

Review Article

Evolution of Signaling Information Transmission

Jasmina Baraković Husić,¹ Himzo Bajrić,² and Sabina Baraković³

¹ BH Telecom, Joint Stock Company, Sarajevo, Directorate BH Mobile, Obala Kulina bana 8, 71000 Sarajevo, Bosnia and Herzegovina

² BH Telecom, Joint Stock Company, Sarajevo, Executive Directorate for Technology and Service Development, Obala Kulina bana 8, 71000 Sarajevo, Trg BiH 1, 71000 Sarajevo, Bosnia and Herzegovina

³ Department for Informatics and Telecommunication Systems, Ministry of Security of Bosnia and Herzegovina, Trg BiH 1, 71000 Sarajevo, Bosnia and Herzegovina

Correspondence should be addressed to Jasmina Baraković Husić, jasmina.barakovic@bhtelecom.ba

Received 27 July 2012; Accepted 17 September 2012

Academic Editors: R.-I. Chang and J. Song

Copyright © 2012 Jasmina Baraković Husić et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Next Generation Network (NGN) faces the challenge of the rapidly increasing amount of signaling. The growing amount of signaling is a consequence of several reasons arising from the fact that signaling is the main source of network intelligence, analysis, and user behavior monitoring. With the increase in signaling load and complexity, the network management becomes a challenging issue that can impact overall Quality of Service (QoS). To confront this issue, there is a need for reliable and forehand signaling transmission in NGN. As there is much confusion about the interpretation of this concept, this paper aims to provide an overview of the evolution of signaling transmission. Migration towards NGN is analyzed from the signaling perspective. The NGN signaling protocols and related transmission requirements are identified. Through the discussion of standard approaches, the paper considers our previously published approach to signaling transmission along with the current issues and emerging opportunities.

1. Introduction

The rapid growth in subscribers, devices, and applications increases the signaling volume that is causing congestion and impacting the Quality of Service (QoS) [1]. In order to deliver the desired levels of service and user's Quality of Experience (QoE), it is necessary to process significant and growing volume of signaling in real time. There is the need to manage signaling updates and to ensure that service performance is maintained. With the migration to Next Generation Networks (NGNs), the need to manage the growth in signaling has become critical to optimize the network and ease congestion in real time [2].

As networks migrate toward all Internet Protocol (IP) based networks, the growing amount of intelligence is sent through the network [3]. Intelligence means more control over network resources. The optimal source of that intelligence is the signaling. Several advantages are offered using the signaling as the main source of network intelligence, for example, signaling contains the valuable and strategic information in the network that is available nowhere except in the NGN control plane. Therefore, with

the growing volume of signaling, the network management becomes the challenging issue. To confront this issue, there is a need for reliable and forehand signaling transmission in NGN. As there is much confusion about the interpretation of this concept, this paper aims to provide an overview of the evolution of approaches to signaling transmission.

The NGN is an IP-based packet-switched network, which does not rely on separate signaling network as in the case of circuit-switched networks [4]. Building such a signaling network is either not possible or it makes no sense. In traditional telecommunication networks, signaling is transported by the separate Signaling System Number 7 (SS7) network fulfilling a number of reliability and performance requirements. Although there is no separate signaling network in the NGN, signaling transport must have similar performance to SS7 [5]. As IP-based transport network is inherently unreliable and lacks the QoS guarantees, there is a need to provide differentiated treatment to signaling traffic. In this respect, several standards and recommendations were proposed by the International Telecommunication Union Telecommunication Standardization Sector (ITU-T),

the Internet Engineering Task Force (IETF), the Institute of Electrical and Electronics Engineers (IEEE), and 3rd Generation Partnership Project (3GPP) [6]. In addition to summarizing the approaches of these standards bodies, this paper aims to discuss our approach [7] emerged from the need for reliable and forehand signaling transmission.

The paper is organized as follows. Section 2 analyzes migration towards NGN from the signaling perspective. The NGN signaling protocols and related transmission requirements are identified in Section 3. Through the discussion of standard approaches, Section 4 describes our previously proposed approach to signaling transmission along with the current issues and emerging opportunities. Section 5 gives concluding remarks and identifies open issues for future work.

2. Signaling Perspective on NGN Migration

The migration to NGN has shown to be an evolutionary rather than a revolutionary transition, which has generated a significant impact on the signaling [3]. As signaling is a starting point of the network's evolution, this section reviews its origins and purpose along with the transmission requirements. Signaling has posed many challenges as its history has been linked to the history of telecommunication and switching [8]. Signaling, as it is known today, began around 1890, with the invention of the automatic switching exchange. The period from 1890 to 1976 was characterized by the use of Channel Associated Signaling (CAS), where the speech (in-band), or a channel closely associated with a speech channel (out-band), is used for signaling. Alternative form of signaling was introduced in 1976, that is, Common Channel Signaling (CCS), where a dedicated channel, completely separate from speech channel, is used for signaling. CCS has been used to implement SS7.

SS7 constitutes a separate network within a telecommunication network. SS7 network includes a number of different types of Signaling Points (SPs). In fact, there can be three different types of SPs in an SS7 network: Service Switching Point (SSP), Service Control Point (SCP), and Signal Transfer Point (STP). SSPs provide the SS7 functionality of a switch. SCPs interface the SS7 network to query telecommunication databases, allowing service logic and additional routing information to be obtained to execute services. STPs are used to transfer signaling messages. In SS7, control messages are used by the relevant SPs for call management purposes [8].

Besides it is the signaling network, SS7 is a protocol suite that provides the mechanisms for exchanging control messages using the packet-switching facilities. The Message Transfer Part (MTP) and the Signaling Connection Control Part (SCCP) provide the transfer protocols. MTP is used to reliably transport messages between nodes, and SCCP is used for noncircuit-related signaling. MTP provides the means of reliable transport and delivery of User Parts (UPs)/Application Parts (APs) information across the SS7 network [8].

In addition to transfer of control messages, the signaling network has to meet a number of reliability and performance

requirements [5]: (1) probability of message delivered with undetected errors less than 10^{-10} ; (2) message loss probability less than 10^{-7} ; (3) signaling network unavailability less than 10 minutes per year; (4) probability of message delivered out-of-sequence (including message duplication) less than 10^{-10} ; (5) message transfer delay in the STPs less than 100 ms. To fulfill these strict requirements, a reliable transfer of messages between any two SPs is offered by the MTP routing functions and network redundancy. The quality of the underlying physical transport and the probability of system internal errors affect the fulfillment of the first requirement. The redundant signaling links and change-over procedure that allows the loss-free switching of traffic from a failed links to other links are deployed by the MTP to fulfill the second requirement. Several procedures supporting redundancy in the signaling network enable the fulfillment of the third requirement. The MTP performs explicitly timer-based sequence control procedures when rerouting traffic via alternative routes or reverting traffic back to the original routes in order to meet the fourth requirement. Several mechanisms to limit outgoing queues and overall signaling transfer delay are provided by MTP to fulfill the fifth requirement.

When SS7 was formalized, direct signaling connections were utilized to transport control messages. In this case, resources were dedicated to the transmission of these messages. In all forms of telecommunications, certain higher layer protocols are required to perform certain tasks. These higher layers had to have a common transport protocol to deliver them efficiently and reliably. In the beginning the Time-Division Multiplexing (TDM) network provided a 64 kb/s signaling link between two elements that could transfer on a link-to-link basis. This is traditionally known as narrowband SS7. Then when the core network model was introduced, Asynchronous Transfer Mode (ATM) was considered to be the common transport mechanism. The use of ATM to transfer SS7 messages is known as broadband SS7. As IP networks become more extensive, the need to utilize the resources becomes more apparent. This is the way the concept of SS7 over IP is introduced, where existing SS7 protocols can be merged into the same common transport mechanisms as speech and data traffic.

In the creation of NGN it is necessary to transfer SS7 information over IP network. This is why Signaling Transport (SIGTRAN) working group of the IETF was created. The SIGTRAN was tasked with defining the architecture for transporting real-time signaling information over an IP network [9]. The working group's effort yielded three key results: (1) new architecture framework centered on a restructuring of the circuit switch into three functional entities that are Media Gateway Controllers (MGCs), Media Gateways (MGs), and Signaling Gateways (SGs); (2) Stream Control Transmission Protocol (SCTP) as a new transport protocol suitable for meeting the requirements of carrying telecommunication protocols, especially SS7, over a packet network; (3) numerous adaptation layers that support the primitives of Switched Circuit Network (SCN) telephony signaling protocols, such as MTP Level 3 User Adaptation (M3UA), SCCP User Adaptation (SUA), MTP Level 2 User

Adaptation (M2UA), MTP Level 2 Peer Adaptation (M2PA), and Integrated Services Digital Network (ISDN) User Adaptation (IUA). Both SCTP and the adaptation layers include mechanisms designed to handle redundant configurations and improve availability required by signaling.

Deploying SIGTRAN brings the IP infrastructure into the signaling network, allowing the interworking between NGN and circuit-switched networks. The SIGTRAN traffic increases the total signaling load in IP network. The estimation of SIGTRAN traffic volume is useful, if it is going to receive preferential treatment in IP network. Both Differentiated Services (DiffServ) and Multiprotocol Label Switching (MPLS) Traffic Engineering (TE) are used for this purpose [10]. These mechanisms provide better QoS to signaling traffic transported over IP network. Additionally, fast rerouting capability of MPLS helps to reduce the number of network failures that have to be recovered by SCTP and M3UA. Since the QoS and reliability are two fundamental requirements for signaling network, DiffServ and MPLS-TE are mechanisms that complement SIGTRAN in order to meet such requirements. Therefore, the necessity of using them should be assessed as part of the overall NGN design.

A challenging issue with the design of NGN architecture is the implementation of a suitable signaling and session control layer that has proven its importance in the SS7 signaling network. The appropriate signaling and session control layer need to be implemented in the NGN to offer highly available signaling with the reliability and scalability of SS7 network. These requirements present an opportunity to introduce IP Multimedia Subsystem (IMS) technology in the NGN requiring the interoperability and integration of different technologies and protocols as the networks evolve [3].

3. Signaling Protocols in NGN

The NGN is being developed using a number of different technologies. The NGN architecture recognizes the differences between the technologies that can be employed in the access part of the network from those used in the core part of the network. The core network consists of four major candidate transport technologies, that is, ATM, Ethernet, IP, and IP/MPLS. The access networks uses various wireless and wireline access technologies, such as Universal Mobile Telecommunications System (UMTS), Long-Term Evolution (LTE), Worldwide Interoperability for Microwave Access (WiMAX), Ultrawideband (UWB), Wireless Local Area Network (WLAN), Wireless Personal Area Network (WPAN), Bluetooth, Ethernet cable, Digital Subscriber Line (DSL), and optical fiber, to provide consistent and ubiquitous services to subscribers [11].

To implement several technologies in a unified infrastructure, there is a need for standards development to optimize the migration to the NGN architecture. Standardization efforts led by the ITU-T and other standards development organizations have generated a consensus on the basic NGN architecture model and services. The NGN architecture is described with a set of Functional Entities (FEs). Each FE is

located in the transport or the service stratum of the NGN architecture. Transport stratum provides user functions for data transfer as well as functions that control and manage transport resources to carry such data among communicating entities. Service stratum provides user functions for service-related data transfer and functions that control and manage service resources and network services to enable user's services and applications [12].

Transport stratum includes Transport Functions and Transport Control Functions. Transport Functions include access network functions, edge functions, core transport functions, gateway functions, and media-handling functions. Transport Control Functions include three groups of functions: Network Attachment and Control Functions (NACF), Resource and Admission Control Functions (RACF), and Mobility Management and Control Functions (MMCF). Service stratum includes Service Control and Content Delivery Functions and Application Support Functions and Service Support Functions [13]. Although the NGN functional blocks have been described in detail, their relationship to the actual protocols used to support these functions have not been described so far. This is the result of assumption that NGN will be based on existing transport-, control-, and management-related technologies. Therefore, there is the need to define the relationship of these protocols to the overall NGN architecture.

Although each NGN stratum is composed of three planes, that is, user plane, control plane, and management plane [12], this paper is focused only on the control plane that includes both routing and signaling protocols, and the associated interfaces. As they constitute a separate and distinct category of protocols, there is a need to clarify this separation in NGN architecture in order to facilitate the development of the interoperable protocols and interfaces. A number of control plane protocols are allowed by the NGN architecture. The control plane protocols depend on whether the connectionless or connection-oriented categories of the NGN are to be implemented. This section considers both routing and signaling protocols for connectionless categories of NGN as shown in Figure 1.

3.1. Transport Stratum Control Plane Protocols. The transport stratum control plane is responsible for routing information exchange and label distribution between adjacent devices. Transport Functions use standard routing protocols such as Routing Information Protocol (RIP), Open Shortest Path First (OSPF), Intermediate System to Intermediate System (IS-IS), and Border Gateway Protocol (BGP) to exchange information in order to build IP forwarding table or label forwarding information base in MPLS environment. Labels are distributed by means of Label Distribution Protocol (LDP), as well as Constraint-based Routing Label Distribution Protocol (CR-LDP) that has been deprecated because IETF decided to focus purely on Resource Reservation Protocol-Traffic Engineering (RSVP-TE). Routing protocols are part of control plane regardless of their transport mechanism; for example, RIP runs over User Datagram Protocol (UDP), OSPF runs directly over IP, BGP runs over Transmission

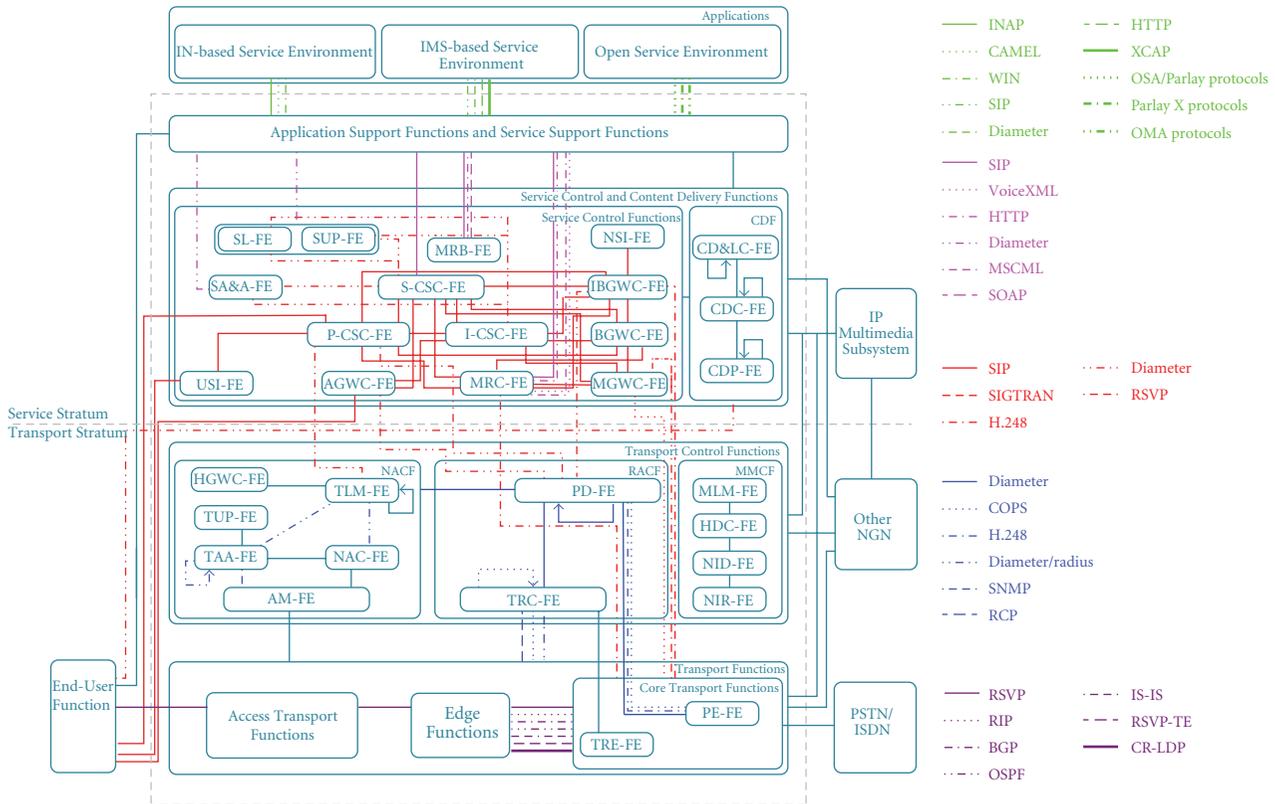


FIGURE 1: NGN control-plane protocols. AGWC-FE (Access Gateway Control FE); AM-FE (Access Management FE); BGP (Border Gateway Protocol); BGWC-FE (Breakout Gateway Control FE); CAMEL (Customized Application for Mobile network Enhanced Logic); CDC-FE (Content Delivery Control FE); CDF (Content Delivery Function); CD&LC-FE (Content Distribution and Location Control FE); CDP-FE (Content Delivery Processing FE); COPS (Common Open Policy Service); CR-LDP (Constraint-based Routing Label Distribution Protocol); FE (Functional Entity); HDC-FE (Handover Decision and Control FE); HGWC-FE (Home Gateway Configuration FE); HTTP (Hypertext Transfer Protocol); IBGWC-FE (Interrogating Breakout Gateway Control FE); I-CSC-FE (Interrogating Call Session Control FE); INAP (IN Application Protocol); IP (Internet Protocol); ISDN (Integrated Services Digital Network); IS-IS (Intermediate System to Intermediate System); MGWC-FE (Media Gateway Control Function FE); MLM-FE (Mobile Location Management FE); MMCF (Mobility Management Control Function); MRB-FE (Media Resource Broker FE); MRC-FE (Media Resource Control FE); NACF (Network Attachment and Control Functions); MSCML (Media Server Control Markup Language); NAC-FE (Network Access Configuration FE); NGN (Next Generation Network); NID-FE (Network Information Distribution FE); NIR-FE (Network Information Repository FE); NSI-FE (Network Signaling Interworking FE); OSPF (Open Shortest Path First); P-CSC-FE (Proxy Call Session Control FE); PD-FE (Policy Decision FE); PE-FE (Policy Enforcement FE); PSTN (Public Switched Telephone Network); RACF (Resource and Admission Control Function); RCP (Resource Control Protocol); RIP (Routing Information Protocol); RSVP-TE (Resource Reservation Protocol-Traffic Engineering); SA&A-FE (Service Authentication and Authorization FE); S-CSC-FE (Serving Call Session Control FE); SIGTRAN (Signaling Transport); SIP (Session Initiation Protocol); SL-FE (Subscription Locator FE); SNMP (Simple Network Management Protocol); SOAP (Simple Object Access Protocol); SUP-FE (Service User Profile FE); TAA-FE (Transport Authentication and Authorization FE); TLM-FE (Transport Location Management FE); TRE-FE (Transport Resource Enforcement FE); TRC-FE (Transport Resource Control FE); TUP-FE (Terminal User Profile FE); USI-FE (User Signaling Interworking FE); XCAP (XML Configuration Access Protocol); XML (eXtensible Markup Language); WIN (Wireless Intelligent Network).

Control Protocol (TCP), and so forth. Routing traffic is important as it keeps the network operational, and it needs to be forwarded in a timely manner. A minimum bandwidth needs to be guaranteed to ensure that routing traffic always receives timely service and probability of packet drop under peak load is very low [14].

Transport Control Functions are responsible for the admission decision and the resource control of the transport function. The RACF comprises Policy Decision Functional Entity (PD-FE) that takes the final decision over the resource and admission control and delivers it to the corresponding Policy Enforcement Functional Entity (PE-FE)

through Common Open Policy Service Policy Provisioning (COPS-PR), Media Gateway Control (MEGACO)/H.248, or Diameter based interface. Derived from its predecessor Remote Authentication Dial-In User Service (RADIUS), Diameter is used by the RACF to interact with NACF, including network access registration, authentication and authorization, and parameter configuration for checking user profiles and Service Level Agreements (SLAs) held by them. Most of these protocols require reliable transport mechanism, for example, Diameter runs over TCP or SCTP, COPS-PR runs over TCP, MEGACO/H.248 runs over UDP, TCP, or SCTP, and so forth. Applications that use these

protocols require a low packet loss, but are relatively not sensitive to delay. Therefore, it is necessary to ensure sufficient bandwidth in the network to provide high assurance of delivery [14].

3.2. Service Stratum Control Plane Protocols. The service stratum control plane provides a variety of functions that control and manage service resources and network services to enable user services and applications. Service Control Functions include resource control, registration, authentication and authorization functions at the service level, and functions for controlling media resources [15]. The actual protocols used to support these functions are Diameter, MEGACO/H.248, and Session Initiation Protocol (SIP). The use of SIP has received substantial attention since it has been selected as the basis of IMS specifications developed by 3GPP and adopted as an integral part of the overall NGN architecture. Many signaling protocols came before SIP, such as ITU-T H.323 umbrella protocol [16]. The Voice over IP (VoIP) signaling protocols evolution is reviewed in [17].

Content Delivery Functions (CDFs) receive content from the Application Support Functions and Service Support Functions and then store, process, and convey this content to the End-User Functions using the means of the Transport Functions. Real-Time Streaming Protocol (RTSP) is intended to control content delivery sessions and provide a means for choosing delivery mechanisms. RTSP can use UDP or TCP as a transport protocol.

Application Support Functions and Service Support Functions include the registration, authentication, and authorization functions at the application level offered to the Applications and the End-User Functions. Cooperation between the Application Support Functions and Service Support Functions and the Service Control Functions offers the requested NGN services to the Applications and the End-User Functions. This cooperation is realized using SIP, Hypertext Transfer Protocol (HTTP), Simple Object Access Protocol (SOAP), Diameter, Voice eXtensible Markup Language (VoiceXML), and Media Server Control Markup Language (MSCML) [15]. SIP is designed to be independent of the underlying transport protocol. Although HTTP definitions presume reliable transport protocol, such as TCP, it has found application even with unreliable protocols, such as UDP. SOAP, as the envelope syntax for sending and receiving XML messages with the Web services can be used over any transport protocol such as HTTP, Simple Mail Transfer Protocol (SMTP), or even TCP. Additionally, HTTP transport is used for VoiceXML, and SIP transport is used for MSCML. This kind of traffic requires sufficient bandwidth in the network to provide high assurance of delivery and low delay. As it does not respond dynamically to packet loss, probability of packet drop or queuing delay under peak load should be very low for this type of traffic [14].

In order to facilitate the implementation of value-added services, NGN capabilities and resources are offered to applications through the use of a standard Application Network Interface (ANI). NGN supports the following three classes of value-added service environments [18]: (1) IN-based service environment (examples of ANI-specific protocols include IN

Application Protocol (INAP), Customized Application for Mobile network Enhanced Logic (CAMEL), and Wireless Intelligent Network (WIN)); (2) IMS-based service environment (examples of ANI-specific protocols include SIP, Diameter, HTTP, and XML Configuration Access Protocol (XCAP)); (3) open service environment (examples of this environment using ANI include Open Service Access (OSA)/Parlay, Parlay X, and Open Mobile Alliance (OMA)). In regard to open interfaces, there are ongoing efforts to provide standardized Application Programming Interfaces (APIs) for RESTful Web services [19], which meets many NGN service provisioning requirements.

4. Signaling Transmission in NGN

The previous section provides an overview of the main signaling protocols in use in current and next generation networks. All of them have the common objective of providing some form of control over and support of user traffic and services, thus contributing to good operation of the network.

Signaling messages can be transmitted using path-coupled or path-decoupled approach. In case of path-coupled approach, signaling nodes must be collocated with routers and signaling messages are routed only through the nodes that are on the data path. In case of path-decoupled approach, signaling nodes should be dedicated and separate from routers and signaling messages are routed through nodes that are not assumed to be on the data path. The RSVP is path coupled and the SIP is path decoupled.

From QoS aspect, signaling can be classified in another two categories: in-band and out-of-band. In-band QoS signaling assumes that signaling traffic is a part of the associated data traffic. It is typically presented in a particular header field of the IP packets, for example, Type of Service (ToS), DiffServ Code Point (DSCP), Traffic Class, and so forth. It neither introduces additional traffic into the network nor incurs setup delay for the data traffic. By definition in-band signaling is path-coupled signaling. Out-of-band QoS signaling assumes that signaling traffic is carried by dedicated packets, separate from the associated data traffic. It entails the use of a signaling protocol and further processing above the network layer. Depending on whether the signaling path is closely tied to the associated data path, it is path-coupled or path-decoupled signaling.

Out-of-band signaling is used in Integrated Services (IntServ), while in-band signaling is used in DiffServ, which are both class-based QoS models proposed by IETF. IntServ is not suitable for signaling transmission due to issues raised by the resource reservation for every flow in intermediate routers across the end-to-end path. This introduces additional delay caused by signaling path establishment process, which is unacceptable due to strict time constraints posed by signaling. DiffServ avoids this processing complexity and signaling overhead in network routers. It concentrates on aggregated flows and Per Hop Behaviors (PHBs) applied to a network-wide set of traffic classes.

The idea of QoS classification appears in a few more standards and recommendations. However, the most important

TABLE 1: ITU-T QoS classes [6].

QoS ^f class	Application (examples)	IPTD ^e	IPDV ^b	IPLR ^d	IPER ^c
Class 0	Real-time, jitter sensitive, high interaction (VoIP ^g , VTC ^h)	100 ms	50 ms	1×10^{-3}	1×10^{-4}
Class 1	Real-time, jitter sensitive, interactive (VoIP ^g , VTC ^h)	400 ms	50 ms	1×10^{-3}	1×10^{-4}
Class 2	Transaction data, highly interactive (signaling)	100 ms	U ⁱ	1×10^{-3}	1×10^{-4}
Class 3	Transaction data, interactive	400 ms	U ⁱ	1×10^{-3}	1×10^{-4}
Class 4	Low loss only (short transactions, bulk data, video streaming)	1 s	U ⁱ	1×10^{-3}	1×10^{-4}
Class 6	Traditional applications of default IP ^a networks	U ⁱ	U ⁱ	U ⁱ	U ⁱ

^aIP: Internet Protocol; ^bIPDV: IP Packet Delay Variation; ^cIPER: IP Packet Error Ratio; ^dIPLR: IP Packet Loss Ratio; ^eIPTD: IP Packet Transfer Delay; ^fQoS: Quality of Service; ^gVoIP: Voice over IP; ^hVTC: Video Telepresence; ⁱU: Unspecified.

QoS classes are defined by ITU-T, IETF, 3GPP, and IEEE [6]. The QoS class is term that is often interchangeably used in the meaning of service class. Services belonging to the same class are described by a specific set of parameters, which can be expressed qualitatively or quantitatively.

ITU-T Recommendation Y.1541 introduced five service classes. There are four QoS parameters to be guaranteed for each service class, that is, IP Packet Transfer Delay (IPTD), IP Packet Delay Variation (IPDV), IP Packet Loss Ratio (IPLR), and IP Packet Error Ratio (IPER). The defined bounds for these QoS parameters are shown in Table 1 and should not be exceeded.

IETF standardized two QoS models, as mentioned above. Within IntServ model, three service classes are defined, that is, Controlled-Load, Guaranteed QoS, and best-effort service classes. DiffServ model introduced a number of PHBs that can be used for realization of various service classes. IETF defined twelve classes in Request for Comments (RFC) 4594, which are listed in Table 2. It recommends how to construct these service classes using DiffServ mechanisms. Although the requirements of each service class are described in terms of a tolerance to packet loss, delay, and jitter, strict values or bounds for QoS parameters are not provided. As defined by RFC 5127, these service classes are aggregated into four QoS classes, that is, Network Control, Real-Time, Assured Elastic, and Elastic Treatment Aggregate.

3GPP specification TS 23.107 defined four QoS classes in Universal Mobile Telecommunications System (UMTS) networks, that is, Conversational, Streaming, Interactive, and Background. Additionally, some features of each QoS class are distinguished in terms of fundamental characteristics, transfer delay, transfer delay variation, low bit error rate, guaranteed bit rate, buffering, and so forth. Fundamental characteristics of each QoS class are shown in Table 3.

IEEE 802.1d standard proposed eight traffic types to be supported in Local Area Network (LAN). The traffic types are mapped to User Priorities (UPs) that are used to differentiate QoS. Requirements for QoS parameter values on traffic types and UPs are not imposed except for voice and video traffic types. For wireless networks, four Access Categories (ACs) are defined to differentiate services in IEEE 802.11e. Mappings between IEEE 802.1d UPs and IEEE 802.11e ACs are shown in Table 4.

Standardization bodies make an effort to define mapping between QoS classes to improve the cooperation between various techniques. It is shown in [6] that, despite some

TABLE 2: IETF QoS classes.

RFC ^f 5127	RFC ^f 4594
Ctrl (CS ^b)	Network control (CS ^b 6)
Real time (EF ^d)	Telephony (EF ^d) Signaling (CS ^b 5) Multimedia conferencing (AF ^a 4) Real-time interactive (CS ^b 4) Multimedia streaming (AF ^a 3)
Assured elastic (AF ^a)	Broadcast video (CS ^b 3) Low latency data (AF ^a 2) OAM ^e (CS ^b 2) High priority data (AF ^a 1)
Elastic (DF ^c)	Low priority data (CS ^b 1) Standard (DF ^c)

^aAF: Assured Forwarding; ^bCS: Class Selector; ^cDF: Default Forwarding; ^dEF: Expedited Forwarding; ^eOAM: Operations, Administration, and Management; ^fRFC: Request For Comments.

inconsistencies, it is possible to find mappings between ITU-T, IETF, IEEE and 3GPP QoS classes. This paper considers only on signaling-specific QoS classes, as shown in Table 5.

The focus is on the Signaling service class, as defined in the IETF RFC 4594. This service class is included in the Real-Time treatment aggregate, as proposed in IETF RFC 5127. It should be supported by the UMTS Interactive QoS class. Signaling service class is marked with DSCP value CS5. This DSCP value can be mapped to IEEE 802.1d UP 5, and IEEE 802.11e AC AC_VI. This service class is mapped to Class 2 that is specified in ITU-T Recommendation Y.1541. Therefore, it should be configured to assure the target values for IPTD and IPLR: the upper bound on IPTD below 100 ms and the upper bound on IPLR below 1×10^{-3} .

RFC 4594 proposed to configure Signaling service class to provide a minimum bandwidth assurance for CS5 marked packets. This service class should use a rate queuing system, such as Weighted Fair Queuing (WFQ) or Weighted Round Robin (WRR). The single rate with burst size token bucket policer should be used to ensure that the signaling traffic stays within its negotiated or engineered bounds. Traffic in this service class does not respond dynamically to packet loss, so Active Queue Management (AQM) should not be applied to CS5 marked packets [14]. Since RFC 4594 is

TABLE 3: UMTS QoS classes.

QoS ^a class	Fundamental characteristics	Example of application
Conversational	Preserve time relation (variation) between information entities of the stream Conversational pattern (stringent and low delay)	Voice
Streaming	Preserve time relation (variation) between information entities of the stream	Streaming video
Interactive	Request response pattern Preserve payload content	Web browsing
Background	Destination is not expecting the data within a certain time Preserve payload content	Background download of emails

^aQoS: Quality of Service.

TABLE 4: IEEE QoS classes [6].

Priority	802.1d			802.11e	
	UP ^b	Designation	Traffic type	AC ^a	Designation
Low	1	BK	Background	AC.BK	Background
	2	—	spare	AC.BK	Background
	0	BE	Best effort	AC.BE	Best effort
	3	EE	Excellent effort	AC.BE	Best effort
	4	CL	Controlled load	AC.VI	Video
	5	VI	Video	AC.VI	Video
High	6	VO	Voice	AC.VO	Voice
	7	NC	Network control	AC.VO	Voice

^aAC: Access Category; ^bUP: User Priority.

TABLE 5: Mapping of signaling-specific QoS classes.

Standard body	Signaling-specific QoS classes	
	Document	QoS class
ITU-T [§]	Y.1541	Class 2
IETF ^f	RFC 4594	Signaling (CS5) ^c
	RFC 5127	Real time (EF) ^d
IEEE ^e	802.1d	UP ^k 5
	802.11e	AC.VI ^b
3GPP ^a	TS 23.107	UMTS ^j /Interactive
	TS 23.203	LTE ^h /QCI ⁱ 5

^a3GPP: 3rd Generation Partnership Project; ^bAC.VI: Access Category_Video; ^cCS: Class Selector; ^dEF: Expedited Forwarding; ^eIEEE: Institute of Electrical and Electronics Engineers; ^fIETF: Internet Engineering Task Force; [§]ITU-T: International Telecommunication Union Telecommunication Standardization Sector; ^hLTE: Long-Term Evolution; ⁱQCI: QoS Class Identifiers; ^jUMTS: Universal Mobile Telecommunications System; ^kUP: User Priority.

to be viewed as industry best-practice recommendation, enterprises and service providers are encouraged to adopt this recommendation, with the aim of improving QoS consistency, compatibility, and interoperability. However, since it is a set of formal DiffServ QoS configuration best practices, and not a requisite standard, modifications can be made to these recommendations as required by specific needs and constraints. To meet the QoS requirements as defined in ITU-T Recommendation Y.1541, we proposed

a modification of these configuration guidelines with regard to Signaling service class [20].

The approach proposed in our previous work is based on configuring Signaling service class by using Priority Queuing (PQ) system to give it absolute preferential treatment over all other User service classes. A packet assigned to Signaling service class should be marked with a new DSCP value, which should be requested from the Internet Assigned Numbers Authority (IANA). This DSCP value should be lower than the one used to configure the Network Control service class and higher than the one reserved for all User service classes defined in RFC 4594 and RFC 5865.

Applying the proposed modification to the Signaling service class configuration, it is possible to improve the QoS performances of real-time services. Since this approach is signaling protocol independent, it is considered in one of our previous works with regard to Real-time Transport Control Protocol (RTCP). The standard approach regarding control information prioritization requires RTCP packets to be marked with the same DSCP as the Real-time Transport Protocol (RTP) packets. Our approach is based on marking the RTCP packets with the higher DSCP than the one used by RTP packets. In other words, RTCP packets are classified into the Signaling service class that is assigned the highest priority level in order to transmit user control information reliably and efficiently and thereby increase QoS. In this regard, Figure 2 shows the impact of control information prioritization on the RTP throughput. Both approaches provide the same throughput when the network is unloaded.

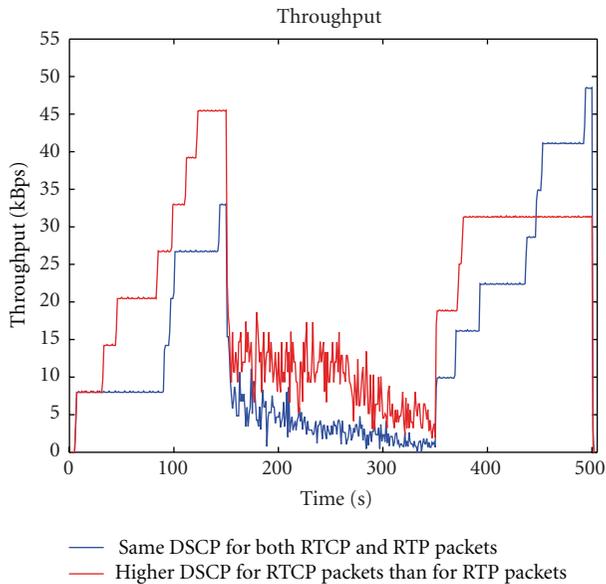


FIGURE 2: Impact of prioritizing of signaling/control information transmission on RTP throughput [21].

Our approach provides higher throughput than the standard approach when the network is heavily loaded. This is because our approach lets RTCP packets adjust the sending rate of RTP packets more dynamically than standard approach and in dependence upon the status of network. Conversely, the standard approach may allow RTCP packets to establish the improper sending rate resulting in higher throughput because QoS statistics determined by delayed RTCP packets may be different than the actual QoS experienced by RTP packets [21].

Signaling service class should be assigned with the highest priority level in order to ensure reliable and forehand signaling transmission. It has been recognized that reliable and forehand signaling transmission is the main condition to guarantee service reachability and enable dynamic end-to-end QoS control depending on the availability of network resources [7].

This concept has also been recognized by 3GPP and incorporated in LTE QoS framework [22]. The Evolved Packet System (EPS) uses QoS Class Identifiers (QCI), which specifies the class to which the bearer belongs. The EPS bearer carrying IMS signaling has a QCI value of 5, which characterizes the highest-priority nonguaranteed bit rate bearer suitable for IMS control messages. Since the bearer is not visible in the transport network, a QCI to DSCP mapping function is implemented. The purpose of this function is to make a translation from bearer-level QoS (QCI) to transport level QoS (DSCP). Using this function, signaling packets on a bearer associated with the highest priority QCI should be marked with the highest priority DSCP for forwarding in the transport network. As mentioned before, the traffic forwarding treatment is determined by queue management schemes and scheduling algorithms for uplink and downlink traffic. For downlink packets, the gateway performs this mapping, while the LTE Radio Access Network (RAN) performs it for uplink packets.

Signaling service class should be used by a number of different applications, such as peer-to-peer IP telephony signaling (e.g., using SIP, H.323), peer-to-peer signaling for multimedia applications, client-server IP telephony signaling (e.g., using H.248, MEGACO, MGCP, etc.), signaling to control IP Television (IPTV) applications (e.g., using Internet Group Management Protocol (IGMP)), and so forth. Different applications have different characteristics and latency requirements. However, in current transport networks, all signaling messages receive uniform treatment regardless of any special application requirements because they are associated with the same DSCP value in their IP packet headers. The transport network uses different DSCP values to apply different forwarding treatment to Signaling service class and all other service classes. The same mechanism could be applied signaling messages having different priorities which are sent over the network. In this case, signaling messages are associated to different DSCP values relating to their priorities. Being linked to various research areas, this idea is hopefully going to become a new starting point for research activities in the future.

5. Conclusion and Future Work

As networks migrate toward NGN, the growing amount of intelligence is sent through network. The main source of that intelligence is the signaling. With the growing amount of signaling, network management becomes the challenging issue that impacts overall QoS. To overcome this issue, there is a need for reliable and forehand signaling transmission. To clarify the interpretation of this concept, this paper provides the overview of evolutionary transition to NGN from signaling perspective.

Since the signaling is a starting point of the network's evolution, its origins and purpose are reviewed along with the transport requirements. Signaling has proven its importance in the SS7, which is separate signaling network fulfilling a number of reliability and performance requirements. Although there is no separate signaling network in the NGN, signaling transport must have similar performance to SS7. Since the NGN transport network is unreliable and lacks the QoS guarantees, there is a need to provide differentiated treatment to signaling traffic. Both routing and signaling protocols in use in current and next generation networks are identified along with their relationship to the overall NGN in order to understand their transmission requirements.

Through the discussion of different standard approaches, this paper finds the mappings between ITU-T, IETF, IEEE, and 3GPP signaling-specific QoS classes. The focus is put on the Signaling service class defined by IETF, which is used as the basis for our approach to signaling transmission. This service class should be assigned with the highest priority level in order to ensure reliable and forehand signaling transmission. This concept has been recognized by 3GPP and incorporated in LTE QoS framework.

Our previously proposed approach could be further improved, if signaling messages having different priorities are marked with different DSCP values to apply different forwarding treatment when transported over the network.

Since the significant growth in Machine-to-Machine (M2M) communications is generating the enormous increases in signaling traffic, this approach could help to meet each M2M's application requirements while protecting the network and making more efficient use of available resources. This idea is hopefully going to become a new starting point for our future research activities.

References

- [1] J. M. Batalla, J. Sliwinski, H. Tarasiuk, and W. Burakowski, "Impact of signaling system performance on QoE in next generation network," *Journal of Telecommunications and Information Technology*, pp. 117–130, 2010.
- [2] T. Magedanz, M. Corici, and H. Coskun, "Telecom signaling evolution for IMS, OTT and M2M services," in *Proceedings of the 13th Annual NGN Signaling Evolution Forum*, Berlin, Germany, 2012.
- [3] E. Mikoczy, I. Kotuliak, and O. van Deventer, "Evolution of the converged NGN service platforms towards future networks," *Future Internet*, vol. 3, no. 1, pp. 67–86.
- [4] B. Orimoloye and H. Hanrahan, "Formalizing the next generation "signaling network"," in *Proceedings of the Southern Africa Telecommunication Networks and Application Conference (SATNAC '04)*, Western Cape, South Africa, 2004.
- [5] D. K. Grandischnig and M. Tuxen, "Signaling transport over IP-based networks using IETF standards," in *Proceedings of the 3rd International Workshop on the Design of Reliable Communication Network (DRCN '01)*, pp. 168–174, Budapest, Hungary, 2001.
- [6] R. Stankiewicz, P. Cholda, and A. Jajszczyk, "QoX: what is it really?" *IEEE Communications Magazine*, vol. 49, no. 4, pp. 148–158, 2011.
- [7] J. Baraković, H. Bajrić, M. Kos, S. Baraković, and A. Husić, "Prioritizing signaling information transmission in next generation network," *Journal of Computer Networks and Communications*, vol. 2011, Article ID 470264, 11 pages, 2011.
- [8] J. G. Van Bosse and F. U. Devetak, *Signaling in Telecommunication Networks*, John Wiley & Sons, Hoboken, NJ, USA, 2nd edition, 2007.
- [9] E. Vazquez, M. Alvarez-Campana, A. Hernandez, and J. Vinyes, "Reliable signalling transport in next generation networks," in *Proceedings of the 6th International Conference on Wired/Wireless Internet Communications (WWIC '08)*, pp. 53–66, Tampere, Finland, 2008.
- [10] J. Baraković, H. Bajrić, and A. Husić, "QoS design issues and traffic engineering in next generation IP/MPLS network," in *Proceedings of the 9th International Conference on Telecommunications (ConTEL '07)*, pp. 203–209, June 2007.
- [11] J. Zhang and N. Ansari, "On assuring end-to-end QoE in next generation networks: challenges and a possible solution," *IEEE Communications Magazine*, vol. 49, no. 7, pp. 185–191, 2011.
- [12] *General principles and general reference model for next generation networks*, ITU-T Recommendation Y.2011, 2004.
- [13] *Functional requirements and architecture of next generation networks*, ITU-T Recommendation Y.2012, 2010.
- [14] J. Bariariz, K. Chan, and F. Baker, "Configuration guidelines for DiffServ service classes," Tech. Rep. RFC 4594, IETF, 2006.
- [15] *Signalling architecture for the NGN service control plane*, ITU-T Recommendation Q.3030, 2008.
- [16] H. Liu and P. Mouchtaris, "Voice over IP signaling: H.323 and beyond," *IEEE Communications Magazine*, vol. 38, no. 10, pp. 142–148, 2000.
- [17] B. Khasnasish, *Implementing Voice over IP*, vol. 2003, John Wiley & Sons, Hoboken, NJ, USA.
- [18] ITU-T NGN FG Proceedings, Part II. Working Group 1 Deliverables—Service Requirements, 2005.
- [19] F. Belqasmi, R. Glitho, and C. Fu, "RESTful web services for service provisioning in next-generation networks: a survey," *IEEE Communication Magazine*, vol. 4912, pp. 66–73, 2011.
- [20] J. Baraković, H. Bajrić, and S. Baraković, "Priority level configuration for signaling service class," in *Proceedings of the 3rd International Conference on Communication Theory, Reliability, and Quality of Service*, pp. 122–127, Athens, Greece, June 2010.
- [21] J. Baraković, S. Baraković, and H. Bajrić, "The impact of control information prioritization on QoS performance metrics," in *Proceedings of the 4th International Conference on Communication Theory, Reliability, and Quality of Service (CTRQ '11)*, pp. 60–65, Budapest, Hungary, 2011.
- [22] M. Alasti, B. Neekzad, J. Hui, and R. Vannithamby, "Quality of service in WiMAX and LTE networks," *IEEE Communications Magazine*, vol. 48, no. 5, pp. 104–111, 2010.

