On the Evaluation of the Extended Generic Internet Signalling Transport Protocol

Rui Campos, Nadeem Akhtar, Cornelia Kappler, Pekka Pääkkönen,
Petteri Pöyhönen, Di Zhou

Abstract — **In the Beyond 3G (B3G) vision convergence and plug&play are central concepts. The convergence of different networks, originally targeting different applicability domains, is envisioned for enabling ubiquitous connectivity and a broader set of services from users' perspective. Also, the advent of small user owned networks dynamically connecting to other networks while users move around and the increasing dynamics of the relationships between network operators, so far statically established, lead to new paradigms. Considering this, the IST Ambient Networks project is investigating two new concepts, Network Composition and Ambient Network, that enable the dynamic and uniform interworking of networks regardless their type or size. The Generic Ambient Network Signalling (GANS) protocol is defined to support this dynamic interworking. GANS is divided in two parts: application layer and transport layer. This paper concentrates on the transport layer which includes the Extended Generic Internet Signalling Protocol (EGIST). The evaluation of EGIST from scalability viewpoint by means of simulations and theoretical analysis is presented.**

Index Terms—**Convergence, Dynamic Interworking, GANS, Extended Generic Internet Signalling Transport**

I. Introduction

Convergence of different networks originally targeting different applicability domains is envisioned to happen in the B3G scope. Along the years, with the exception of the Internet, the guiding principle has been to have networks and the application services they support tightly connected; cellular networks have been an illustrative example of this paradigm. Also, independent and specific technologies for data transportation and network control have been deployed for each type of network, leading to heterogeneities within both the data plane and the control plane. In the B3G context, this heterogeneity is expected to disappear. A clear separation between the network infrastructures and the services they support, and the uniform control over the multiple network types is being proposed. On the one hand, this vision considers an evolution towards the model followed in the development of the Internet – the network just transports packets back and forth and is application service agnostic; the Internet Protocol (IP) acts as the basis for the integration of the current and future wireless and wired technologies. On the other hand, it considers convergence at the control plane in addition to today's data plane convergence; the control plane framework becomes the same independently of the underlying networking technology being used.

In addition to the move towards convergence, another growing trend is plug&play networking [1,2,3]. Nowadays, it is common for users to have quite a few communication devices, such as laptop, mobile phone, and Personal Digital Assistant (PDA). In the future, these devices may fully self-organise to form small low-complexity networks, such as Personal Area Networks (PANs), that may dynamically interwork with each other or with access networks which, in turn, may dynamically interwork with core networks. This leads to a new communication model, where networks connect/disconnect dynamically to other networks, and the relationships established between network operators and/or network operators and service providers are dynamic. In this dynamic environment, where cooperating and competitive operators coexist, and technology should adapt almost instantly to the rapidly changing networking contexts, plug&play of networks becomes crucial.

The concepts of Ambient Network (AN) and dynamic interworking of networks, also known as Network Composition, are being studied in the Ambient Networks project [3,4] – one of the integrated research projects of World Wireless Initiative financially supported by the European Union – as a mean of enabling convergence and plug&play interworking between networks of different types and sizes. A key enabler for dynamic interworking of networks is the coordination of control spaces, such as coordination of address spaces and addressing schemes or synchronization of network and policy databases. This requires control signalling between the networks involved. Thus, a generic signalling protocol suite, called Generic Ambient Network Signalling (GANS), for solving this problem has been proposed in [5,6]. GANS is a suite of signalling protocols designed to enable efficient and secure message exchange between related control functions, e.g., QoS and Mobility, of the control spaces of different networks. It follows a two-layer approach. The lower layer comprises two protocols: the Destination Endpoint Exploration Protocol (DEEP) [7] and the Extended Generic Internet Signalling protocol. Herein, we focus on EGIST, namely in its evaluation from a scalability perspective.

The remainder of this paper is organised as follows. Section II presents the state of the art with respect to signalling

Manuscript received April 17th, 2006.
Rui Campos is with INESC Porto, Portugal (email: rcampos@inescporto.pt).
Nadeem Akthar is with the University of Surrey, UK (email: N.Akhtar@surrey.ac.uk).
Cornelia Kappler, is with Siemens Communications, Berlin, Germany. (e-mail: cornelia.kappler@siemens.com).
Pekka Pääkkönen, is with VTT Technical Research Centre of Finland, Oulu, Finland (e-mail: Pekka.Paakkonen@vtt.fi).
Petteri Pöyhönen is with Nokia, Helsinki, Finland (e-mail: petteri.poyhonen@nokia.com).
Di Zhou is with PSE/Siemens, Vienna, Austria (e-mail: di.zhou@siemens.com).
Note: Names of co-authors have been listed in alphabetical order.

transportation over heterogeneous networks. Section III describes the EGIST protocol design, and Section IV depicts the simulation and theoretical analysis used as basis for the evaluation presented in Section V. Section VI draws the conclusions.

II. State of the Art

With the advent of Quality of Service (QoS) requirements and the introduction of middleboxes, such as Network Address Translators (NATs) and Firewalls, the Internet has evolved towards a new network with architectural changes. As a result, the installation, maintenance, and removal of control state information in certain nodes using signalling protocols became a reality. However, standard transport protocols, e.g., Transport Control Protocol (TCP) and User Datagram Protocol (UDP), simply provide transport services and require that signalling applications pick up the proper destination addresses for a signalling message. Hence, concerning, for instance, signalling along a data path that contains one or more firewalls and NATs requires a per signalling application mechanism for discovering each of these nodes, and the addressing of separate signalling messages towards each of them. This brings up inefficiencies since the discovery process has to be defined and carried out by each signalling application independently; a generic and application-independent discovery mechanism could be defined instead.

Some signalling protocol suites have been designed in the IETF in order to overcome these problems. The Resource ReServation Protocol (RSVP) [8] has been designed to provide a mechanism for QoS signalling across a data path. Nonetheless, it is an application-specific signalling protocol and was not originally designed for accommodating various signalling applications besides QoS signalling. Considering the shortcomings of RSVP, and namely the lack of modularity and extensibility, the Cross-Application Signalling Protocol (CASP) has been proposed in [9]. CASP enhances the RSVP model namely to allow reliable and secure signalling transportation and signalling peer discovery along a data path. It defines a modular and extensible two-layer model, where the lower layer, CASP Message Layer (ML), implements the features common to all signalling applications, and the upper layer, Client layer, comprises the signalling applications themselves and a special purpose client for signalling peer discovery. CASP represented the first steps in the development of a currently prominent protocol suite – Next Steps in Signalling (NSIS) – under research in the IETF NSIS WG. The NSIS protocol suite [10] considers a similar approach, but the distribution of tasks across the layers and some design decisions are slightly different, e.g., signalling peer discovery mechanism is built-in in the lower layer. Besides, the terminology used is different – the lower layer is called NSIS Transport Layer and the upper layer is called NSIS Signalling Layer. In [11] the GIST protocol is proposed as a solution for the NSIS Transport Layer defined in the NSIS framework [10]. Like the CASP ML, GIST provides the general message transport services, such as finding signalling peers, establishing security associations, and transporting messages; it relies on standard TCP/IP protocols for actual message transportation. In spite of its modularity and extensibility, so far the GIST protocol has been mostly designed to provide control signalling related to data flows, where the control signalling follows the data path. In order to have GIST also supporting non data flow related signalling, namely signalling for dynamic interworking of networks, it has to be extended.

III. Extended Generic Internet Signalling Transport Protocol Design

The EGIST protocol is defined within the GANS Transport Layer Protocol (GTLP) of the GANS protocol suite [5,6]. GANS, designed to meet special requirements of Ambient Networks [3,4], represents a backwards compatible extension to the basic NSIS protocol suite; its structure is shown in Fig. 1 where all new functionalities extending NSIS are illustrated by dashed lines. Consequently, EGIST itself represents a backwards compatible extension to GIST. It comprises all features already deployed by GIST plus additional features required for GANS concerning dynamic interworking of networks. EGIST represents the actual transport framework provided by GTLP. As illustrated in Fig. 1, it interacts with the signalling applications – those defined within the NSIS context (NSLPs) and the applications to be defined in the GANS context (GSLPs) – through the so-called GTLP API, and uses standard protocols to provide the transport and security services to the signalling applications; it also interacts with the DEEP protocol via the DEEP internal interface in order to resolve the abstract names used by the signalling applications for identifying the destination signalling endpoints.

The new features added by EGIST on top of those provided by GIST are:

- support for abstract names (e.g., *GSLPx.Network1*) to address signalling endpoints, in addition to IP addresses already supported by GIST;
- interaction with a new protocol implementing a name resolution scheme;
- support of single-hop oriented signalling applications, in addition to multi-hop signalling applications supported by NSIS, e.g., QoS signalling [12], where more than two NSIS nodes along the data path are involved in the signalling process;
- name binding state storage and maintenance without modifying the message routing state (MRS) table specified by GIST for storing state information related to message routing towards the right EGIST peer;
- name binding state update procedure;
- indirect signalling configuration.

GIST defines the concept of Message Routing Method (MRM) [11] for routing signalling messages towards the proper adjacent peer. The MRM defines the algorithm for discovering the route that the messages should follow between the source and destination endpoints at the GIST layer; in order to forward messages between GIST nodes, GIST relies on the IP layer routing. GIST supports multiple MRMs; the default MRM is data flow related, so-called "path-coupled" MRM [11]. In conjunction with the MRM concept, the protocol defines the Message Routing Information (MRI) object, which allows signalling applications to specify the concrete MRM that GIST should use to forward their signalling messages, apart from identifying, e.g., the flow for which signalling is being performed. In addition, GIST specifies two operation modes named datagram mode (d-mode) and connection mode (c-mode); UDP and TCP, respectively, are the standard transport protocols supporting them. When using c-mode, a so-called message association between the signalling endpoints is established. GIST defines a table, called Message Association (MA) table, for storing the state information related to each established MA; for each MA table entry a MRS table entry is created and linked to the former. MA reuse is specified to enable multiplexing of multiple signalling sessions over a single association. Further details on GIST are presented in [11].

Considering the nature of the GANS signalling applications (GSLPs), a new MRM is defined in EGIST, called “path-decoupled/single-hop” MRM; it supports exchange of non data flow related signalling between signalling endpoints (GSLPs) that are one “GSLP hop” apart. This new MRM requires modifications concerning the internal processing of the messages received from the signalling applications, namely regarding the way messages are routed and the way the two operation modes of GIST are used; the newly MRM has been defined according to the guidelines provided in the GIST specification, so that it can be seamlessly integrated in the GIST protocol specification. The new corresponding MRI object includes the source and destination abstract names and the corresponding IP addresses that can either be filled in by the signalling applications, when they use IP addresses to identify the destination of the signalling messages, or by EGIST after the DEEP protocol is invoked for name resolution and discovery of signalling destination.

The support of abstract names in EGIST requires the storage of state information concerning the mapping between those names and the corresponding IP addresses, so-called Name Binding State (NBS). For this purpose, EGIST uses the MRS table already specified by GIST [11] and encapsulates the NBS inside the new defined MRI object stored in the MRS table.

Another extension to GIST is related to the update of the NBS. When there is a relocation of signalling endpoint, which causes a change in the IP address and/or in the port being used by one of the communicating signalling endpoints, there is a need for updating the corresponding NBS. For this purpose, EGIST reuses the GIST refresh procedure specified in [11] that is based on the exchange of two messages (GIST-Query/GIST-Response); this procedure is still used by EGIST for refreshing the NBS periodically.

A further extension is introduced in EGIST to avoid revealing the internal network topology when two networks are dynamically interworking. For that purpose, EGIST defines the so-called indirect signalling configuration. In this configuration, instead of having signalling endpoints inside each interworking network directly connected, one or more intermediary nodes in each network, called signalling gateways, act as signalling contact points from the outside point of view.

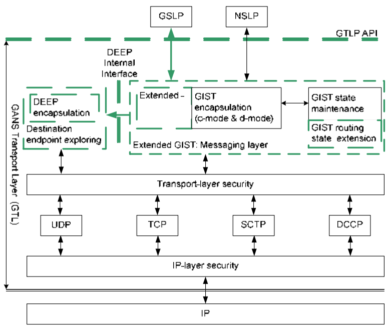

Fig. 1. GANS protocol suite

IV. Simulation and Theoretical Analysis

EGIST is intended to be used in a wide variety of scenarios ranging from simple interworking between devices to interworking between operator networks, i.e., EGIST has to be scalable. Then, the scalability analysis of the protocol is of utmost importance. In the NSIS scope, an overhead and performance study of the GIST protocol, earlier called General Internet Messaging Protocol for Signalling (GIMPS), has been performed; this study is presented in [13]. Herein, our goal is to evaluate the new features added to the base GIST protocol and compare our results with those presented in [13]. The evaluation of EGIST from scalability standpoint was performed by means of simulations; we concentrated on memory consumption and signalling overhead. Suitable simulation scenarios and the corresponding simulation models were defined. The EGIST protocol was implemented over the Network Simulator (NS-2) [14] and the simulation models instantiated accordingly. Apart from the simulation analysis, a theoretical analysis was performed subsequently. For that purpose, some mathematical equations modelling the EGIST procedures were derived; equations providing the memory consumption due to the EGIST MRS and EGIST MA tables, as well as equations providing the total signalling overhead introduced by EGIST concerning the message sending, MA establishment, and the refresh and update procedures were derived [6]; these equations were then used to draw some curves presented in Section V.

The simulation scenarios illustrated in Fig. 2 were defined. Scenario 1 was used to evaluate the EGIST memory consumption, the signalling overhead introduced by the MA establishment and the message transportation procedures, and how MA reuse could improve the EGIST scalability. Scenario 2 was used to evaluate the signalling overhead introduced by the NBS refresh and update procedures.

The first scenario models communication between two interworking networks. Multiple signalling sessions were created between the networks modelled by two nodes in NS-2, and signalling messages were exchanged between them; EGIST c-mode was considered in order to simulate the MA establishment procedure. The varied parameters were the number of sessions created between the two networks and the frequency at which the signalling messages were sent; the measured parameters were the memory consumption due to the MRS and MA tables and the number of signalling messages exchanged between the interworking networks.

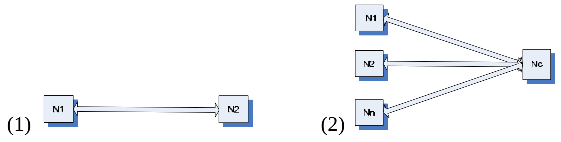

Fig. 2. Simulation scenarios

The second scenario models the dynamic interworking between a varying number of networks (N1, N2, …, Nn) – peer networks – and a central network (Nc); again, each network was modelled by a network node in NS-2. In this scenario, each peer network establishes multiple signalling sessions with Nc and EGIST d-mode is used. Afterwards, each peer network performs periodic refreshes of the NBS and updates the NBS once. The varied parameters were the frequency of refreshes, the number of sessions established between each peer network and Nc, and the number of peer networks; the measured parameters were the signalling overhead due to the refresh and update procedures.

V. Results and Evaluation

This section presents the results obtained in the simulations and through the mathematical equations aforementioned; due to space limits just a subset of the results is presented. Based on the results, the evaluation of the EGIST protocol from scalability standpoint is accomplished.

A. Memory Consumption

Fig. 3 provides the MRS table memory consumption for Scenario 1 when for each signalling session a new MRS table entry is created. In addition, it presents the MA table memory consumption, when MA reuse is applied in Scenario 1 (zero slope line), and the theoretical results obtained for two cases. The first case considers that the signalling sessions are established between a single source network but different destination networks (higher slope line); the second case, considers that 5 sessions each are established between distinct source-destination pairs, which allows MA reuse. The rightmost graph shows that the memory consumption due to the MA table is significantly reduced when MA reuse is applied. Even when only partial reuse is employed (second case) the gain is evident; the same conclusion can be drawn for memory consumption due to the MRS table since there is a one-to-one relationship between the two tables, i.e., for each MA table entry there is a corresponding MRS table entry. On the other hand, the leftmost graph shows that the MRS table memory consumption increases linearly with the number of sessions established between two interworking networks when no MA reuse is applied; since EGIST c-mode was considered in this scenario, this conclusion can be extrapolated for the MA table memory consumption, as illustrated in the rightmost graph.

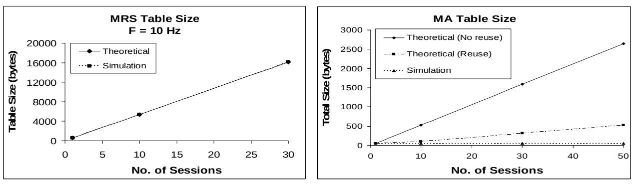

Fig. 3. MRS table and MA table memory consumption

In [13] the authors demonstrate that the memory consumption for the GIST protocol also increases linearly with the number of sessions. Nevertheless, when we compare the size of the MRI object related to the default MRM (24 bytes for IPv4) considered in [13] with the size of the MRI object related to the new defined MRM (225 bytes for IPv4) in EGIST, we conclude that the latter in average (assuming an average abstract name length of 100 bytes) incurs nine times more memory consumption, rendering EGIST scalability poorer than GIST scalability.

B. Signalling overhead

Fig. 4 shows the signalling overhead at Nc in Scenario 2 due to the refresh procedure, when each peer network performs NBS refreshes at two different frequencies (0.1Hz, 0.01Hz) and the number of peer networks and number of sessions established by each of them is varied. Fig. 5 illustrates a comparison between the signalling overhead due to the same process when direct and indirect signalling configurations are used in alternative; the frequency of refresh is 0.01 Hz; the curves invisible in the graph have the same slope as the curve with lower slope. In the indirect configuration, all incoming and outgoing signalling in each network traverses a signalling gateway.

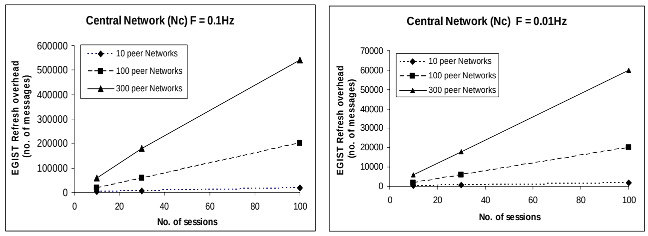

Fig. 4. Signalling overhead due to the refresh procedure

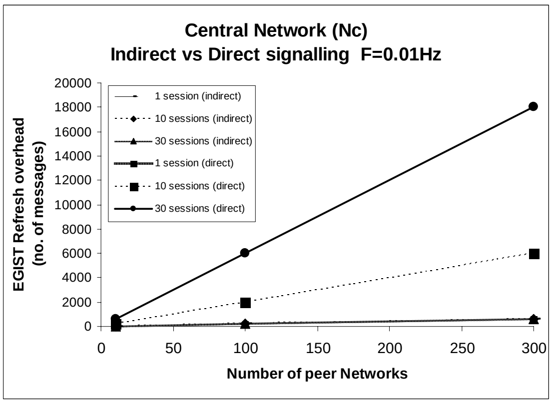

Fig. 5. Signalling overhead for direct and indirect signalling configurations

The whole set of results obtained during the simulation and theoretical analysis [6] demonstrated that the EGIST relevant procedures from scalability standpoint are the MA establishment and the refresh procedures. Concerning the MA establishment signalling overhead no results are provided herein, since the evaluation of the process can be performed based on the conclusions drawn for the MA table memory consumption; the creation of each MA table entry requires a 3-way handshake between the interacting signalling endpoints; therefore, like the MA table memory consumption, the signalling overhead introduced by this procedure increases linearly with the number of sessions and with the number of networks a specific network interacts with; this overhead can also be reduced by applying MA reuse. Concerning the refresh procedure, from Fig. 4 we verify that the signalling overhead is direct proportional to the frequency of refresh – e.g., a reduction of a power of 10 in the frequency of refresh represents a similar reduction in the signalling overhead. Thereby, the frequency of refresh has to be carefully adjusted, so that the signalling overhead does not have a high increasing rate, and the scalability of the protocol is not compromised. The scalability of the protocol is improved when the indirect signalling configuration referred in Section III is used; Fig. 5 shows the gain introduced by this configuration in comparison with direct signalling. Using this approach, a single MRS table entry is created at the signalling gateway in Nc for each peer network it is interworking with; the refresh signalling overhead becomes independent of the number of sessions created between the networks; implicitly, the indirect signalling approach enables MA reuse, as the signalling information between networks can be transported over a single MA established between the signalling gateways of each interworking network. In this context, a trade-off between signalling overhead and fault-tolerance is in place, however. The direct signalling decreases the scalability of the protocol; on the other hand, the indirect signalling configuration has the disadvantage of a single point of failure. For that reason, we claim that the scalability is improved, without affecting the fault-tolerance feature, by adopting a hybrid approach where more than one signalling gateway is deployed in each interworking network.

In [13] the GIST signalling overhead is compared against the RSVP overhead but no GIST scalability analysis is performed at all. Nonetheless, the size of the various GIST messages is provided concerning the default MRM. When comparing such message sizes with the sizes obtained for the new defined MRM, we conclude that the overhead in the latter case is approximately nine times higher. Hence, from signalling overhead perspective, the EGIST scalability becomes poorer than GIST scalability when the new defined MRM is used.

We then conclude that the new defined MRM degrades the EGIST scalability when compared to GIST scalability. However, we verify that EGIST scalability is significantly improved when the indirect signalling configuration is used and MA reuse is applied in EGIST c-mode; such scalability degradation can thus be mitigated. On the other hand, in most of the cases the size of the abstract names is envisioned to be under 100 bytes. As such, further memory consumption and signalling overhead referred above will be lower.

VI. Conclusion

Next generation networks will be characterized by convergence and plug&play interworking between networks of different types and sizes. In the Ambient Networks project the GANS protocol suite has been developed to complement existing protocols to enable this. EGIST implements the transport layer of that protocol suite and represents a backwards compatible extension to the GIST protocol. It does not modify any of the GIST features; rather, it extends the applicability domain of the GIST protocol by providing further features originally not included in it. This paper evaluated EGIST by means of simulations and theoretical analysis and presented a comparison between the results obtained for EGIST and the performance results available for GIST. We verified that the new defined MRM degrades the EGIST scalability when compared to the GIST scalability. However, we concluded that employing the new defined indirect signalling configuration and applying MA reuse renders EGIST significantly more scalable.

Acknowledgment

The authors would like to thank Robert Hancock for enlightening discussions on GIST.

This work is supported by the Ambient Networks Project, partially funded by the European Commission under its Sixth Framework Programme. It is provided “as is” and without any expressed or implied warranties, including, without limitation, the implied warranties of fitness for a particular purpose. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Ambient Networks Project or the European Commission.

References

[1] R. Campos et al., *Dynamic and Automatic Interworking between Personal Area Networks using Composition*, Proc. of the 16th IEEE International Symposium on Personal Indoor and Mobile Radio Communications, Berlin, Sep. 2005.

[2] C. Kappler et al., *A Framework for Self-organized Network Composition*, Proc. 1st IFIP TC6 WG6.6 International Workshop on Autonomic Communication (WAC 2004), Springer LLNC-IFIP Series Vol. 3547., Berlin, Oct. 2004.

[3] N. Niebert et al., *Ambient Networks: An Architecture for Communication Networks beyond 3G*, IEEE Wireless Communications, vol. 11, pp.14-22, April 2004.

[4] IST Ambient Networks Project, http://www.ambient-networks.org.

[5] N. Akhtar et al., *GANS: A Signalling Framework for Dynamic Interworking between Heterogenous Networks*, submitted to 64th IEEE Vehicular Technology Conference, Montreal, Canada, Sep. 2006.

[6] I. Svinnset, Ed., et al., *Connecting Ambient Networks – Final Architecture, Protocol Design and Evaluation*, IST–2002-507137-AN/WP3/D/3-3, Ambient Network WP3 Deliverable 3-3, Dec. 2005.

[7] P. Pöyhönen et al., *DEEP - A Generic Name Resolution Protocol for Heterogeneous Networks*, Proc. 2nd IEEE International Conference on Information & Communication Technologies: from Theory to Applications, Damascus, Syria, April 2006.

[8] R. Braden, Ed., et al., *Resource ReSerVation Protocol (RSVP) – Version 1 Functional Specification*, RFC 2205, Sep. 1997.

[9] H. Schulzrinne et al., *Design of CASP – a Technology Independent Lightweight Signaling Protocol*, Proc. International Workshop on Inter-domain Performance and Simulation, Salzburg, Austria, Feb. 2003.

[10] R. Hancock et al., J. Loughney and S. van den Bosch, *Next Steps in Signalling (NSIS): Framework*, RFC 4080, June 2005.

[11] H. Schulzrinne and R. Hancock, *GIST: General Internet Signalling Transport*, draft-ietf-nsis-ntlp-09, Internet Draft (work in progress), Feb. 2006.

[12] J. Manner et al., *NSLP for Quality-of-Service Signalling*, draft-ietf-nsis-qos-nslp-10, Internet Draft (work in progress), March 2006.

[13] X. Fu et al., *Overhead and Performance Study of the General Internet Messaging Protocol for Signaling (GIMPS)*, to appear in Proc. 25th Conference on Computer Communications (INFOCOM’06), Barcelona, April 2006.

[14] NS-2 website, http://www.isi.edu/nsnam/ns/