

6.3 Ontology Requirements to Support Business Management Applications

On top of QoS operation and QoE supervision, business management still needs additional information and deals with complex concepts that can be included in the MOI for sake of integral systems support. The use case of addressing network and content resources to cope with the SLA signed with all kinds of clients in the chain value of telecommunication business may be complex but it is to figure out.

Unlike simple network operators, that just respond for connectivity services, integral operators need to know which kind of client (in both senses, the human and the service one) is affected by a traffic trouble. If this is to be achieved by means of TMA, the MOI has to provide information about:

- Segment of the network, namely whether a probe, host, etc. is in the core, metro, access or residential section of the e2e connection.

- Extended classification of network elements including service gateways (for residential networks), sensors, end devices (Set-top-boxes in IPTV, Softphones in VoIP, etc.) and private content producers (thus named despite they may be simple hosts for simple traffic monitorization).

The "self" paradigm in network management will also require importing new concepts like agent, autonomous system, etc. Furthermore, the virtualization of network elements could give rise to specific concepts but it is still too early to establish which ones. Thus far, it is just convenient to point out the concept itself of virtual element like an extra attribute that could, in principle, be applied to any element.

7 Additional Input for Privacy Protection Approaches

Intuitively, since network monitoring depends by default on the collection and processing of information, it raises issues related to the protection of personal data. Even more, network monitoring activities are in particular interesting compared to other domains as far as privacy protection is concerned, for a number of reasons:

- Privacy-sensitive information is not limited to the payload of the network packets, i.e. the content of the monitored communications. In fact, this case could be even considered as trivial from a privacy protection point of view, since the confidentiality of the content can be adequately achieved by using strong end-to-end encryption. The focus of passive network monitoring is on the collection of so-called context data [i.27]. In [i.28], such data are characterized as "semi-active", in the sense that data collection occurs transparently for the user; this type of transparent, implicit collection tends to raise greater privacy concerns than those initiated by the data subject [i.29].

- While the various protocols' headers already reveal much information (e.g. a visited web-site or the peers of a VoIP call), a huge amount of personal information can be further extracted from their processing, even if they have been anonymised. Literature has shown that once a flow can be examined in isolation, fingerprinting techniques allow deriving personal information from as little as the basic statistics (i.e. packet sizes and inter-arrival times' correlation) of the delivered packets [i.30], [i.31] and [i.32]. Moreover, as [i.33] demonstrated, SSL/TLS does not resist statistical traffic analysis, while meaningful data can be finally extracted even from "unsuspicious" header fields, such as the IP ID alone [i.34].

- The network monitoring activities, as well as the underlying categories of data, have been subject of specific regulations, such as [i.35] and [i.36] in Europe and [i.37] in the USA. Additionally, in many countries, independent data and communication protection authorities regulate and audit privacy protection in communications.

Indeed, the legislation plays a crucial role in the determination of data collection and processing policies in the context of network monitoring; the underlying principles originate not only from the data protection domain, but also from provisions related to public welfare, such as public security. In this context and with respect to personal data protection, legislation should be the starting point for obtaining the corresponding requirements to be reflected by an ontology for IP traffic. The milestone Directive 95/46/EC [i.38] plays a particular role, since it sets the fundamental basis. Moreover, in order to obtain the requirements related to personal data protection for a monitoring ontology, the features that should characterise monitoring systems and procedures should be considered beforehand, since one objective of such an ontology should be to support the enforcement of these needs [i.39]. A brief analysis follows:

- **Lawfulness of data processing:** A monitoring system should be able to evaluate the lawfulness of each request for personal data with applicable laws and regulations. This implies the assessment of 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 should be evaluated against the type of collected information and the purposes for which it was collected. It follows that the system should be configurable with a set of data types and purposes deemed to be lawful and, for these specified and preidentified data types and purposes, the system should allow the processing of the personal data. All such configuration of the system with respect to the lawfulness of stated processing purposes should 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 should be denied. If the request is concerned with a certain kind of network monitoring on the basis of the specific purpose of a monitoring activity, the system should be able to apply the other mandatory legal requirements. For example, the use of data in anonymous or identifiable form should be permitted or not permitted depending upon the specific monitoring function to be carried out.

- **Purposes for which data are processed:** A monitoring system should provide the means for identifying the purpose of each request, in order to comply with the "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, explicit and legitimate purposes. In addition, the system should prohibit that personal data collected for some specific and legitimate purposes are used for other purposes, incompatible with those for which the data have been originally collected. The purpose principle also implies that the controller should act transparently. This means that the controller should specify and make explicit to the data subjects the reasons why the personal data are used. 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 kind of negotiation with the party submitting the personal data processing request 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 subjects.

- **Necessity, adequacy and proportionality of the data processed:** A monitoring system should 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 are collected by the system. The system in practice should be able to determine the amount of personal data that may be processed within a specific monitoring function and the type of data may be processed within the same function. 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 data subjects. 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 make anonymous any data that are redundant or no longer needed for a specific monitoring function.

- **Quality of the data processed:** A monitoring system should ensure that the data processed are correct, exact and updated. Moreover, the system should be able to perform corrective actions in order to delete or correct inaccurate data and to delete or update data that are outdated or redundant. In addition to these corrective remedies, the system should also allow periodic audits on the personal data that it stores, so as to verify the legitimacy of said data.

- **Minimal use of personal identification data:** A monitoring system should 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 case of mandatory data retention obligations under the Directive 2006/24/EC [i.36].

- **Storage of personal data:** A monitoring system should keep personal data in an identifiable form only for the time that it is strictly necessary to the specific 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.

- **Data retention:** A monitoring system should comply with the requirements set forth by applicable data retention regulations. This implies that the system should store the specific data that are subject to the data retention regulations 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 applicable legislation. It should be stressed that compliance with data retention law requirements implies additionally that the system should fulfil specific and mandatory security requirements to be applied for the storage of the data and relevant access. For example, the data stored for data retention purposes should be kept logically separated from the other data stored by the system.

- **Access limitation:** A monitoring system should authenticate all users of the system, should provide different levels of access to the stored data and should provide for the logging of all access to the stored data in order to detect attempted or successful unauthorized access. These levels of access should be granted based on the authentication of individual users, the need to know associated with each individual user's role, as well as the types of 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.

- **Information to and rights of the data subject:** With regard to the requirements relating to providing to the data subjects adequate information on the purposes, conditions and features of the data processing, as well as to the requirements relating to offering to the data subjects the possibility to obtain information on their data and to actively intervene on the data processing enforcing privacy rights such as the rights to access data, ask for data updating, integration, deletion and others, the core of the matter is that these requirements may be fulfilled only by the entity having direct contact with the data subjects. The scenario would change according to the entity that performs the monitoring. In case the network provider itself performs monitoring on its customers, the provider should comply with applicable legislation with regard to information to the data subjects and enforcement of their privacy rights. In case the monitoring is performed by an entity having no direct contact with users, these requirements should be addressed by negotiation between the network provider and the entity that intends to perform the monitoring. The subset of mandatory information that the data subject should receive varies across different jurisdictions, 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 entities in charge of the data processing acting as data controller and data processor. As to the privacy rights of the data subjects, we may recall for example that the data subject should be provided 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 or 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 to object the processing of his/her personal data for legitimate reasons.

- **Consent of the data subject:** For compliance with the consent requirements, please refer to the above comments. The monitoring system should 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 expressed by the data subject. The data subject should be enabled 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). 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 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 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).

- **Data security measures:** A monitoring system should adopt appropriate technical and organizational measures with the purpose of protecting the personal data that are collected and processed 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, as well as 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 should be able to ensure an adequate level of security. The adequacy should be assessed having regard to the risks represented by the nature of the personal data to be protected and the processing 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 system subject to these laws should 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 should be duly informed about said concerns.

- **Special categories of data:** A monitoring system should 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 the processing of traffic data and location data, which are of particular interest for passive network monitoring, the Directive 2002/58/EC [i.35] 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 the case where 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. For these requirements relating to information to the data subjects and consent of the data subjects, please further refer to the clauses above. The Directive 2002/58/EC [i.35] also imposes specific security requirements for the processing of traffic and location data. For instance, 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, while it should 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 behaviours and individuals' profiles. Overall, the monitoring system should 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).

- **Coordination with competent Data Protection Authority:** A monitoring system should monitor compliance with the notification requirement and with the provisions of the authorizations of the competent Data Protection Authorities, as ruled under applicable data protection legislation. Moreover, the system should allow communications between the system and the competent Data Protection Authorities in order to validate and verify that the notification and/or authorization requirements have been duly complied with. This kind of interaction with the competent 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 Data Protection Authority to verify accomplishments of the due legal conditions.

- **Supervision and sanctions:** A monitoring system should provide the competent 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 Data Protection Authorities encounter difficulties in auditing the processing of personal data carried out through technical means and over the Internet; this is due to the peculiar nature of the technical means deployed, that allow the hiding of the data processing activities performed. The system would not act as an enforcement authority, since it would lack the necessary competence; instead it should provide information to the competent 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 monitoring system and the competent Data Protection Authorities, so that the system provides the aforementioned Authorities with a log of data processing activities performed.

- **Communications confidentiality and Lawful Interception:** A monitoring system should 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. Therefore, the monitoring system should guarantee confidentiality in the communications, but should also be able of complying with the lawful interception requests coming from the competent 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. The monitoring 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 data 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 means of retention and communication of the personal data representing the subject matter of the interception.

- **Flexibility and adaptability of legal compliance provisions:** Given the complexity of the legal environment in which a monitoring system operates, the different legal requirements across different jurisdictions and the nature of the law to change from time to time, the system's design should to the extent possible be flexible and adaptable with respect to all the provisions described in this clause. Specifically, the system should encode as much of these provisions in dynamic policies.

8 Main Features and Global Requirements for MOI

8.1 Data Types Support Requirements

Formal data serialization languages, such as XML, have a predefined set of allowed data types. Although this set is fairly big and covers the different notations for different types, it normalizes the data in terms of its syntactical structure and not in their semantics. In a Measurements Ontology not only the information about the data type needs to be modelled, as could be done using xsd types, but the semantics of the data need to be established within the ontology to be able to share and maintain the information across application or business boundaries.

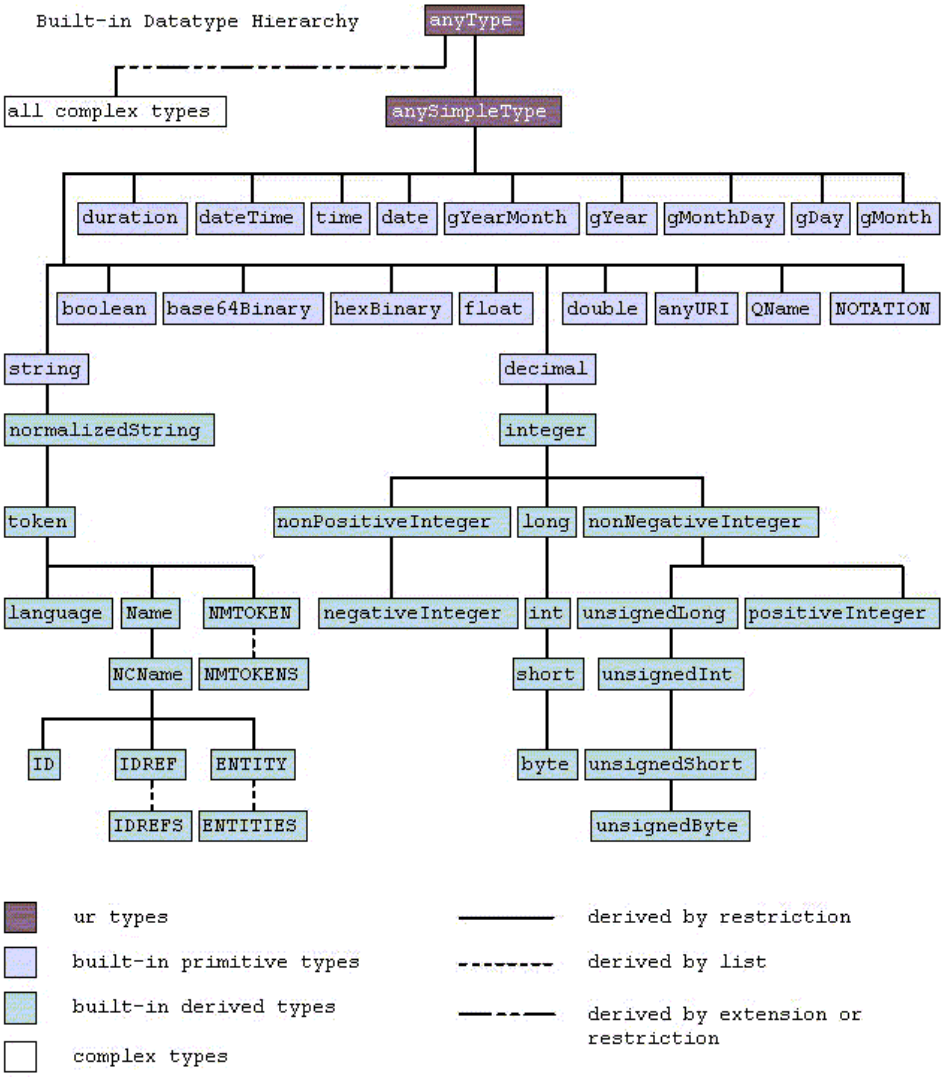

Figure 1: XSD types from W3C XML Schema [i.40]

This is mainly caused by the fact that measurements of the same metric, can be expressed in different ways (different units, i.e. seconds or microseconds) but all of them being the same xsd type (i.e. xsd:decimal). Therefore an important part of the ontology should describe the units of the measurements and the relations between them. This process of enhancing the information contained in the data-types, being able to express the unit and the representation format used for the measurement is called augmentation.

This augmentation of the types can be introduced in three different ways:

- Creating new data types, extending the semantics of the existing ones (i.e. xsd) and use them as standard data types (i.e. "1"^^MOI:Megabitpersec). This way the semantics of the value, the unit information, is directly attached to the value. This way the ontology does not require any additional structure to hold this information. The main disadvantage of this approach is that applications not supporting MOI data types, only xsd ones, won't be able to process these data and process it as a simple xsd data type.

- Creating only a new data type, quantityAndSymbol and to include the unit symbol, which should be standard, inside the literal. "10 Mbps"^^xsd:quantityAndSymbol is an example on how to include the unit inside the literal. This creates a new disadvantage which is that applications need to parse and process those literals in a different way than usual XSD types.

- The other option is to create containers, new classes, for the measurement values. This way any related information of the measurement values can be expressed as properties of the container. This approach increases the complexity of the ontology because we are introducing new concepts but applications which are unaware of MOI ontology will recognize the data as valid xsd types, and those applications using MOI can extract useful information from the container.

Several approaches exist to design measurement unit ontologies (e.g. [i.41], [i.42] and [i.43]), but all of them are related to Physics measurements and therefore cannot be directly imported to the MOI ontology. All of these ontologies use the second approach to augmentation, they create new classes to represent the Unit concept, and model the relations as the factors to transform between units.

The only reference to Units for network measurements is the MOMENT Units Ontology [i.44], which is used in the MOMENT project to provide a unified framework in which automatic unit transformation can be made in terms of the semantic relations.

The MOI should then accommodate existing data-types from the language which is used for its serialization as well as create new concepts, referred as containers, to be able to extend the semantics of those values, including the unit and any additional information which is needed.

8.1.1 Requirements for Application-specific Data Types

Representation Formats

In network measurements, not all collected data is numeric (i.e. "13" Megabits per second) but some parts of the measurements are usually expressed in a human readable format (i.e. Ip addresses in dotted format). Those concepts do not have a unit, in the same sense that numeric measurements have, which we are able to transform to related types using numerical factors.

With IP Addresses, MAC Addresses and other non-numeric measurements a concept is necessary to specify the format of representation which is being used for the measurement, or it will be impossible to compare and analyze different data in different formats. This format representation could also apply to numerical values if decimal bases are not always used for representation. A clarifying example: "10"MBps,"A"Mbps and "ten"^^lang:en Mbps are in the same unit, the only difference is the representation format for the numerical value (decimal base, hexadecimal base and natural language in English).

As the number of different representation formats for the same value is not manageable, the MOI should define the standard or default unit and representation format for each concept included. It would be also useful to include other formats and include the relations between them, if possible. For example, in IP addresses, dotted notation is as popular as integer notation and therefore choosing which one is the default one, discarding completely the other would mean that a lot of information is going to be lost, or very difficult to integrate. In such cases where the default unit or representation format is not clear, providing the mechanisms to relate equivalent values in different representation formats is a key issue to be addressed in MOI.

Statistical operators

Many of the network measurements are generated from automated tools and programs which perform periodically a set of predefined measurements. After that, if those measurements are used to perform statistical analysis of the network, parameters such as the time between measurements, the number of measurements, etc. are useful information related to the measurement that should be modelled into the ontology. Moreover, special units to represent the results of those statistical experiments such as: ratio, percent, percentile, etc. need to be included to be able to compare and compute the results of those statistical measurements.

Very few work on this aspect has been done in the semantic web. Projects such as [i.45], [i.46] and [i.47]perform a preliminary analysis on how to model and include statistical information in an appropriate way to share it on the web. Although the basis has been established, there is no information on how to relate the raw measurements to the statistical measurements, and the statistical operator which were applied to obtain the statistical measurement. A first approach is included in [i.44] where statistics are concepts that applied to raw measurements create new measurements. This concept can also be useful for other non statistical measurements, but ones that are created by the information of other measurements. This process can be seen as a cause-effect trigger, which useful in modelling events and alarms in the ontology.

Application parameters

When a network measurement tool is used, usually it consumes a set of key-value pairs, known as parameters which modify the behaviour of the algorithm to perform the measurement. In some cases those are directly related to the individual measurement itself, but in others they are related to network or system parameters which are needed to be able to perform the measurement in certain situations.

In those cases where parameters affect directly the measurement, they are a key aspect of the performed measurement, and therefore should be included with all the related information. Those application parameters can be categorized in the following clauses:

- *Application configuration*: these parameters include interface numbers, memory requirements, disk requirements, option selection, input and output configuration, etc.
- *Measurement configuration*: these parameters include measurement attributes, measurement thresholds, measurement policies, measurement objectives, etc.
- *Planification configuration*: including start time of the measurement, estimated end time of the measurement, number of iterations of the measurement, etc.
- *Previous information*: including network measurements from a previous measurement used by the tool or algorithm to adapt or configure its internal behaviour.

8.2 Operational Requirements

From the point of view of the data owner: the format, unit and representation of the information will be the one that best suits it's own application. In most cases this information will be stored in relational databases or log files which are not directly transformable to OWL/RDF or any other ontological serialization language. Several projects are tackling the problem of publishing existing information to the semantic web as RDF, and will benefit from the common information model that the MOI ontology can provide.

Therefore the ontology should accommodate the structure of previous information models for Network Measurements. As it has been analyzed in [1] most of the information models use either a tabular approach or a OOP (Object Oriented Paradigm) approach. The first fits very well to large repositories, as storing and retrieving the information is critical, it leverages the capacities of existent relational database technologies. The second one leverages the technologies from UML (Unified Modelling Language) and dynamic code generation when the transmission or processing of the information are the key aspects.

Ontologies extend the OOP approach, giving the semantics to the concepts and relations. This way existing UML information models can be migrated to their ontological counterpart without further problems. For relational databases and tabular data, the heterogeneity of solutions applied when designed, make impossible to provide a fixed structure which will fit any network measurement relational database. Therefore the ontology should be able to adapt, and reproduce the structure of a generic table, with a variable number of columns.

8.3 Requirements for Integral Privacy Protection Provisions

Having as starting point the high-level legal and regulatory provisions that have been outlined in clause 7, this clause provides a codification of the specific requirements characterizing an ontology for IP measurements. The identified requirements are the following:

- Any data abuse presupposes access to the data; therefore, MOI should provide the means for the specification of access control policies, or be integratable with access control mechanisms in order for effectively regulating access and minimizing the danger for illicit use.
- The policies regulating data management should be characterized by flexibility, in order to meet the need for adaptability to different data protection jurisdictions, as well as to be easily and seamlessly adapted to changes of laws and regulations.
- The semantics of personal identifiable information determine the norms for its use; indeed, each data item should be treated according to its particular type. In that respect, MOI should provide the means for the semantic categorization of personal data based on their type.
- The requirement for the specification of the semantic type of information is complemented by the requirement for the organization of the personal data types. This implies the taxonomy of data so that the associated relations (e.g. AND and OR) to be able to be specified.
- The notion of "purpose" (behind data collection and/or processing) plays a critical role in privacy protection; therefore, the MOI should provide the means for the explicit specification of the purposes for which some collection or processing of information takes place. In this context, purpose can be seen as the detailed determination of the underlying monitoring task or activity.
- Intuitively, since monitoring purposes can be characterized by complex relations between them (e.g. a purpose can be the composition of sub-purposes), the means for the specification of such relations should be provided by MOI.
- MOI should appropriately treat special categories of monitoring purposes that are activated for reasons of law enforcement; this includes Lawful Interception activities.
- Apart from the semantics of data and purposes, MOI should include or be integratable with third models for the incorporation of other contextual parameters that ultimately should be taken into consideration when taking decisions for data management. A fundamental parameter is the role of the entity that executes a monitoring task; i.e. the policies defined by MOI should be role-based. Additional contextual parameters include (but should not necessarily be limited to) the following: temporal parameters; spatial parameters; history-based parameters; parameters related with the age of the data.
- Regarding the need for a role-based model as described above, there are some roles the importance of which should not be neglected; these role include the data subjects themselves, as well as the entities of the competent Data Protection Authorities and the Law Enforcement Agencies.
- Since the retention periods of data have emerged as an important regulatory requirement, the means for the specification of such periods should be provided by MOI.
- In addition, to the data retention specification requirement, MOI should enable the description of the necessary action to be executed upon the expiration of the data retention period, e.g. the automatic deletion of the data or their anonymization following some anonymization pattern.
- MOI should enable the definition of anonymization strategies to be applied to data prior to their disclosure or being stored, or in any other case required.
- MOI should provide the means for the specification of "privacy obligations". This implies actions that should take place before or after an event, such as collection, disclosure, storage, processing, etc.
- Since security constitutes the bottom-line for privacy protection, MOI should enable the specification of the appropriate security measures that should be taken for the protection of the data under consideration. This includes the association of the data with the protective mechanisms that should be applied both at the communication phase and during their storage.

9 Ontology Architecture and Structure Requirements

This clause presents the set of requirements for the architecture of the MOI ontology to set a coherent ontology within the established limits of its practical application, but also open to further extensions or links to other information models, as expected for Future Internet development.

9.1 Requirements of Expandability

As stated above, it is necessary that the MOI ontology can evolve as new developments are put in production networks. The use of ontology languages solves this issue with the use of the following capabilities:

- Concept inheritance. This is the main way of extending definitions. A child class would extend the already defined information, making more specific the information defined previously in the parent class.
- Intersection, union and complement. It is possible also to use AND, OR and NOT operators to define new classes that are intersection, union or complement of already defined classes.
- Properties as "first class" elements. Most ontology languages define properties as "first class". They can be defined outside of a class, and later, specify one or more classes as the domain of a property. This will let a specification of properties of an already defined class.
- Ontology import statements. Ontology languages such as OWL allow the definition of import statements to reference an ontology containing definitions, whose meaning is considered to be part of the meaning of the importing ontology. New ontology defininitions can use this import clause to leverage previous specifications.
- Ontology versioning and compatibility. Ontology languages such as OWL can include a facet to describe the version of a class or a property. In this way, different versions of a definition can coexist if necessary. At the same time, it is possible to specify if these versions are compatible or not, and even if a definition is deprecated.

9.2 Requirements of Interoperability

The MOI ontology should be able to include information about other sources of information that have already specify network monitoring information. It should also support synonyms if necessary (e.g. host/node/station or gateway/router) For these requirements, it is possible to use the following capabilities of the semantic web languages (OWL, RDFS):

- Definition of similar classes and properties. Languages such as OWL let define equivalent classes and properties, which is useful to state that their descriptions have the same extension. That is, they contain the same set of individuals. Equivalent classes and properties have the same values but may have different intentional meaning. Then, it is possible to define equality, where two references actually refer to the same thing. This equality (sameAs in OWL) can be used to define mappings (see below).
- Human-readable names. Sometimes, it is not necessary to define equivalent or equal classes and concepts, and it is just necessary to have several labels for the same construction. This can be useful for synonyms such as node/host/station. For this, the ontology language should let define several labels per concept or property.
- Specify the original source of information. Many definitions of the MOI ontology will come from already defined specifications. In these cases, it will be useful to have annotations for these definitions, including in the ontology information to know their original source. RDFS constructions such as seeAlso or isDefinedBy can be used with this purpose. The former provides additional information about the subject resource whereas the latter specializes the former to indicate the resource defining the subject resource.
- Mapping ontologies. As stated before, it is possible to specify that some classes or properties are similar. This can be useful to define mapping rules between the MOI ontology and another information model. Sometimes, the mapping rules are more complex than defining two concepts as equal. In this case, it is necessary to use mapping ontologies which can express other type of relationships between classes and properties from different ontologies.

9.3 Requirements of Ontological Processing Performance

If the information is defined using ontologies, it is necessary to know which is the performance toll that has to be paid. Several performance issues have to be taken into account:

- *Memory*: ontologies are usually represented as graphs. This can consume a high amount of memory. Moreover, network measurement will have millions of instances. Thus, it is necessary that the ontology can be stored in backends that deal with this issue. Hopefully, there are backend implementations that currently solve this problem.

- *Reasoning and query delay*: provided that a MOI ontology can have millions of instances, it is necessary that backends can be queried, providing answers in a reasonable time. There are implementations, such as OWLIM [i.48], that are working on this issue to provide results with low delay.

Annex A (informative): Authors & contributors

The following people have contributed to the present document:

Rapporteur:
Dr., Georgios, Lioudakis, ICCS/NTUA

Other contributors:
Prof., Jorge, Lopez de Vergara, UAM

Mr., Giuseppe, Tropea, CNIT

Dr., Daniel, Morato, UPNA

Prof., Javier, Aracil, UAM

Mr., Alfredo Salvador, UAM

Mr., Felix, Strohmeier, SRFG

Dr., Angel, Ferreiro, Telefonica

Mr., Antonio, Cuadra, Telefonica

Mrs., Francesca, Gaudino, Baker & McKenzie

History

Document history		
V1.1.1	July 2012	Publication