QUALITY OF SERVICE IN THE FUTURE INTERNET

Jorge Carapinha[1], Roland Bless[2], Christoph Werle[2], Konstantin Miller[3], Virgil Dobrota[4], Andrei Bogdan Rus[4], Heidrun Grob-Lipski[5], Horst Roessler[5]

[1]PT Inovação, [2]Universität Karlsruhe (TH), [3]Berlin Institute of Technology, [4]Technical University of Cluj-Napoca, [5]Alcatel-Lucent

ABSTRACT

Whatever the Network of the Future turns out to be, there is little doubt that QoS will constitute a fundamental requirement. However, QoS issues and the respective solutions will not remain unchanged. New challenges will be raised; new ways of dealing with QoS will be enabled by novel networking concepts and techniques. Thus, a fresh approach at the QoS problem will be required. This paper addresses QoS in a Future Internet scenario and is focused on three emerging concepts: Network Virtualization, enabling the coexistence of multiple network architectures over a common infrastructure; In-Network Management, improving scalability of management operations by distributing management logic across all nodes; the Generic Path based on the semantic resource management concept, enabling the design of new data transport mechanisms and supporting different types of communications in highly mobile and dynamic network scenarios.

Keywords— QoS; Network Virtualization; Generic Path; In-Network Management

1. INTRODUCTION

Quality of Service (QoS) has played a relatively minor role in the innovations proposed in the framework of Future Internet research initiatives. Paradoxically, the importance of QoS, in a wide range of application scenarios, is likely to become more crucial than ever before – pervasiveness of network-based applications and emerging trends, such as cloud computing, will contribute to exacerbate the need for QoS and to make network performance more crucial than ever before for an increasing number of application scenarios. In fact, the capability to guarantee deterministic QoS is usually included in the "wish list" for the Future Internet.

In spite of considerable research effort devoted to QoS technologies in last few years, QoS mechanisms have been deployed in large scale commercial environments to a limited extent. Part of the problem lies in the complexity of implementing QoS models, which often encourages network operators to use "brute-force" solutions based on resource overprovisioning.

While best effort can be considered good enough in many cases, it is clear that for a lot of highly important applications, either for their potential business value (e.g. voice, interactive video), or for their role in crucial aspects of social welfare and quality of life (e.g. telemedicine, security), strict fulfillment of QoS parameters is a crucial requirement.

Several recent trends have had a significant impact on QoS. Overwhelming growth of P2P traffic has had a disruptive impact on ISP network backbones. In addition, the increase of access network capacity (which traditionally represented the main bottleneck), combined with the explosion of video-based applications, tends to provoke a huge traffic growth in core networks which will likely be exacerbated in the near future [5].

On the other hand, the adequacy of the original Internet design principles to cope with future Internet requirements has been put into question in the last few years. Despite the huge success of the Internet, largely enabled by the simplicity and scalability properties of the IP protocol, it is becoming increasingly clear that novel ideas and fresh technical approaches are needed to fulfill the requirements of future applications and services. Inevitably, this will have a major impact on how QoS can be provisioned, managed and controlled.

Traditionally, QoS is handled by the proper combination of network resource provisioning with techniques such as admission control, packet scheduling and active queue management. In relatively static networking environments, this kind of approach is adequate to control QoS in most cases. However, in the future this is likely to be challenged by higher elasticity, dynamicity and scalability requirements.

There is no doubt that QoS will remain a fundamental requirement in the Future Internet. But it is also clear that QoS is a moving target and new challenges will be raised by emerging trends, for which the traditional QoS tools may be no longer adequate.

The FP7 4WARD Project [1] has addressed key challenges of dynamic and scalable internetworking posed by the Future Internet and aims at creating a new architectural approach, more flexible and better adapted to present and future requirements [7]. Innovative networking concepts have been developed, referring to different views or behavioral aspects of the network, and paving the way to new approaches for provisioning, managing and controlling traffic and network resources.

The following sections are focused on three major concepts developed by 4WARD and analyze their potential impact on QoS, both in terms of new challenges and new capabilities:

- Network Virtualization to enable the deployment of multiple networks and architectures over a common infrastructure and foster the emergence of novel Internet paradigms;
- The Generic Path based on the semantic resource management concept to enable the design of new data transport mechanisms in order to flexibly establish and manage connectivity, supporting different types of communications in highly mobile and dynamic network scenarios;
- In-Network Management to simplify and improve scalability of management operations by pushing management intelligence into the network and distributing management logic across all nodes.

A detailed discussion of these concepts is beyond the scope of this paper. Readers are advised to look for relevant information in [2], [3], [4], respectively.

2. NETWORK VIRTUALIZATION

Network Virtualization is the concept of sharing physical resources (the substrate), i.e. nodes and links, in order to create virtual networks (VNets) on top of this shared infrastructure. 4WARD considers Network Virtualization in a competitive environment [13] and especially considers shared infrastructure in an inter-provider setting. Within VNets, arbitrary network architectures, which are not necessarily based on IP, may be deployed. Figure 1 shows an example VNet instantiated on top of the physical resources of three Infrastructure Providers with some end users attached to the virtual network.

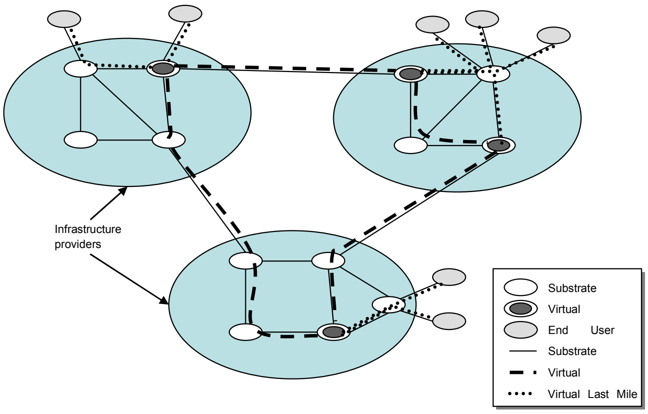

Figure 1. Basic Network Virtualization Scenario

This example already implies that there are at least two levels of QoS involved in network virtualization:

1) QoS at the substrate level

At the substrate level, QoS mechanisms are required for multiple reasons. From a security point of view, a well-defined degree of isolation has to be provided between virtual networks in order to prevent VNets from mutually affecting each other adversely — be it on purpose or inadvertently. Therefore, a differentiation between traffic of different virtual networks is necessary and the associated resources need to be isolated from each other to a certain degree. A stringent resource isolation requires resource reservation before use as well as policing during use. Thus, before the creation of a new virtual network, admission control is performed and the required resources are reserved for use by the virtual network afterwards. While the virtual network exists, usage of these resources needs to be policed accordingly in order to guarantee the SLA negotiated for this virtual network. These SLAs can comprise, e.g., the three classic levels of QoS assurance depending on the actual requirements on the VNet:

(a) Best effort: A virtual network is asserted only best effort service as usual in today's Internet. Virtual networks in the best effort category would then share with each other the resources remaining from higher QoS classes. This behaviour may be acceptable for some virtual networks that are not too sensitive to QoS parameters and that need to be operated very cost-efficiently. Real isolation is not provided and thus a DDoS flooding attack on one best effort VNet may adversely affect other best effort VNets on the same physical link.

(b) Statistical Multiplexing: In order to allow the Infrastructure Providers, which operate the substrate, to make efficient use of their resources, they may apply statistical multiplexing between virtual networks if the associated SLAs permit them to do so. For instance while providing a minimum throughput for a virtual link, additional bandwidth may be shared between several other VNets and may be utilized if available.

(c) Hard QoS Guarantees: For virtual networks that are very sensitive to QoS parameters or for which strict isolation is of uttermost importance, the substrate must enforce those hard QoS guarantees. That usually implies the use of admission control procedures on the respective resources.

2) QoS at the virtual network level:

The second level of QoS supporting mechanisms is located inside of virtual networks and inherently depends on the QoS guarantees given to the VNet by the substrate. Based on those guarantees, the Operators of virtual networks can then apply QoS models and mechanisms they prefer inside the virtual network. This includes especially all QoS mechanisms spanning multiple virtual links, including those providing end-to-end QoS guarantees. We note that QoS guarantees at this level are not possible if the substrate level QoS guarantee for virtual links is only best-effort.

This two-layer QoS model leads to the following implications with respect to a globally deployed virtualization framework:

(a) In order to provide virtual networks with network resources spanning multiple infrastructure provider domains with QoS guarantees, an inter-domain QoS solution needs to be standardized and deployed. This at least implies a standardized notion of interoperable QoS descriptions as for example provided by the QoS NSLP QSPEC template. Whereas the QoS models and mechanisms by which the QoS constraints are enforced inside a domain may be left to the responsible parties, at the inter-domain boundaries, a common language for QoS specifications and interoperable mechanisms are required.

(b) Operators of virtual networks may deploy their preferred QoS models inside their seemingly homogenous VNets at potentially global scale without requiring global agreement. This, however, comes at the cost of an agreed-upon QoS solution at infrastructure provider level, which has to be highly flexible, extensible, and scalable, e.g., by considering aggregates of virtual links wherever possible, in order to support future needs and to deal with a vast amount of virtual networks running in parallel.

Summing up the above, many of the QoS-related problems emerging with the advent of network virtualization can be solved by applying existing QoS approaches, such as basic mechanisms of Differentiated Services. Some of them are already deployed whereas some others have been rather investigated from a scientific point of view only and are not widely deployed in practice or used within a limited scope only. For instance, from an inter-domain perspective, standardised interfaces, deployable business models, and inter-provider agreements are necessary in order to get network virtualization deployed in a global manner. With this prerequisite met, however, global deployment of QoS provisioned virtual networks could foster the development of service-tailored networks for innovative future applications.

3. GENERIC PATH

Aiming to overcome limitations present in the current Internet through a set of radical architectural approaches, we design a new end-to-end communication abstraction, the Generic Path (GP) [3]. One main aim of the GP clean slate concept is to support various types of communications in highly mobile and dynamic network scenarios between two or more end systems. It aspires to adapt transport and QoS procedures to the capabilities of the underlying network when multiple routes as well as advanced techniques such as network coding and different types of diversity are considered for improving efficiency of information transport.

3.1. The Generic Path Architecture

The current Internet architecture is founded on the layered model. At the time, this was one of the powerful software engineering foundations for abstracting functionality and separating engineering concerns. This model, however, seems to be less than adequate today. The main reasons are its rigid opacity and the lack of recursiveness that is often needed to explain and capture the repetition of functionality in different contexts and scales. This means that in the current layered model, functionality, semantics, APIs and algorithms are not re-usable or cooperative across different levels of abstraction; thus leading to an explosion of APIs, repetition of functionality in overlays, and uncontrolled feature interactions due to competitive objectives across layers. Last but not least new functionality is impossible to introduce outside the scope of a layer without violating the architecture.

Moreover, in today's Internet architecture the socket level API is supposed to be the main generic abstraction that hides the complexity of an underlying network sub-system and offers the network service developer a generic access mechanism to different layer functionalities. This, however, is frequently not the case. As a result, today, most network application developers resort to the use of higher level network APIs that are offered by middleware platforms. Those, although they partially solve the problems, are non-standard and system centric as opposed to network centric.

In our work on the Generic Path, we focus on the aspect of resource management by applying shared semantic concepts and formalizing the heterogeneous communication technology with ontologies. At the same time, apart from the foundational work, we back the proposed architectural framework with mechanisms and actually apply them in a number of key application domains such as routing and mobility.

3.2. Semantic Resource Management Based On Ontologies

Generally, an ontology defines a "network" of information associated with logical relations used for structuring and as a means for data/information exchange. In this way, information and mechanisms can be shared between different systems.

The use of ontologies in future networking is particularly suitable to support and facilitate network interoperability. In the current situation, there subsists a huge semantic gap between the service and the transport layer. So the general objective should be to improve the vertical interplay by introducing semantic methods (ontologies) into the network. Ontologies additionally open up opportunities to amend the horizontal resource handling of heterogeneous network technologies. Semantic methods are already established in applications (Web services) and in the QoS area. Web services are supported by ontologies describing the concepts of an application domain. QoS and traffic characteristics are represented by QoS ontologies. However, network resource ontologies still need to be developed to map and match the capabilities of different technologies in order to support the requesting applications.

3.2.1 Ontologies for Traffic Profiles and QoS

The consideration of QoS requirements in the GP context leads to an approach similar to the semantic Web services. The aim is to support automated QoS-aware network resource selection and composition, which addresses the service QoS requirements and objectives.

There is a substantial need to represent GP QoS features with a QoS ontology in order to adopt it for the GP traffic profile construction. More precisely, the QoS profile parameters (bandwidth, delay, error tolerance) are deduced from the QoS ontology descriptions. Figure 2 shows an example for a QoS metric ontology, where a metric is separated into static and dynamic metrics [6].

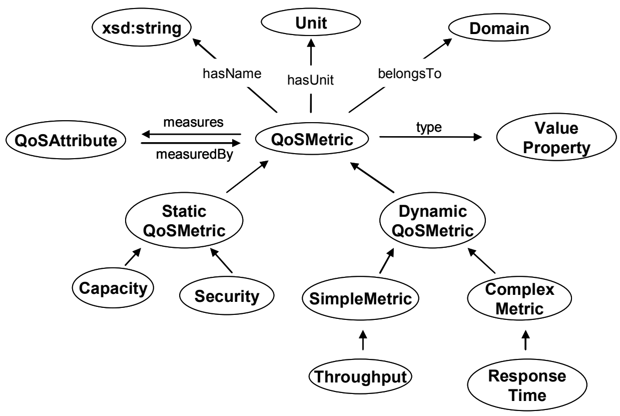

Figure 2. QoS Metric Ontology Example

3.2.2 Network Resource Ontology

In the first step, a generic representation of the network resource parameters has been developed. It provides a standard model defining the characteristics of associations between attributes and the way they are measured based on the following main classes:

- Network Resource Parameter representing a property of the resource within a specific compartment;
- Metric defining how each parameter is assigned with a Value;
- ConversionFormula class enabling the transformation from one metric to another;
- Impact representing the way the parameter value contributes to the perceived quality;
- Type specifying the category of the ontology vocabulary, where the parameter belongs to;
- Aggregated describing the property of aggregation.
- Relationship characterizing the way a Parameter is correlated with others.

Figure 3 shows an example of a network resource ontology model for arbitrary network resource parameters.

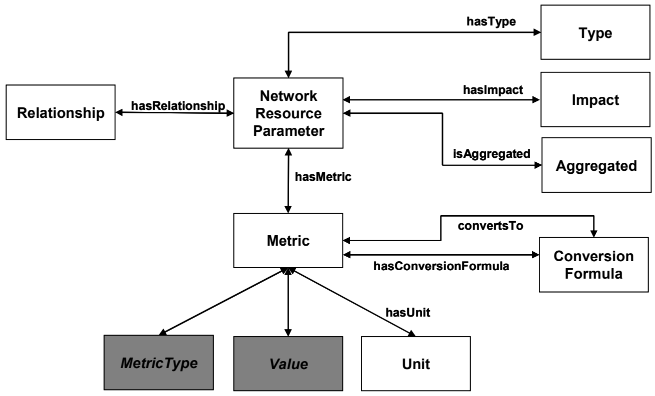

Figure 3. Network Resource Parameter Ontology

If the discovered network resources are accompanied with descriptions of their non-functional properties, automated network resource selection and aggregation is possible. Figure 4 sketches an exemplary network resource ontology representing the relationship between diverse resource types on different levels.

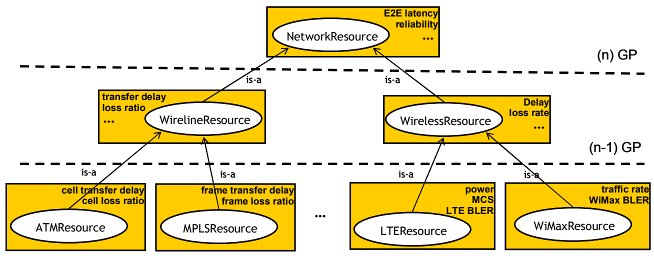

Figure 4. Exemplary Network Resource Ontology

It also shows examples for resource type associated network resource parameters, instantiated as subclasses of the network resource parameter class. The WirelineResource and the WirelessResource classes are examples for an abstract resource type providing an encapsulation for resource aggregation. Generally, the ResourceObject represents physical and abstract resources at a specific GP level.

3.2.3 Semantic Resource Management

The semantic resource management supports fair resource strategies by combining best effort and strict allocation policies to a hybrid approach.

Arriving packet flows are classified according to the QoS profile, resulting in a traffic resource request. This request arrives at the top-level (GP) ResourceObject (Figure 5), representing the current status of the abstracted E2E Resource. Based on the current route path the resource assignment starts and checks if the status of the ResourceObject is UpToDate. In this case, the process verifies whether the available resources fulfill the requirements. If the required resources are available, resource assignment can be performed. Otherwise, a re-routing has to be triggered. In case re-routing does not succeed, the request has to be delayed or rejected. In this way, the GP-using application gets immediate feedback to a dedicated resource request by accessing the top-level (GP) ResourceObject.

At the bottom-most GP level ResourceObjects represent real physical resources, which are located and distributed directly at the physical resource level. There technology specific resource functionalities are performed and information about the current link state, e.g. link degradation etc, are available. As the physical resource is a highly shared object, it is very beneficial to provide a pro-active behavior at each layer by iteratively advertising its resource status to all ResourceObjects in the compartment (abstraction) layers above. In Figure 5 the resource status is periodically advertised starting from the bottom-most ResourceObject level to vertical and horizontal adjacencies. The advertising process can be triggered from a resource request in the case when the ResourceObject is not UpToDate.

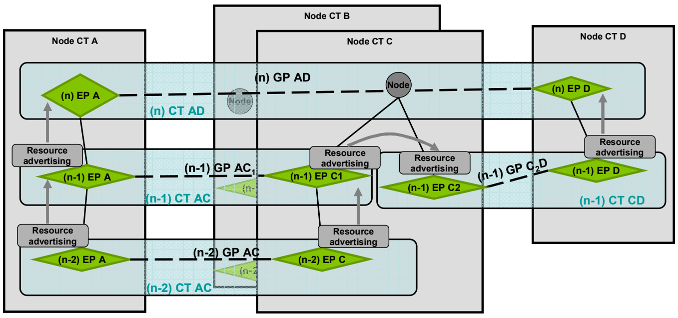

Figure 5. Resource status advertising within the GP architecture

In vertical direction, the advertising process leads to an aggregation (concatenation) of different and possibly heterogeneous ResourceObjects (level (n-1)) resulting in a more abstract representation of the ResourceObject (level (n)). Aggregation and abstraction are supported from the network resource ontology. In particular specific resource parameters are converted to a new metric with the help of the conversion formula depicted in Figure 3.

3.3. Defining QoS for challenged networks

One of the innovations incorporated in the Generic Path architecture is support for intermittent connectivity in Delay-Tolerant Networks. The feasibility of Delay-Tolerant Networking (DTN) and the efficiency of its implementation are closely related to features provided by the GP architecture. In the case of DTN message switching, the GP represents and manages the delivery of the message (or, more generally, of the Information Object) from a source application to a destination application.

The main question that a DTN routing algorithm has to answer is: "How many copies of a message should be distributed into the network?". The answer may range from "one" to "as many as the present nodes are". We argue that this question has to be answered taking into account the desired service outcome, which is not the same in all cases.

In conventional Internet communications end-to-end connectivity exists at all times and therefore, QoS techniques are investigated accordingly. For example, different service priorities, over-provisioning and queuing management, just to name a few, have been of great interest the past few years. In the Internet, however, QoS techniques can be triggered reactively, i.e., reside above the network layer of the TCP/IP protocol stack, since this is indeed possible given the small end-to-end delays. The rules change, however, when discussion comes to challenged, partitioned, disconnected, delay-tolerant networks. Thus, we attempt to re-define the term "QoS" for such networks.

In particular, there exist two ultimate goals a DTN forwarding algorithm attempts to achieve: (i) (high) delivery ratio, (ii) (low) delivery delay. We call these two goals the service targets of a DTN system; we contend that these service targets form the main QoS guarantees that a routing/forwarding algorithm should be able to provide. Furthermore, given that DTNs consist mainly of mobile, battery-powered devices, they will be constrained with regard to: (i) energy consumption, (ii) storage space. We refer to these constraints as system constraints. Although delivery delay may sound contradicting as a QoS service target for delay-tolerant networks, we note that delay-tolerant may be an application that can tolerate one minute of delay, but delay-tolerant may also be an application that has to tolerate one week of delay. In that sense, some sort of service differentiation seems to be essential.

Since the DTN "killer app" is yet to be found [8], we pick some well-known, but still DTN-applicable applications in order to illustrate the difference. Email requires 100% delivery ratio, but is not strict in terms of delivery delay. In contrast, "web-on-the-move", or non-critical telemetry data becomes useless if not delivered within (relatively) strict deadlines. In that sense, QoS targets appear to be diverse among DTN applications.

Applying similar forwarding policies to the above applications may drain the system's energy in case of e-mail flooding, or saturate the system's storage in case of reliable telemetry delivery, for instance. In other words, QoS guarantees may be difficult (if possible at all). We argue that holistic, "one-fits-all" approaches to DTN routing, which appear to be the norm until now, will lead to QoS deadlocks, similar to the ones that the research community faces presently in the conventional Internet (e.g., mice vs. elephants). In contrast, proactive, service-target driven QoS designs have the potential to provide, inherently, supply according to demand, instead of reactive patches that would regulate demand according to supply. A first step on that direction has been made in [9].

4. IN-NETWORK MANAGEMENT

The network of the future is believed to be a network with converged services. The same network will provide access to data, voice or high quality video content to the end users. Because not all the data flows require the same traffic parameters, we have to classify them according to the transported information type and treat them differently. All these operations are made in order to maintain a specific quality of the service, provided by the operator to the end user and specified in the SLA between those two entities. Because a usual traffic flow crosses different communication domains, each one with its own rules, it is still a challenge to guarantee end-to-end QoS services on a communication channel, because each domain can implement different QoS mechanisms that are not always compatible, or worse, are not offering QoS at all. The reason that many network administrators choose not to enable these functions in their network is that they are difficult to configure properly, requiring a thorough understanding of the mechanisms behind.

We consider that QoS is an important management functionality that should be supported by future networks. The architecture of each network node should contain several quality of service related management capabilities. They are included into an SE (Self-Management Entities) and have two types of interfaces. The organizational interface (ORG) is used by a manager or another entity to send high level commands to a specific INM (In-Network Management) entity. The collaboration interface (COLL) is dedicated to facilitate the communication between two management entities residing either in the same or in different nodes

The INM CLQ (Cross-Layer QoS) accesses the hardware directly, through the collaboration interface and it has actually two approaches:

- Bottom-up approach: will enable collecting traffic parameters like: ATR (Available Transfer Rate), OWD (One-Way Delay), BER (Bit-Error Rate),, and other information that is able to characterize a specific physical link. This is an objective way of evaluating a communication channel. The results could be obtained directly from the hardware driver where the technology will permit, or using different dedicated tools that will perform passive or active measurements between nodes.

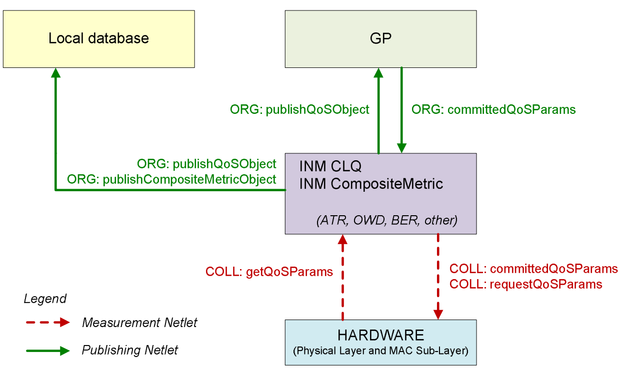

Figure 6. Interaction between INM Cross-Layer QoS, INM Composite Metric, hardware and other managed entities

- Top-down approach: will impose a specific transfer rate to the hardware through INM platform and hardware driver. The requested traffic parameters will be received from different applications (e.g. GP) through the organizational interface and sent directly to the hardware, using the collaboration interface.

An immediate example of INM CLQ's beneficiary would be the real-time composite metric (CM) calculation. The preliminary formula used for an overall perspective of the links with the neighbors (for hop-by-hop data transport) was:

$$CM = \frac{k_0}{ATR[bps]} + \frac{OWD[s]}{k_1} + k_2 \times BER \qquad (1)$$

where $k_0 = 10^9$ *[bps]*, $k_1 = 10^{-5}$ *[s] and* $k_2 = 10^{12}$. This CM could help the management as criteria for triggering network-coding based GP activation, QoS-aware routing, etc. The formula should be interpreted in a similar way Cisco's EIGRP composite metric is used in Network Layer. This means that the maximum ATR envisaged was 1 Gbps, the minimum OWD was 10 microseconds and the BER was not involved in fixed networks-based testbed. Obviously, additional work is needed to demonstrate that this formula seizes the dynamicity of the physical links. However, as a first step we may consider the composite metric provided as useful.

This paper proposes and demonstrates that a combination between INM and GP is feasible. Based on our evaluation of existing solutions in [10], we designed Cross-Layer QoS as a particular example of in-network management, according to the new paradigm. To resume, INM supposes inherent management (or at least integrated) into the monitored network, and not external as in the legacy solutions. The generic path is an abstraction facilitating the development of applications that use the data transport and enhancing the communication's reliability and quality.

By combining Cross-Layer QoS with a simplified generic path, i.e. a multi-point-to-multi-point communication based on network-coding (NC), congestion control is one of its immediate applications. Note that our approach addresses the case of preserving the performances of the running services, despite the congestion which cannot be eliminated. Thus we offer an enhanced distributed routing on a real-time implementation on the existing hardware (no dedicated platforms are needed). This is valid for all types of networks and it involves dedicated software that could be automatically activated/ deactivated by the Cross-Layer QoS. An example of employing the in-network management capabilities for building a GP based on network coding (NC) is discussed in [11], [12]. Network coding was intended for improvement of the multicast transfer rates. It was mathematically demonstrated that NC could improve significantly the multicast transfer rate when congestion is present in the network. Even if the principle of the NC is a relatively simple one, the implementation of such a technique is difficult due to the supplementary operations required: selection of the encoding nodes and of the interconnecting links (shortly the coding network topology), appropriate choosing of the links transfer rates and of the local coding coefficients in the encoding nodes, network wide synchronization of the flows transmitted between neighbor encoding nodes.

The QoS management element collects information about available resources in each strategic node, i.e. in a node that includes GP, CLQ and major management capabilities (neighbor discovery, registry, resource control, event handling, security etc.). The testbed, presented in Figure 7, includes six routers with cross-layer & network coding capabilities (R1 - R6), each one running in a Linux-based virtual machine. The data flow generators (S1, S2) and the destinations (D1, D2) are PCs performing cross-layering only. Specialized software is running on each node to monitor the substrate resources, i.e. the transfer rates and the one-way delays between the neighboring nodes, in order to assist congestion control mechanisms to get a global perspective and to have statistics on link status.

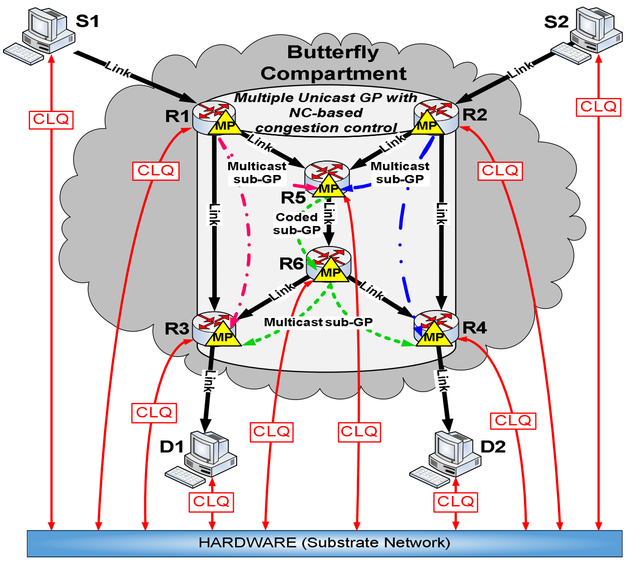

Figure 7. Testbed for Cross-Layer QoS and NC-based GP

Briefly, the experiments included three cases: a) Case 1 (no congestion, no NC-based GP): due to enough available transfer rates on link R5-R6 in Figure 7, the quality of experience at the destinations was very good. b) Case 2 (congestion on link R5-R6, no NC-based GP): both receiving nodes R3 and R4 experienced a bad quality of the movies because of the packets lost in the congested link. c) Case 3 (congestion on link R5-R6, with NC-based GP awareness): the link between R5 and R6 remained congested as in Case 2, but a NC-based GP was instantiated whenever CLQ triggered the situation. Note that a pre-congestion was experimentally detected using the INM algorithm. Thus the mechanism was activated in advance, before severe congestion might occur. The measurement results presented in Figure 8, according to [12], show that the number of packets lost is very low (0.75%, compared to about 18% in the previous case).

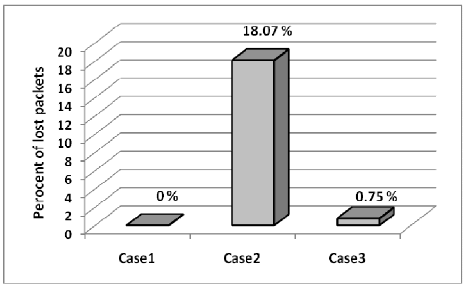

Figure 8. Summarized results proving the efficiency of INM CLQ combined with GP in case of congestion [12]

5. CONCLUSION

Considerable research effort will be necessary to address the challenges raised by the design of the Networks of the Future.

In this paper, three concepts developed in the framework of the FP7 4WARD project have been analyzed from the point of view of the potential impact on QoS – Network Virtualization, Generic Path Semantic Resource Management and In-Network Management.

While these novel networking concepts have not been specifically targeted at handling QoS issues, they undoubtedly enable new QoS approaches and solutions for the networks of the future, especially in view of requirements such as dynamicity, flexibility, adaptability and scalability.

Network virtualization decouples networks from infrastructure and allows infrastructure resources to be shared among multiple isolated networks. By enabling the capability to build service-tailored networks and overcoming the limitations of the traditional "one-size-fits-all" approaches, network virtualization inherently brings potential advantages in terms of QoS. However, a two-layer QoS model will be required, which raises new challenges, particularly in multi-domain scenarios.

The Generic Path is a new end-to-end communication abstraction that aims at overcoming the inadequacies of the traditional layered network model. Resource management is accomplished by applying shared semantic concepts and formalizing the heterogeneous communication technology with ontologies. The QoS features of the Generic Path can be represented by an ontology from which QoS profile parameters, such as bandwidth, delay and error tolerance can be derived. Semantic resource management enables fair resource strategies by combining best effort and strict allocation policies.

Finally, in-network management enables the incorporation of QoS management capabilities in the network elements, thus facilitating QoS configuration, even without a thorough understanding of the usually complex mechanisms behind it. Cross-Layer QoS has been presented as a particular application of in-network management in a scenario combining cross-Layer QoS with a multi-point-to-multi-point communication based on network-coding. In INM network management is inherent, or integrated, into the network, and not external as in legacy solutions.

Integration of these new concepts with traditional QoS mechanisms will require further study. The need for interoperability will surely make standardization play a fundamental role in this scenario.

ACKNOWLEDGMENT

The authors would like to thank colleagues in the 4WARD Project, especially in Work Packages 3, 4 and 5 for fruitful discussions. In particular, the authors are grateful to Ioannis Psaras for his contribution to the DTN section.

This work has been carried out in the framework of the IST 7th Framework Programme Integrated Project 4WARD, which is partially funded by the Commission of the European Union.

REFERENCES

[1] The FP7 4WARD Project, http://www.4ward-project.eu/

[2] S. Baucke et al, “D-3.1.1 Virtualisation Approach: Concept”, http://www.4ward-project.eu/index.php?s=file_download&id=68, September 2009

[3] S. Randriamasy et al, “D-5.2 Mechanisms for Generic Paths”, http://www.4ward-project.eu/index.php?s=file_download&id=78, May 2010

[4] G. Nunzi et al, “D-4.2 In-Network Management Concept”, http://www.4ward-project.eu/index.php?s=file_download&id=37, March 2009

[5] “Approaching the Zettabyte Era”, Cisco White Paper, http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-481374.pdf, 2008

[6] L. Lin, S. Kai, S. Sen, “Ontology-Based QoS-Aware Support for Semantic Web Services”, Technical Report at Beijing University of Posts and Telecommunications, 2008

[7] M. Johnsson et al, “Introduction to a 4WARD Architectural Approach”, 4WARD Deliverable 0.4, February 2009

[8] J. Crowcroft, E. Yoneki, P. Hui, T. Henderson, “Promoting Tolerance for Delay Tolerant Network Research”, SIGCOMM Comput. Commun. Rev., Vol. 38, No. 5, 2008, pp. 63-68

[9] I. Psaras, N. Wang, R. Tafazolli, “Six years since first DTN papers. Is there a clear target?”, ExtremeCom, 2009

[10] A. B. Rus, V. Dobrota, “Overview of the Cross-Layer Paradigm Evolving towards Future Internet”, ACTA TECHNICA NAPOCENSIS, Electronics and Telecommunications, ISSN 1221-6542, Vol.50, No.2, 2009, pp.9-14

[11] Z. Polgar, Z. Kiss, A. B. Rus, G. Boanea, M. Barabas, V. Dobrota, “Preliminary Implementation of Point-to-Multi-Point Multicast Transmission Based on Cross-Layer QoS and Network Coding”, 17th International Conference on Software, Telecommunications & Computer Networks IEEE SOFTCOM 2009, Split-Hvar-Korkula, Croatia, September 24–26, 2009, pp.131-135

[12] A. B. Rus, M. Barabas, G. Boanea, Z. Kiss, Z. Polgar, V. Dobrota, “Cross-Layer QoS and Its Application in Congestion Control”, 17th IEEE Workshop on Local and Metropolitan Area Networks LANMAN 2010, Long Branch, NJ, USA, May 5-7, 2010, http://www.ieee-lanman.org/

[13] G. Schaffrath, C. Werle, P. Papadimitriou, A. Feldmann, R. Bless, A. Greenhalgh, M. Kind, O. Maennel, L. Mathy, “Network Virtualization Architecture: Proposal and Initial Prototype”, p. 63-71, ACM, Barcelona, Spain, August 2009.