Dimensioning of Next Generation Networks Signaling Gateway for Improving a Quality of Service Target

Cherif Ghazel, and Leila Saïdane
The National School of Computer Sciences.
University of Manouba - Tunisia
Cherif.Ghazel@Email.ati.tn, Leila.Saidane@Ensi.rnu.tn

Abstract

In this paper, we propose a set of dimensioning rules, which deliver high quality session-based services over a Next Generation Network based IP/MPLS transport infrastructure. In particular, we develop a detailed dimensioning methodology for improving a target QoS requirement. The proposed methodology outlines an optimal equipment allocation strategy for a requested capacity. The benefits of operating a network under the paradigm of generous dimensioning, for converged multiservice traffic flows, include target QoS guarantee, scalability, and network resilience. We present and discuss experimental results which illustrate a practical implementation of the proposed dimensioning strategy and its benefits.

Keywords: *NGN QoS, dimensioning, signaling traffic surveying, signaling Gateway, IP/MPLS.*

1. Introduction

The Next Generation Networks (NGNs) are developed to supports a broad range of services in any media, over any facilities, anytime, anywhere, and in any volume. Despite their advantages, NGNs are not without limitations and further research and development may be required when they become in service. The next step once NGNs are deployed is to monitor them so that the promised QoS can be delivered. One of the main issues related to NGNs, which has been the focus of several works, is the end-to-end QoS guarantee for multiservice traffic. For instance, a dimensioning method for a Multiservice Gateway (MSG) voice service was proposed in [1]. This method was limited to defining dimensioning rules related to a MSG carrying voice traffic. However, it did not discuss the NGN traffic modeling neither did it propose a solution for data and signaling traffic computing and carrying. Also, in [2], various optimizing techniques for dimensioning the MPLS networks as the underlying transport technology for NGNs are outlined. However, these techniques were proposed for a specific category of networks and lack several functionalities, which are necessary for a standard NGNs architecture dimensioning. A method limited to defining dimensioning rules related to the future transport networks for real time telephony application was also proposed in [3]. Several other network dimensioning methods were published in the literature [4], [5] and [6].

In this work, we build on previous contributions, such as in [1-6] and propose a novel methodology for improving NGNs QoS through dimensioning. The proposed methodology takes several additional factors into considerations and provides the desired improvement in order to satisfy requirements in terms of QoS. We propose additional rules and analytically derive solutions for dimensioning one of the main NGN entities, which is the Signaling Gateway (SG). We also study and analyze, in details, the signaling traffic for an NGN based IP/MPLS transport network.

For our practical purposes, we will also present and discuss experimental results which illustrate the practical implementation and advantages of the proposed dimensioning method in terms of performance gain and entity dimensioning benefits. In particular, we shall investigate the performance of the SG dimensioning strategy by experimentally assessing the SG capacity for the generated signaling traffic, which simplifies the selection of adequate range of equipments, for the required capacity.

The organization of this paper follows a standard format consisting of three sections. In the first section, we model the NGNs multiservice traffic, where we mainly survey the voice and data and signaling traffic, and treat the NGN traffic distribution and behavior. The second section is focused on Signaling Gateway entity dimensioning and the proposed QoS improvement. The third section presents experimental results' discussions and performance analysis illustrating the efficiency of our method. Summary and concluding remarks are presented in the last section.

2. NGN traffic modeling and analysis

In a telecommunication network, users transmit and receive various forms of data including speech, images and text. The analysis of these various natures and forms of demands of data transmission is conducted at two different time-scales: *packet* and *flow* [2]. The packet traffic between any pair of nodes may be modeled in terms of flows or aggregates [2]. For the flow traffic, it is typically modeled as a Poisson process [5], [5] and [8]. Practically, the user *session* traffic consists of successions of high-volume flow interrupted by periods of inactivity. The use of average traffic, which may be interpreted as the *demand*, allows us to better estimate the required dimensioning model. A vector may be used to denote the entering demand:

$$D = \left(d_1 \cdots d_Q\right)^T \qquad (1)$$

Where the superscript T stands for the vector transpose operation, and each entry d_q represents the traffic value required for each source-destination pair $q = (m, n)$, for $q=1, 2 \ldots Q$, and (m, n) is the node pairs, with $m \neq n$ and $Q = N(N - 1)$.

2.1. Network graph layout and traffic cases.

The NGN architecture and various types of NGN traffic flows cases are illustrated in Fig.1.

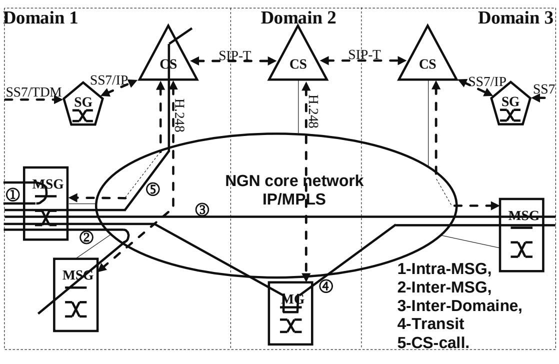

Figure 1. NGN architecture and traffic cases

Typically, the NGN physical network is modeled as a directed graph $G = \{V, E\}$ where V is the set of nodes indexed, $i = 1, 2\ldots n$ and E is the set of directed links indexed, $j = 1, 2\ldots m$. Suppose that, link l corresponds to a transmission line with capacity c_l, and C denotes the m-dimensional capacity vector:

$$C = \left(c_1 \cdots c_m\right)^T \tag{2}$$

Generally, an NGN network considers multiple paths connecting a pair of nodes. Let the node pair *(m, n),* with $m \neq n$, be indexed by $1, 2\ldots Q$. The total number of node pairs may be written as follows:

$$\mathrm{Q} = \mathrm{N}(\mathrm{N} - 1) = \sum_{i=1}^{n-1}\sum_{j=i+1}^{n} q_{ij} \tag{3}$$

Where N represents the set of the graph nodes.

Moreover, let the total traffic-path matrix denoted by A represents the different acyclic traffic-paths P_{ij} indexed by $a_{ij}^1, a_{ij}^2, \ldots, a_{ij}^{P_{ij}}$ and implemented for connecting a source entity i, to a destination entity j, for $i=1, 2\ldots m$ and $j = 1, 2\ldots n$, with $i \neq j$ and $m \neq n$. The source and destination entities could be one of the following types: MG, CS, SG or core node. This total traffic-path matrix A corresponding to the different acyclic paths linking a node i to a node j over the NGN network graph, can be represented as follows:

$$A = \left(\begin{array}{ccc|c|ccc} a_{11}^1 & \cdots & a_{1n}^1 & \cdots & a_{m1}^1 & \cdots & a_{mn}^1 \\ \vdots & \cdots & \vdots & \cdots & \vdots & \cdots & \vdots \\ a_{11}^{P_{11}} & \cdots & a_{1n}^{P_{1n}} & \cdots & a_{m1}^{P_{m1}} & \cdots & a_{mn}^{P_{mn}} \end{array} \right) = \left(A_1 \middle| A_2 \middle| \ldots \middle| A_m\right) \tag{4}$$

Where, $a_{ij}^{P_{ij}}$ is the traffic-path value identified for each path linking a source entity i to a destination entity j. The matrix A represents the total path-traffic matrix corresponding to the total number of node pairs. Note that the diagonal elements of A are non-zero since we consider the intra-site traffic [5].

Note that the superscript is not an exponent, but it is an index of the acyclic paths linking the node pairs *(i, j)*.

3. Signaling gateway dimensioning

Physically, the Signaling Gateways (SGs) are implemented in a distributed way and represent the concentration points of the signaling traffic. They convert the traffic signaling exchanged between the NGN networks and the external connected networks. Notably, they ensure the signaling adaptation between the TDM networks and the packet-based IP/MPLS transport network. The dimensioning of the SG is realized by determining the volume and capacity of the signaling traffic conversion and adaptation as well as the number and type of interfaces in the SG access and core sides.

In the following subsection, we determine the capacity of the SG in signaling traffic conversion and adaptation in order to assess its performance.

3.1. Signaling traffic at the access level

For the signaling traffic, each type-k access indexed $1, 2\ldots k$, may use one of its corresponding q_k signaling (*ISUP*, *TUP*, etc.) indexed $1, 2\ldots m_k$. For each call, the number and length of used signaling exchange messages (in bytes/s) depend on the direction (uplink/downlink) and the call status (succeeded/failed), as shown in [1, 9]. In fact, the signaling traffic and its corresponding *E1* may be calculated as follows:

a) First, the number of calls per second, which use different types of *q*-signaling is:

$$Y_{sig} = \sum_{k=1}^{n} \sum_{q=1}^{m_k} Y_{k,q} \tag{5}$$

Where *q* represents the used signaling (e.g. *ISUP*, *TUP*, *INAP*, etc) per voice access service or Intelligent Network (*IN*) service and *k* signifies the different access service types.

b) Then, the required bandwidth for the signaling message traffic handled by a *SG* is determined as:

$$A_{sig} = \sum_{i=0}^{1}\sum_{k=1}^{n}\sum_{q=1}^{m_k} Y^i_{k,q} Li_{q,i} Ni_{q,i} + \sum_{i=0}^{1}\sum_{k=1}^{n}\sum_{q=1}^{m_k} Y^i_{k,q} Lo_{q,i} No_{q,i} \tag{6}$$

c) Thus, the corresponding *E1* may be derived from the above signaling message traffic *BW* as follows:

$$N_{E1,sig,acc} = \left(A_{sig,\ \phi/s} \times 8.10^{-6}\right)/2.048 \tag{7}$$

Where, the factor 8.10^{-6} is used to convert results to Mbps, and *E1* interface is part of a higher order Line Interface Board. It composes of 30 voice channels, 01 signaling, and 01 synchronization channels. Besides, the quantities Li_q and Lo_q represent the length of the *q* signaling messages (in bytes/s) sent in the uplink and downlink directions, respectively. Similarly, Ni_q and No_q, represent the number of such messages. The indices *q* and *i* represent the used control signaling and the call status (succeeded /failed), respectively. Note that the superscript *i* is not an exponent, but it is an index of the call status.

After determining the total signaling message traffic BW and the required interfaces in the SG access side, next we address the same issue at the SG core level.

3.2. Signaling traffic at the core level

After conversion, the signaling traffic is carried on packet switching-based SCTP/IP/MPLS Tunnels or LSPs established through the core network towards the Call Server (CS). The choice of SCTP for carrying the control signaling messages is justified by its potentialities and advantages over UDP or TCP, in establishing secure and reliable connections. The SCTP packet is composed of a common header, with size of 12 bytes and one or multiple bundled chunks. Each chunk is formatted with a block header, with size of 16 bytes and a variable data chunk payload "*D*". Furthermore, the IP packaging and MPLS labeling require an additional 20 and 4 bytes of overhead, respectively. The total length of these headers is then 52 bytes, which is transmitted each time a signaling packet is sent. As illustrated in Figure 2, in the simple case, we assume that an SCTP packet consists of a header followed by only one data chunk with fixed "*D*" payload length in bytes. As discussed in [10], the total chunk-length should be a multiple of 4 bytes. If this is not the case, the sender needs to pad the chunk with "*P*" bytes, where *P = 1, 2, or 3*.

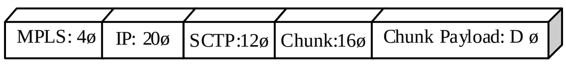

Figure 2. Signaling traffic encapsulation

In view of this discussion, the required BW corresponding to the *E1* circuit emulation of the access side at the IP/MPLS core side can be computed using the following principle:

a) First, a voice signaling channel with 64 Kbps requires a bandwidth (BW) equal to:

$$64 \times (52+D+P)/D, \text{ in Kbps} \tag{8}$$

Where D is the data chunk payload and P is the number of added padding bytes.

b) Thus, the BW (in Mbps) reserved for the control signaling message traffic (e.g. SS7/SCTP/IP/MPLS messages) exchanged between the Call Server (CS) and the Signaling Gateway (SG) is given by:

$$BW_{sig\_core} = \left(\sum_{i=0}^{1}\sum_{k=1}^{n}\sum_{q=1}^{m_k} Y_{k,q}^{i} Li_{q,i} Ni_{q,i} + \sum_{i=0}^{1}\sum_{k=1}^{n}\sum_{q=1}^{m_k} Y_{k,q}^{i} Lo_{q,i} No_{q,i} \right) \times (52+D+P)/D \text{ x } 8.10^{-6} \tag{9}$$

c) Indeed, the total BW requirement (in Mbps), at the IP/MPLS core level, needed for carrying the total signaling messages traffic is given by:

$$BW_{sig\_core} = A_{sig,Ø/s} \times 8.10^{-6} \times (52+D+P)/D \text{ Mbps}$$
$$= N_{E1,sig,acc} \times 2.048 \times (52+D+P)/D \text{ Mbps} \tag{10}$$

d) Accrdingly, the corresponding *E1* can be derived from the above BW as follows:

$$N_{E1,sig,core} = \frac{BW_{sig\_core}}{2.048} \tag{11}$$

Note that P may be used only if the total chunk-length is not a multiple of 4 bytes. Also, the superscript i is not an exponent, but it is an index of the call status.

3.3. Experimental results and discussion

In this section, we present and discuss experimental results which illustrate the practical implementation and advantages of the proposed dimensioning method in terms of performance gain and entity dimensioning benefits. To generate the experimental results, we developed an NGN entities' dimensioning software tool, which was used for dimensioning the different NGN components based on the different access network traffic and corresponding signaling. For our dimensioning purposes, we implemented an NGN architecture based on a transport network composed of several IP/MPLS edge and core nodes and four Media Gateways (MGs), and a control level consisting of two Signaling Gateways (SGs) and one Call Server (CS).

It should be noted that the selected values of all input parameters used in this simulation are not absolute fixed values. They are only suitable and practical values used for our case study and practical purpose. Also, for simplicity, it is assumed that the message number and length in the uplink and downlink directions for various signaling traffic flows are the same. As well, the selected services may vary from one entity to another.

3.3.1. Signaling gateway dimensioning implementation: As listed in Table 1, for the SG dimensioning, we specified the number of IN service calls per second, the number and length of service messages for the cases of success and failure service calls. Also, Table 2 lists the selected number and length of each call signaling message for the cases of success and failure calls, and Table 3 illustrated the chosen number and length of the control signaling messages for the cases of success and failure calls. For all practical purposes, the proportions of failed and successful calls are set to be 0.2 and 0.8, respectively.

Table 1. Intelligent network calls features

IN access type	Call number	MSU length in octet (ϕ)	Nb. MSU per success call	Nb. MSU per failure call
Prepaid	20	500	2	1
Free phone service	30	430	2	1
Premium rate service	5	400	2	1

Table 2. Service calls features

Voice access type	Call number	MSU length in ϕ	Nb. of MSU per success call	Nb. of MSU per failure call
RTPC	5 000	15	3	2
ISDN-BA	3 000	30	3	2
ISDN-PRA	2 000	30	3	2
DIAL-UP	800	50	3	2

Table 3. Control signaling features

Control Signaling with the CS	MSU length in ϕ	Number of MSU per success call	Number of MSU per failure call
SS7/SCTP/IP/MPLS	52 + 45	5	3

Once the parameters concerning the IN calls' signaling and control signaling have been appropriately specified, we have completed the dimensioning stage. The output of the SG dimensioning, consisting of the call signaling processing (traffic conversion and adaptation) and corresponding interfaces, is illustrated in Table 4.

Table 4. SG Call signaling and interfaces

Call signaling /s	Signaling traffic in ϕ	E1 Interfaces per side
48	28 973	01

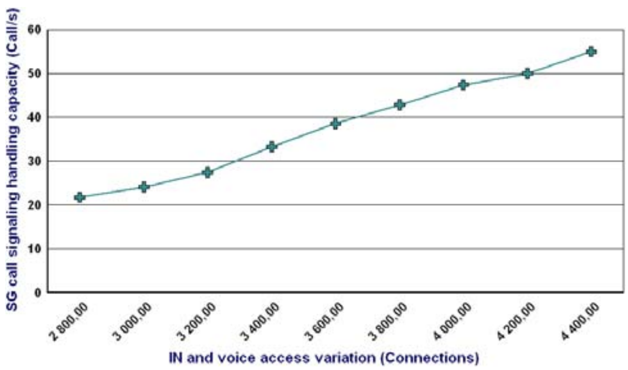

Figure 3. SG performance evaluation as a function of the IN and call signaling variation

3.3.2. Performance evaluation: At this stage, we varied the value of inbound traffic volume generated by various access networks connected to the Media Gateways in order to assess the variation of the SG capacity as a function of generated signaling traffic. Fig. 3 illustrates the variation of the SG capacity in call signaling traffic conversion and adaptation according to the generated traffic variation. This figure clearly shows that the SG processing capacity and performance variation strongly depend on the overall active voice and IN connections in the access network level. In practice, assessing the dependence of the capacity of the SG on the signaling traffic volume generated by different voice and IN access types can provide us with a strategy for selecting the adequate range of equipment and meeting the required capacity. Other benefits may include avoiding under and over dimensioning, minimizing incoming traffic blockage and congestion, and enhancing the required QoS.

4. Conclusion

In this work, we investigated the fundamental multi-service traffic relationships between the performance, capacity and QoS requirements for high-speed NGN networks. Dimensioning a SG entity involves exploiting available related traffic flows in order to estimate the necessary capacity. This capacity may be in terms of physical devices performance.

We presented and discussed experimental results which illustrate the practical implementation and advantages of the proposed dimensioning method in terms of performance gain and entity dimensioning benefits. In particular, we investigated the performance of the SG dimensioning method by experimentally assessing the SG capacity for the generated signaling traffic. In practice, the selection of adequate range of equipment, which meets the required capacity, is determined based on the dependence of the capacity on generated signaling traffic.

References

[1] "Multiservice Gateway MSG Dimensioning", Ericsson AB, 2002.

[2] S. Beker, and D. Kofman, "Techniques d’Optimisation pour le Dimensionnement et la Reconfiguration des Réseaux MPLS", Télécom Paris-ENST, 2004, pp. 25-73.

[3] G. Van Hoey, S. Van Den Bosch, H. De Neve, and G. Petit, "Le dimensionnement de futurs réseaux de transport pour les applications téléphoniques en temps réel", Alcatel, 2001.

[4] W. Ogryczak, A. Wierzbicki, and M. Milewski, "Fair and efficient network dimensioning with the reference point methodology", Journal of Telecommunications and Information Technology, April 2006.

[5] Z. Zhao, Q. Liu, and H. Guan, "A Method and Simulation Study: Network Dimensioning of the Broadband Residential Ethernet-based Access Network", Nokia white paper, 2004.

[6] T. Borland, P. Olivier, and J. Roberts "Dimensioning high speed IP access networks", France Telecom R&D, 2004.

[7] M. Nabe, M. Murata, and H. Miyahara, "Analysis and Modelling of Worldwide Web Traffic for Capacity Dimensioning of Internet Access Lines Performance Evaluation", 1998, pp. 246-271.

[8] B. Krithikaivasan, K. Deka, and D. Medhi, "Adaptive Bandwidth Provisioning Envelope Based on Discrete Temporal Network Measurements", Proceedings of IEEE INFOCOM 2004, Hong Kong, March 2004.

[9] "Signaling System No. 7 - Survey", Ericsson, 2001.

[10] R. Stewart, Q. Xie, K. Morneault C. Sharp, H. Schwarzbauer, T. Taylor, I. Rytina, M. Kalla, L. Zhang, and V. Paxson, "Stream Control Transmission Protocol", IETF RFC 2960, October 2000.

Authors

Mr. **Cherif Ghazel** received his Master's Degree in Telecommunication from The School of Telecommunications Engineering, at University of Carthage, Tunisia in 2004. He is currently a PhD candidate at the National School of Computer Sciences, at The University of Manouba, Tunisia, and member of the CRISTAL Research Laboratory. He has been involved in many research projects on computer sciences, networking and communications systems. His research interests include: Next Generation Networks (NGN), Network Architectures, Access and Core Protocols, Admission Control, Resource Control and Management, and QoS.

Professor **Leila A. Saïdane** is the Director of the National School of Engineering Computer Science (ENSI), at The University of Manouba, in Tunisia. She is the co-director of RAMSIS pole of CRISTAL Research Laboratory (Center of Research in Network and System Architecture, Multimedia and Image Processing) at ENSI. She collaborated on several international projects. She is author and co-author of several papers in refereed journals, magazines and international conferences. She is the supervisor of the Master's Degree program in Networks and Multimedia Systems and the Chairperson of the PhD and Habilitation Commission at ENSI.