# Supporting an architecture for cross-layer optimization

Gianmarco PANZA[1], Luca SIOLI[2]
[1]*CEFRIEL – Politecnico di Milano, Via Fucini 2 Milano, 20133. Italy*
*Tel: +39 02 23954326, Fax: + 39 02 23954526, Email:gianmarco.panza@cefriel.com*
[2]*CEFRIEL – Politecnico di Milano, Via Fucini 2 Milano, 20133. Italy*
*Tel: +39 02 23954326, Fax: + 39 02 23954292, Email: luca.sioli@cefriel.com*

**Abstract:** Due to the lack of layer interaction between networked applications and the underlying network during service provisioning, many end user applications and services cannot efficiently utilize the network capabilities, nor can achieve the desired quality of service objectives. As applications grow in areas such as Video on Demand, Video Gaming, or network storage, the management needs to be smarter to pick multi-layer topologies that are efficient across the entire network.

In this paper, the general problem of cross-layer optimization is deeply analysed, and the need for it pointed out together with the most relevant and challenging issues.

In brief, it aims at the overall optimization of application layer and network resources including transmission ones. To achieve such a goal, an effective interaction and exchange of information between the lower and upper layers are required.

Design options for effectively addressing the data collection, synchronization and provisioning, as well as the scalability and backward-compatibility issues are discussed in detail, in order to specify a cross-layer communication architecture that enables full optimization across multiple layers, devices, domains and technologies in the Future Internet.

**Keywords:** Cross-layer communication, Cross-layer design, Cross-layer optimization, Future Internet, Multi-device, Multi-domain, Multi-layer, Multi-technology, scalability.

## 1. Introduction

Nowadays, more and more applications exploit the Internet and the underlying network connection for providing an enhanced user interaction. Upper layer services can rely on a variety of application layer resources such as data storage, computation power, specialized server capabilities, large data sets, and can make dramatic demands on network resources, such as bandwidth, and Quality of Service (QoS) guarantees.

However there is a lack of layer interaction between networked applications and the underlying telecommunication infrastructure during service provisioning and consequently many upper layer services make poor use of network resources or do not achieve their overall QoS objectives, including being wrongly denied of requested resources actually available.

Currently, if an application client can obtain a desired large data set (file, video, database, etc.) from a server selected between many different options, the application service will take into account the current status and load on the possible servers but only minimal network considerations, such as topological proximity, connectivity, ping latency

(i.e. the issued Round Trip Time), rather than actual link bandwidth utilization or other relevant QoS parameters (e.g. reliability, delay and jitter).

The lack of communication between the application and lower layers (from physical to network strata) across the Internet does not allow cross-layer optimization to be applied for the followings:

- coordinated application and network requests for available resources of all layers (including transmission ones)
- efficient recovery from failures based on policy related to all levels of the protocol stack
- monitoring and provisioning processes relying on synchronized information about application and lower layer resource availability
- consistent operations on information coming from different layers, devices and domains independently from employed technology (e.g. optical or electrical, wired or wireless)
- detailed knowledge and control on expected and actually provided end-to-end QoS

Adding more details, application Resources are those critical for achieving the upper layer functionality. Examples include caches, mirrors, application specific servers, contents, large data sets, and computing power. While, network Resources are those of layers 3 or below, such as bandwidth, links, connection processing (i.e. creation, deletion and maintenance) capabilities and network databases.

Without communication and optimized management across multiple layers, poor service provisioning or resource exploitation are likely to happen at all levels.

This work[1] provides evidence of the need for cross-layer optimization (and communication) in order to achieve an effective support of value-added services and applications in the Future Internet, which can be seen as a confederation of autonomous systems [1], having each its own architecture, topology, technology and facility, over an IP infrastructure. In practice, the overall optimization of application layer and network resources, including transmission ones, requires the interaction and exchange of up-to-date and consistent information between the lower and upper strata across multiple devices, domains and technologies.

The main issues to address for the purpose, such as data indexing, collection, synchronization and distribution are also outlined and deeply discussed. Furthermore, aspects related to efficient communications, scalability, interoperability and backward-compatibility are yet considered in order to specify an effective architectural solution for cross-layer communication enabling a full optimization across the whole Internet.

The remainder of this paper is organized as follows. First, the limitations of current Internet for an effective support of networked application are pointed out, as a basis for the problem statement and objectives definition. Second, main challenges and necessary capabilities are discussed together with the inadequacy of already deployed control planes and proposals in the field. Then, design options for addressing the identified issues are considered in detail, as basis for the specification of an architecture supporting cross-layer communication and hence optimization over the Future Internet.

## 2. Problem identification and final objectives

In the following, the issues related to popular applications and their effective support are presented with the purpose of underlining the need for cross-layer communication and clarifying the ultimate optimization objectives.

[1]The research leading to these results has received funding from the European Unions Seventh Framework Program ([FP7/2007-2013]) under grant agreement n° 288502.

*2.1 –File distribution systems*

Internet content and file distribution systems have been set up as overlays on existing network infrastructures. Commonly encountered optimization problems with network implications include cache and Mirror placement, efficient transfer of content to servers and client to server assignment.

The cache placement problem concerns what content to allocate to which cache servers based on their proximity to clients and their load [2]. Mirrors differ from caches in that a client is only directed to a mirror if it has the desired content [3]. The mirror and cash servers placement problem concerns where to place them given a fixed number of possible sites [3][4].

The employed algorithm necessarily works with knowledge of application topology and some type of network topological information, such as relative link cost models.

Synchronization of application and network data (in general, of every layer) is key to optimization. Actually, exact network models are not always necessary to achieve significant performance improvements [5].

The efficient transport of the original content to the replication servers becomes more and more important for optimization with the increase of the amount of material to transfer. When dealing with a large set of replication servers and quantity of data, the delivery benefits from point-to-multipoint concurrent communication path selected on the basis of the current network load conditions. Therefore, the cross-layer optimizing process must have visibility into the underlying network resources and operating on up-to-date and consistent (hence, synchronized between layers) information.

In the assignment or selection of a content server for a given client, both server load (application layer information) and transfer latency between the client and server [5] should be taken into account. This highlights the need for synchronized multi-layer monitoring and configuration.

*2.2 Content streaming distribution systems*

Basically, streaming services can be either live or on-demand. However, many variants in between these two extremes are created when pause or replay functionality is included in a live streaming service. Such services are quite different from file downloading ones. first, the beginning of content consumption does not require an entire file to be transferred. Second, minimum bandwidth and possibly other QoS requirements are to be guaranteed wen delivering content between the server and client.

In live streaming, the client is willing to receive the content at its current play out point rather than at some pre-existing start point. A key network issue is whether multi-casting takes place at the application or network level. Both the options are being adopted. For example, in carrier operated IPTV telecommunication infrastructures IP multi-casting is spreading [6]. While, in the case of an independent live video distribution service, overlay networks of servers provide the multicasting facility.

In the end, optimization problems for a live streaming service are: either server selection (application based multi-cast) or leaf attachment (network based multi-cast) [7], and either server placement (application based multi-cast) or tree construction (network based multi-cast).

On-demand services entail additional technical challenges. From one hand, long start up delays should be avoided to retain customers, but batching together requests to save on server costs is desirable, on the other. Therefore, further optimization decisions and problems typically arise in the on-demand applications in addition to those seen in live streaming, such as client stream sharing technique and, batch or Multicast Server selection.

Actually, the on-demand streaming service concerns are similar to those in file distribution: data allocation (when and where to pre-stock video files), on-demand server placement (where to put and how much capacity) and efficient (cost effective and timely) transfer of content to servers. Therefore, the need for cross-layer synchronized data is yet critical.

### *2.3 Conferencing and gaming*

Conferencing and gaming imply further complexity with respect to the cases above, in the related application connectivity and the need for cross-layer synchronization of monitoring, configuration, Operation Administration and Management (OAM). First, the issues associated with streaming services are also applicable to conferencing and gaming. Second, bi-directional and asymmetric bandwidth connections between the server and the user host are concerned. Third, gaming requires multipoint-to-multipoint communications with hard QoS constraints on latency.

This increased complexity over the point-to-multipoint case of streaming content distribution brings additional problems. Firstly, multipoint-to-multipoint data path formation and re-formation can be very inefficient without considering the underlying network resources. Secondly, addressing QoS constraints on latency and bandwidth guarantees for multipoint-to-multipoint connectivity requires coordination across the layers in terms of both path selection and reservation.

Finally, massively multi-player online games (MMOGs) have the multipoint nature and QoS needs of conferencing but with additional concerns on scalability. For this [8] and due to constraints imposed by player preferences [9], an optimized server selection can be more complicated than in streaming content distribution.

### *2.4 Grid computing*

Grid computing supports extremely large transfers of files and data-streams (live or on-demand). The volume of the traffic makes it critical to synchronize changes at the application and network. In addition grid computing may have delivery requirement similar to those of streaming content distribution systems. Therefore, optimization objectives related to grid computing include effective instantiation of connectivity with high data rates and/or data set size, as well as ability to efficiently control very high speed networks.

## 3. Analysis and state-of-the-art

### *3.1 Challenges and needed capabilities*

In the previous section, problems that occur when resource allocations for both application and network are not synchronized in their action have been outlined together with the associated optimization objectives.

Hereafter, a more detailed discussion about the challenges and needed capabilities for achieving cross-layer optimization is provided with the aim of understanding the key issues to address in the design of an architectural solution supporting efficient communication and management across multiple layers, devices, domains and technologies.

*Synchronized reception of multiple real-time topologies and Traffic Engineering related databases*

As explained in the previous section, the processes of server selection and content placement can have dramatically better outcomes if current network and application topologies are known at the same time. It is critical to have quite detailed data about where

the application clients and servers are located and how they are connected at network layers (from physical to IP ones).

The ability to capture up-to-date and consistent information allows planning during rapidly changing conditions or short traffic burst. For example, location selection for servers and clients requires that the performance estimates about the network and application layers align, in particular when stringent QoS parameters are to be guaranteed.

As 90% of traffic is represented by short flows [10][11], out-of- date information or inconsistent between different strata, will provide an inaccurate representation of the network traffic and status in general. This can cause the selection of non-optimal paths across multi-layer network topologies. The key point is that the data needs to be synchronized at all levels.

*Cross-layer cooperative load and traffic monitoring*

Load and traffic monitoring can be facilitated having available information from Cross layer Optimization and management entities in all layers. Indeed, it is critical to have up-to-date and synchronized) data about QoS and load at every level.

Furthermore, as conditions change or problems occur, it may be important to adjust the granularity of these measurements. For example, bandwidth used and allocated/reserved, network delay statistics, existing client-server relationships and data regarding the allocation of clients to servers should be made available with different levels of detail, as needed.

*Cross-layer synchronization of configuration changes*

Re-optimization over the whole protocol stack end-to-end requires synchronized configuration at all levels. Otherwise, flows at a given layer may fall outside the planned network traffic patterns at other layers.

*Cross-layer provisioning*

The cross-layer connectivity entails the ability to provision additional communication paths and resources at some layers. For example, in MPLS-TE [12] and GMPLS [13] networks, the responsible IP entity is required to initiate connection setup across multiple-layers on behalf of the application entities (such as clients and servers).

### *3.2 Inadequacy and limitations of current control-planes and proposals*

In the last few years, there has been a lot of research in the field of cross-layer communication and optimization over wireless networks. Many solutions have come out with the aim of adapting Physical and MAC layers together with the Application one (e.g. joint source-channel (de-)coding). Various approaches can be followed:

- application layer-based [14], where the application considers the lower layers as a black box and modifies the transmission according to a small set of feedbacks it can receive from them;
- Layer-Centric, where the application directly [15] or through the mediation of an intermediate layer [16] can access the internal protocol parameters of the lower layers and uses them for optimization purposes;
- Centralized [17], where the layers are connected to a middleware or are monitored by a single optimizer that estimates resource availability and coordinates the resource allocation, adapting each protocol’s algorithms and parameters;

- Foresight oriented [18], where each layer decides autonomously for the most appropriate operation mode, using its own data and information gathered from messages exchanged with other layers.

All the above strategies do not take into account the IP and upper levels, because the main issues related to the transmission on wireless links concern the MAC and PHY layers. However, in a general scenario there are a lot of feedbacks and information pieces that the network and upper layers can provide in order to improve the service provisioning, as clearly shown in the previous section. Actually, some attempts have already been made to address this point In the research world.

IETF ALTO WG [19] has been focusing on overlay optimization among peers by utilizing information about topological proximity and appropriate geographical locations of the underlying networks, including related resource usage and availability. With this method, an application may optimize selecting peer by location. But, the current scope of Alto work does not cover multi-layer data synchronization and communication aspects for a full optimization, and it considers slowly time-variant information only.

Yet, cross-layer strategies addressing specific application type or network scenario are being developed into some FP7 EU projects: Alicante [20], Envision [21] and Adamantium [22]. The former provides Content-awareness to the Network Environment, Network- and User Context-awareness to the Service Environment, and adapted services and content to the End- User by relying on advanced middleware and network components that indeed, do not have a full multi-layer view. The latter ones follow a cross-layer approach where the problem of supporting multimedia services is solved cooperatively by service providers, ISPs, users and the applications themselves by delivering both content-aware networks and network-aware applications. However, ENVISION considers only peer-to-peer applications and the being designed complex set of interfaces can make it difficult to be deployed in practice over the multiple devices, domain and technology environment of the Future Internet. While, Adamantium is based on a user/customer-centric approach and not on an engineering one. It proposes extensions to be applied to IMS-compatible architectures [23] specifically. Furthermore, the synchronization aspects related to the data exchanged across multiple layers are not explicitly tackled by either of them.

Regarding existing network management solutions, significant limitations are present and critical extensions should be conceived and applied. SNMPv3 [24][25] provides the idea of a SNMP context, which is defined as a collection of management information accessible by an SNMP entity. An SNMP entity potentially has access to many contexts. The missing piece is a Context with a management information view that allows synchronization of actions across multiple layers, devices and domains for read-view, write-view, notify-view and actions. Netconf [26] supports an XML based access to SNMP MIB data. The same concepts found in the SNMP Access models of context for viewing data are implemented; therefore, the same lack of functionality for synchronization across all levels still exists. MPLS [12] is regarded as one of the underlying network transport technologies that could enable cross-layer optimization with application layer; however, current scope of MPLS OAM [27] does not encompass any non-MPLS device for its configuration and provisioning functions. In addition, UNI interface for GMPLS-controlled telecommunication infrastructures [28] is currently defined for network equipment only, not including direct interaction with the application layer services. ITU-T Y.2011 NGN and Y.2011 Resource and Admission Control Functions [29] discuss the service and transport strata separation. ITU-T Y.2012 [30] defines an Application-Network Interface (ANI), which provides a channel for interactions and exchanges between upper and lower layers. However it does not address any issues on the synchronization of exchanged data.

## 4. Architecture Design

As explained in Subsect. 3.1, the target solution has to deal with several challenges: synchronized reception of multiple real-time topology and Traffic Engineering related databases, cross-layer cooperative load and traffic monitoring, cross-layer provisioning and synchronization of configuration changes across all layers. Therefore, the novel architecture should enable the indexing, collection, synchronization and retrieval of distributed, heterogeneous, time-variant data. Such a data could be organized in different views and made available locally or remotely in a scalable manner, for use by interested parties.

Furthermore, integration and backward compatibility with existing systems, databases, interfaces and protocols is critical for addressing straightforward extension and improvements over multiple layers, devices, domains and technologies in the Future Internet.

In this section, design options for such issues are first specified and discussed.

Then, due to inadequacy and limitations of current control-planes and proposals a novel architecture encompassing all the required components and functionality, enabling cross-layer communication, and hence optimization is presented. A key benefit of such architecture is that no constraints are imposed on the existing infrastructure. Therefore, it can be easily and progressively deployed for consistently collecting, synchronizing and providing data on demand, in a scalable manner.

### *4.1 Information type and operational*

The information type clearly depends on the regarded resource. Examples are the delay on the IP interface and throughput at the transport layer of a given network node. The set of relevant information that is made available by a component (producer entity) is constrained by the concerned control-plane and policy put in place by the ISP of the domain where the said component is located. Furthermore, the same information type can be provided with different levels of granularity (e.g. the queuing delay can be per flow, user or aggregate).

A critical point is the consistency and validity of information. An information is valid if it actually refers to the status of the associated resource (not of others) and is up-to-date; while, it is also consistent if it is representative of the status of that resource (e.g. it could not be the case for an instantaneous value of the SNR on a wireless interface), possibly in conjunction with the status of resources in the same layer or of others in the network at the same time.

More specifically, the status of a resource can be identified by both static and dynamic data. The former can be notified once, while the latter needs to be refreshed with a proper frequency, possibly in association with filtering and hysteresis techniques in order not to generate too much overhead and have still reliable values for extremely time-variant information (e.g. if related to wireless interfaces). The data processing can be performed at a collecting server or at the issued component (depending also on its capabilities). The scalability and efficiency of the overall system are directly connected with the granularity and refreshing frequency; therefore, a trade-off should be found between accuracy in the knowledge of the network status and implied overhead. The validity of information is strictly related to its nature, it can become stale for several reasons, such as failure of the issued components or of the cross-layer communication system, and long latency in the distribution-collection mechanism. Gathered Data can be automatically deleted when an associated timeout expires.

Atomic or composite data can be generated periodically in an automatic way (according to policy) or as a result of explicit request (in this case, a delay in the data availability is present). For example, performance for a flow of traffic aggregate along a network path can be measured by an end-point only when requested or with a given frequency, directly (e.g.

by passive or active probing) or by combining relevant information provided by the nodes along the issued path.

The way data and possibly related meta-data, are coded can be different, also in association with the specific information type,. The important matter is that the information producer and consumer align on that.

To be noticed that the architecture being defined in this work is fully agnostic to the information type and coding format.

### *4.2 Indexing*

Information must be uniquely identifiable for proper management and retrieval. Basically, the IP address of the device where the concerned resource has been instantiated is needed.

In addition, an object ID can be used for complete specification (contingently, also for selecting a given database view). An option is to rely on a hierarchical organization of such an ID, also for scalability reasons (see the structured adopted in the SNMP MIB queries [24][25], for example). Alternatively, it can be represented by an N-upla of elementary identifiers (such as, protocol ID, port number, interface ID, service class name and queue number). In detail, different granularity for the same information should be separately and uniquely named. To be noticed that the adopted format is not required to identify the location of the related resource information when such information can be also retrieved from given databases built for the purpose, other than directly from the concerned producer entity. However, data regarding real-time flows can be accessed directly from the producer entity as for re-direction by the said databases (see Subsect. 4.5).

For the indexing issue, standardization is critical in order to achieve a global spreading of cross-layer communication and optimization over the Future Internet.

For interoperability and efficiency reasons, information addressed and made available by a given control and management system (e.g. SNMP, MPLS OAM) could be provisioned to a requesting consumer entity all over the multiple devices and domains where that system is supported by developing the needed data adaptation and communication interfaces. Likewise, content of existing databases (e.g. for Traffic Engineering, associated with the implementation of MIH standard [47], about overlay networks) should also be identified in conformance with the adopted reference format for the indexing, in both the publishing and access processes, as well as properly interfaced in order to enable their use.

### *4.3 Distribution-collection*

The critical issue is the design of a scalable solution that supports the distribution by producer entities of resource information for use by consumer ones; from one hand, for efficient retrieval and on the other, introducing limited overhead. It appears appropriate to figure out in each domain some Collection Points (CPs), possibly in a hierarchical organization for scalability reasons.

Information is to be accessible by means of CPs, locally or remotely. Data about the network status (at all levels) should be stored locally in databases (associated with the respective CPs) located in the domain of that network. While, data related to resources of the other domains can be stored in databases belonging to those domains. Information can be simply notified as available to the said CPs rather than actually delivered to them. This is for latency reduction in the provisioning, as well as for saving bandwidth and memory storage in the issued databases. This is most reasonable for extremely time-variant information that could be provided on demand to the consumer entity directly by the concerned producer entity (e.g. when located at a wireless interface), for data about resources in different domains and in the case of leveraging on existing databases.

Therefore, every CP is associated with both a database and a directory, for actual information provisioning and re-direction, respectively.

Also for backward compatibility issues, existing databases (e.g. for Traffic Engineering, associated with the implementation of MIH standard, about overlay networks, information base of deployed control and management planes) can be integrated in the architecture for cross-layer optimization by properly registering their content (i.e. the related indexing information) to the CPs of the respective domains. In this way, by a re-direction process that available data can be accessed by interested parties. Likewise, CPs of a given domain should register their content to the CPs of the others, as for retrieval of information from anywhere about the whole Internet, specifying the set of indexing information (which includes the IP address) of the resources associated with the network components that they serve in the respective domain. For the purpose, anycasting [31] and multicasting [32] facilities of IPv6 can be exploited for supporting scalability and efficiency. The former for registration to the closest (in terms of network proximity or policies enforced by the concerned ISP) CP of the neighbouring domains, and the latter for distribution to the CPs internally to a domain. Both type of addresses can be statically assigned (under IANA [33] approval). The anycast addresses can be spread across the Internet by means of Interior and Exterior Gateway Protocols (e.g. OSPF and BGP [34], respectively); while, the CPs of a given domain belong to the same multicast group.

### *4.4 Synchronization*

As already extensively highlighted, resource information needs to be synchronized across all layers for optimization purposes.

As for 4G networks and beyond, the telecommunication infrastructure is synchronized.

Standard protocols, such as IEEE 1588 [35], can be used for frequency and time synchronization within the Internet. Therefore, it is likely that the Future Internet will be a synchronized network. For what concerns backward compatibility, former solutions (e.g. NTP [36]) can be deployed as well. Having a global synchronization can help in dealing with the issue. However, the very concept is that information about different resources taken as input for a decision making process needs to be mutually consistent across all layers and up-to-date (other than available).

The target precision in the time binding of a data set depends on the type of regarded information about the status of the network. For example, the shorter the refreshing period for a given information (which is related to its time-variant nature), the higher the required precision in order to get a correct picture of the status of the concerned resource (set). As explained in the next paragraph, the consumer entity can also (periodically) retrieve data directly from the provider entity according to policy and the time-variant nature of the issued information (in this case no delay is introduced by accessing a CP, which collects the needed data and is queried only the first time the given information is requested).

Certainly, values from the same device can be mutually synchronized locally before distribution to CPs. A timestamp of the generation instant should be tied up with the provided data in order to create temporal associations between information related to resources located in different layers, devices and even domains. Of course, if such a timestamp is not delivered with the data or considered unreliable, the reception time can be kept as the reference for synchronization purposes, possibly diminished of an estimate of the transferring delay between the provider and the consumer entities when available (e.g. as from a CP or combining information gathered from CPs of different domains if the provider and consumer entities are actually in different domains). If the information is retrieved from a CP (i.e. the CP operates as provider entity) and no timestamp was generated for it, then the time spent in the associated database together with an estimate of

the transferring delay between the producer entity and the issued CP should also be taken into account.

Depending on the specific decision making process and policy in place (related to the concerned service, application, domain(s), etc.), a retrieved information can be considered out-of-date. In the case, different actions can be performed to compensate for the lack of updated information, including using a default value, calculating a likely value and mostly asking the producer entity directly or the CP for an updated value (the same CP that provided the considered out-of-date information).

When the timeout associated with given information expires, the CP should undertake the necessary steps to obtain an up-to-date value (i.e. prompting the corresponding producer entity).

### *4.5 Retrieval*

The goal is to efficiently obtain the needed information. From one hand, the delay in receiving data after a request for it should be as short as possible and on the other, the regarded communication model(s) should be properly selected.

Concerning the first issue, the closest (in terms of network proximity and policies in place) CP can be reached by using IPv6 anycasting [31]. Of course, the anycast address of the CPs in a given domain is to be advertised all over the served network (for use by all its components). When a CP receive resource information by a producer entity, it can efficiently spread it among all the other CPs in that domain by using a multicast address [32] (the same employed also for spreading information about data accessible by CPs of other domains – see the Subsect.4.3). Indeed, the said CPs belong to a multicast group that can also be statically configured. If the information to retrieve is stored in the database of another domain, then the CP interrogated locally will re-direct the request to it. A re-direction is likely to happen as well, when issuing extremely time-variant data (see again Subsect. 4.3). In such cases, a direct communication between the consumer and the producer (or first provider when accessing a remote database) entities should be established for efficiency reasons.

Of course, if the needed information is not available, a proper notification message is sent back by the closest CP in response to a request for it.

Different communication models can be used for retrieving information. For efficient and scalable cross-layer communication, needed data should be provided on demand or when specific events happen (e.g. availability of a new value from a producer entity) in the case of subscription to a notification service. When using one or the other (actually, both of them can be simultaneously employed as well) depends on the specific nature of the issuing decision making process and concerned resource, but also on policies enforced by the ISP (e.g. about security aspects, backward compatibility constraints, limitations dictated by the supported functionality). For example, quite time-variant information to be used for a period of time in network or application operations could be delivered to the interested party every time it changes (possibly using filtering and hysteresis techniques, as also mentioned in Subsect. 4.1) by leveraging on a subscription-notification service, instead of explicitly and continuously querying a CP database or directly the producer entity for it (which would be certainly inefficient).

### *4.6 Communication protocols*

A large variety of options are available for the purpose. TCP [37] and UDP [38], together with their counterparts with security facilities (i.e. TLS [39] and DTLS [40], respectively) can be used at the transport layer, also for backward compatibility issues. Security could be addressed at application level, including credentials in the resource information request as

in SNMP [24][25]) or as proposed in the IETF ALTO WG [19] for authentication, integrity protection and encryption. While, additional features of UDP-Lite [41], DCCP [42] and SCTP [43] do not appear really necessary for cross-layer communications in general. More timers and different types of connections can be used in parallel (as in SIP [44]), for efficiency reasons when an ack about the success of a provisioning operation is required. ICMP [45] relying directly on IP, can also be exploited.

Finally, hop-by-hop and destination options of IPv6 [46] should be employed when an IP flow is already active between the producer entity and the issued CP, the CP and the consumer entity or, the producer and the consumer entities, when data delivery is required between the same end-points. Under the network layer, IEEE 802.21 MIH with some already proposed extensions for communication with upper layers [47] is an option (though, its scope is still limited to a 802.x network).

Routing protocols with support for Traffic Engineering, such as OSPF_TE [48], IS_IS_TE [49], BGP [50], are not actually a good options, due to their flooding nature that would entail a considerable overhead.

The selection among the mentioned protocols is to be made by considering the resulting trade-off between aspects related to communication reliability, introduced bandwidth overhead, expected provisioning delays, backward compatibility, complexity, security and flexibility. Depending on the case, a choice can reveal better than another. in general, simple solutions are preferable. Therefore, UDP, ICMP or IPv6 options (when applicable and yet addressing security issues) should be more appropriate (possibly, with periodic retransmission for reliability reasons).

There are no constraints on the payload format at application layer. Certainly, the indexing data needs to be included in the message for either delivering or retrieving resource information.

Specific port numbers at transport layers, type and code values for ICMP, as well as Type Length Values (TLVs) in IPv6 options can be assigned by IANA [33] as needed for supporting the being defined cross-layer communication architecture.

### *4.7 Proposed architecture*

In Figure 1, the architecture for cross-layer communication as resulting from above discussions is depicted.

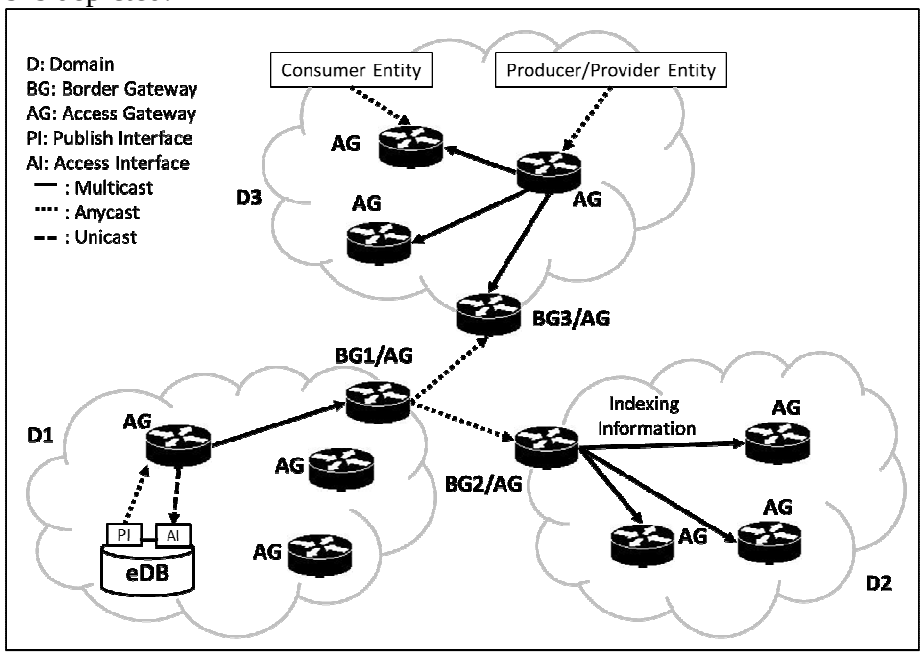


*Figure 1 - Architecture for cross-layer communication all over the Internet*

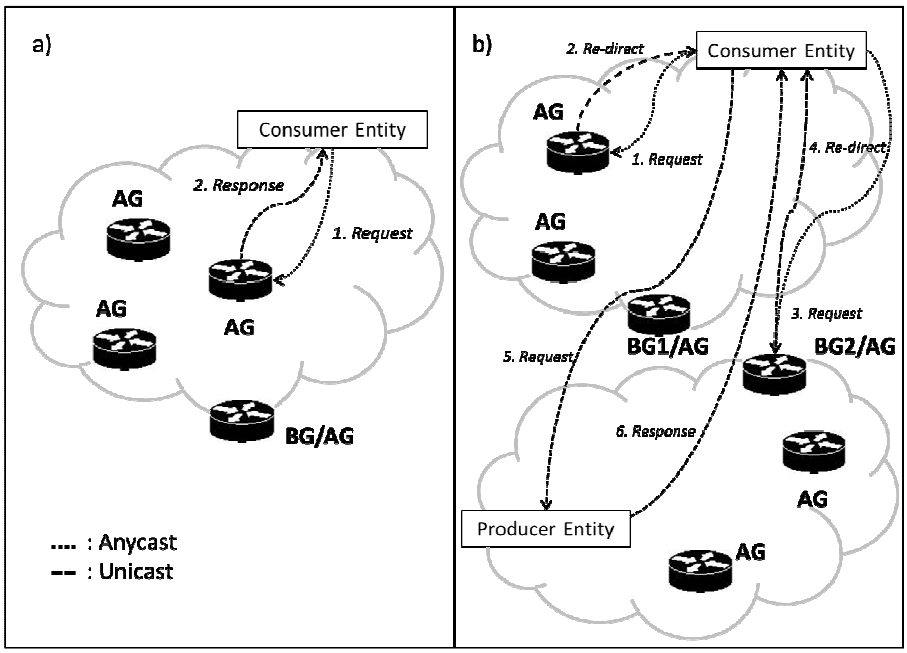


*Figure 2 – examples of information retrieval: (a) directly from the closest AG and (b) from the producer entity in a remote domain after re-directions from the local and remote AGs*

The key components for collecting and provisioning data across multiple layers, devices, domains and technologies are the Access Gateways (AGs) that play the role of the CPs. A pool of AGs is deployed in each domain (e.g. Autonomous System) for the direct delivering of information about resources of the associated domain and the re-direction of requests for information about resources in the others (indeed, re-direction is also possible and reasonable in the local domain when dealing with resources having an extremely changing nature, as pointed out in Subsect. 4.5).

The number and location of AGs In a domain is a trade-off between several aspects, including efficiency, scalability, additional overhead, cost and complexity. The larger the number, the shorter the delay in getting the needed information, but also the higher the introduced overhead, in terms of bandwidth, storage memory and data processing. For small domains even a single AG could be enough.

Among each pool of AGs, a special role is covered by the Border Gateway (BG) that collects indexing information about resources of the other domains. In principle, a hierarchy of BGs can be employed for scalability reasons. In this case BGs of different domains can not communicate directly, but through the BGs that connect them at a higher level in the hierarchy.

As already explained, the pool of AGs (i.e. including the associated BG) in a domain belongs to a multicast group for efficient distribution of an AG database content to the other AGs within that domain. In addition, an anycast address is also assigned to the said pool in order to collect resource information by the AG closest to the producer entity, as well as to provide resource information to the consumer entity by its closest AG. This anycast address is also used by the BGs of the other domains to globally spread the indexing information about the resources in their respective area of competence. The BG of a domain should be advertised to the others as the closest gateway by EGPs for that anycast address. Therefore, each domain has assigned well-known anycast and multicast addresses for its pool of AGs.

An existing database (eDB in Figure 1) can be integrated in the architecture by publishing the indexing information about its content to the closest AG (which will then distribute it within the pool of AGs of the same domain; while the reference BG for that domain is in charge of global spreading). Furthermore, an access interface is associated with the eDB in order to retrieve resource information from it.

In the end, both, efficiency, scalability and backward compatibility issues are addressed with the proposed architecture by leveraging on simple interfaces and usage of IPv6 anycast and multicast addresses.

For the sake of clarity, in Figure 2 the message exchange for the retrieving of a resource information is shown for two cases. In (a), the consumer entity obtains the requested information directly from the closest AG in its domain. While in (b), the consumer entity is first re-directed to the closest AG (i.e. the BG) in the (remote) domain where the issued resource is located, and then finally re-directed to the target producer entity.

## 5. Conclusions

In this paper, the need for cross-layer optimization in the Internet has been outline together with the related key challenges. Also, the limitations of current control and management planes, as well as of proposals in the field have been highlighted.

By a deep analysis of the problem, the critical issues to enable a full optimization across all layers have been identified and the most appropriate design options for addressing them have been discussed.

As a result, an architecture for cross-layer communication supporting data collection, synchronization and provisioning have been designed, yet assuring efficiency, scalability, backward compatibility and flexibility. Indeed, relying on a few additional elements (namely, the AGs, a sort of registrar and directory servers), anycasting and multicasting facilities of IPv6 and publishing-access interfaces for existing databases (e.g. for Traffic Engineering, associated with the implementation of MIH standard, about overlay networks, information base of deployed control and management planes), a consumer entity can get up-to-date and consistent resource information from producer entities anywhere in the Internet at all levels, directly or through the closest (as for network proximity and policies in place) AG of the local or remote domains. This is the very basis for cross-layer optimization across multiple layers, devices, domains and technologies.

The benefits for networked applications in terms of sharp achieving of target QoS guarantees, and for the underlying network in relation to more efficient resource utilization and management, reflect on direct advantages for service and internet providers, who can better interact (timely and consistently exchanging all the relevant information) in a synergic way, for the support of enhanced experience for the users at a low cost and complexity.

In future, a quantitative analysis of the introduced bandwidth and storage memory overheads will be carried out for different application and network scenarios over either a single or multiple domain(s). the collected results will provide helpful data for a complete evaluation of the proposed cross-layer communication architecture and for understanding the most effective way of deploying it in the Future Internet (e.g. position and number of AGs, direct provisioning of data versus re-direction, signaling protocol(s) for message exchange, communication model and refreshing frequency for a given resource information).

## References


[1] G. Tselentis et al., “Towards the Future Internet - Emerging Trends from European Research”. IOS Press, 2010

[2] P. Krishnan, D. Raz and Y. Shavitt, "The cache location problem". IEEE/ACM Trans. on Networking, vol. 8, 2000, pp. 568-582

[3] E. Cronin et al., "Constrained mirror placement on the Internet". IEEE Journal on Selected Areas in Communications, vol. 20, 2002, pp. 1369-1382

[4] L. Qiu, V. Padmanabhan and G. Voelker, "On the placement of Web server replicas". INFOCOM 2001. Twentieth Annual Joint Conference of the IEEE COMSOC. Proc. IEEE, 2001, pp. 1587-1596 vol.3

[5] S. Ratnasamy et al., "Topologically-aware overlay construction and server selection". INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE COMSOC. Proc. IEEE, 2002, pp. 1190-1199 vol.3
[6] A. Mahimkar et al.,"Towards automated performance diagnosis in a large IPTV network". Proc. of the ACM SIGCOMM 2009 conference on Data communication, Barcelona, Spain: ACM, 2009, pp. 231-242
[7] Z. Fei, M. Ammar and E. Zegura, "Multicast server selection: problems, complexity, and solutions". IEEE Journal on Selected Areas in Communications, vol. 20, 2002, pp. 1399-1413
[8] P. Quax et al., "Dynamic server allocation in a real-life deployable communications architecture for networked games". Proc. of the 7th ACM SIGCOMM Workshop on Network and System Support for Games, Worcester, Massachusetts: ACM, 2008, pp. 66-71
[9] S. Gargolinski, C.S. Pierre and M. Claypool, "Game server selection for multiple players". Proc. of 4th ACM SIGCOMM workshop on Network and system support for games, Hawthorne, NY: ACM, 2005, pp. 1-6
[10] M. Suznjevic, M. Matijasevic and O. Dobrijevic, "Action specific Massive Multiplayer Online Role Playing Games traffic analysis: case study of World of Warcraft". Proc. of the 7th ACM SIGCOMM Workshop on Network and System Support for Games, Worcester, Massachusetts: ACM, 2008, pp. 106-107
[11] B. Krishnamurthy, C. Wills and Y. Zhang, "On the use and performance of content distribution networks". Proc. of the 1st ACM SIGCOMM Workshop on Internet Measurement, San Francisco, California, USA: ACM, 2001, pp. 169-182
[12] J. Agogbua, M. O'Dell and J. McManus, “Requirements for Traffic Engineering Over MPLS”. IETF RFC 2702, September 1999
[13] K. Shiomoto and D. Brungard, ” Generalized MPLS (GMPLS) Protocol Extensions for Multi-Layer and Multi-Region Networks (MLN/MRN)”. IETF RFC 6001, October 2010
[14] A. Reibman and M.-T. Sun, “Compressed Video Over Networks”.New York: Marcel Dekker, 2000
[15] M. van der Schaar and S. Shankar, “Cross-layer wireless multimedia transmission: Challenges, principles, and new paradigms”. IEEE Wireless Commun., vol. 12, no. 4, Aug. 2005, pp. 50–58
[16] X. Wang, Q. Liu and G. B. Giannakis, “Analyzing and optimizing adaptive modulation coding jointly with ARQ for QoS-guaranteed traffic”. IEEE Trans. Veh. Technol., vol. 56, no. 2, Mar. 2007, pp. 710–720
[17] T. Holliday, A. Goldsmith and P. Glynn, “Optimal power control and source-channel coding for delay constrained traffic over wireless channels”. In Proc. IEEE Int. Conf. Commun., May 2002, vol. 2, pp. 831–835
[18] F. Fu and M. van der Schaar, “A New Systematic Framework for Autonomous Cross-Layer Optimization”. IEEE Trans. On Vehicular Technology vol. 58, no. 4, May 2009
[19] IETF ALTO WG charter, http://datatracker.ietf.org/wg/alto/charter/
[20] FP7 ICT ALICANTE home page, http://www.ict-alicante.eu/
[21] FP7 ICT ENVISION home page, http://www.envision-project.org/index.html
[22] FP7 ICT ADAMANTIUM home page, http://www.ict-adamantium.eu/
[23] 3gpp IMS home page, http://www.3gpp.org/article/ims
[24] D. Harrington et al., "An Architecture for Describing SNMP Management Frameworks". IETF RFC 2261, January 1998
[25] B. Wijnen et al., "View-based Access Control Model (VACM) for the Simple Network Management Protocol (SNMP)". IETF RFC 2265, January 1998
[26] R. Enns, “NETCONF Configuration Protocol”. IETF RFC 4741, December 2006
[27] T. Nadeau, “A Framework for Multi-Protocol Label Switching (MPLS) Operations and Management (OAM)”. IETF RFC 4378, Februry 2006
[28] G. Swallow, H. Ishimatsu and Y. Rekhter, “Generalized Multiprotocol Label Switching (GMPLS) User-Network Interface (UNI): Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Support”. IETF RFC 4208, October 2005
[29] ITU-T Y.2011, “General principles and general reference model for Next Generation Networks”. October 2004
[30] ITU-T Y.2012, “Functional Requirements and architecture of the NGN”. April 2010
[31] K. Lindqvist, “Operation of Anycast Services”. IETF RFC 4786, December 2006
[32] T. Pusateri, “Protocol Independent Multicast - Sparse Mode (PIM-SM)”. IETF RFC 4602, August 2006
[33] Internet Assigned Numbers Authority (IANA), http://www.iana.org/
[34] Y. Rekhter, T. Li and S. Hares, ” A Border Gateway Protocol 4 (BGP-4)”. IETF RFC 4271, January 2006
[35] IEEE 1588-2002, “IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems”. 2002
[36] D. L. Mills, “Internet Time Synchronization: The Network Time Protocol”. IEEE Trans. Commun., vol. 39, no. 10, Oct. 1991, pp. 1482–93
[37] J. Postel, "Transmission Control Protocol". IETF RFC 761, January 1980

[38] J. Postel, “User Datagram Protocol”. IETF RFC 768, August 1980
[39] T. Dierks, ”The Transport Layer Security (TLS) Protocol Version 1.2”. IETF RFC 5246, August 2008
[40] E. Rescorla and N. Modadugu, “Datagram Transport Layer Security”. IETF RFC 4347, April 2006
[41] A. Larzon et al., “The Lightweight User Datagram Protocol (UDP-Lite)”. IETF RFC 3828, July 2004
[42] E. Kohler et al., “Datagram Congestion Control Protocol (DCCP)”. IETF RFC 4340, March 2006
[43] R. Stewart, “Stream Control Transmission Protocol”. IETF RFC 4960, September 2007
[44] H. Schulzrinne et al., “SIP: Session Initiation Protocol”. IETF RFC 3261, June 2002
[45] A. Conta and S. Deering, “Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification”. IETF RFC 2463, December 1998
[46] S. Deering et al., “Internet Protocol, Version 6 (Ipv6) Specification”. IETF RFC 2460, December 1998
[47] E. Piri, T. Sutinen and J. Vehkaperä, “Cross-layer Architecture for Adaptive Real-time Multimedia in Heterogeneous Network Environment”. In Proc. EW2009, Aalborg, Denmark, May 2009
[48] K. Ishiguro et al., “Traffic Engineering Extensions to OSPF Version 3”. IETF RFC 5329, September 2008
[49] T. Li and H. Smit, “IS Extensions for Traffic Engineering”. IETF RFC 5305, October 2008
[50] H. Ould, D. Fedyk and Y. Rekhter, “BGP Traffic Engineering Attribute”. IETF RFC 5543, May 2009