IEICE TRANS. COMMUN., VOL.E90–B, NO.5 MAY 2007

INVITED PAPER *Special Section on Advanced Transfer Technologies for the Next Generation Network*

Introduction to the Functional Architecture of NGN

Naotaka MORITA[†a)] ***and*** **Hideo IMANAKA**[†]**,** ***Members***

SUMMARY In July 2006, International Telecommunication Union-Telecommunication Standardization Sector (ITU-T) Study Group 13 initiated the approval process for a batch of framework Recommendations on the Next Generation Network (NGN) Release 1. One of the new Recommendations, Y.2012, illustrates the NGN from the viewpoint of a functional architecture consisting of various functional blocks, namely functional entities. In conjunction with this Recommendation, this paper explains how the NGN can be built and how the NGN utilizes functional entities to provide expected services and required capabilities. This paper also identifies open issues for extending the functional architecture towards Release 2.

key words: *NGN, IMS, functional architecture, stratum, emulation, simulation*

1. Introduction

The Next Generation Network (NGN), which has been excessively used as a commercial catch phrase for any new network technology, is now seen by major network operators and service providers as a key infrastructure. It is expected to replace the aging telephone networks in an economical manner. It is also expected to provide a new service platform that will enable operators to offer converged services for the fixed and mobile communication businesses and bridge telecommunication and broadcasting businesses. This should create a new source of revenue to offset the current decline in revenues. NGN study accelerated in 2003, driven by major carriers in Europe, and there have recently been several aggressive public announcements from major carriers who are making commitments to introduce NGN within a few years [1], [2].

When it comes to the renewal of an entire network, which implies entering the next generation, standards are very important. For manufacturers, interoperability between their products and those of other vendors is important because products that comply with the standard will be easily adopted by operators and end users. This is accelerating market competition, which will lead to less expensive equipment being provided. For operators, interconnectability with other operators and other business partners is important in expanding business opportunities; there is no need to establish unique systems for each partner. For regulators, identifying the roles of operators is important to ensure that the arrangement of service provisioning is correct. Finally, for end customers, flexible selection of terminals and operators is important because it enables them to choose the most attractive services without being locked to specific vendors or operators.

International Telecommunication Union - Telecommunication Standardization Sector (ITU-T) answered the demand for new standards by establishing a timed focus group for NGN (FGNGN) in 2004. Its mission was to formulate a series of fundamental specifications by the end of 2005. The series is expected to contain the coverage of the first set of the release (i.e., Release 1), expected services, network capabilities, and functional architectures that clearly characterize the NGN. Study group 13 (SG13), its parent SG, encouraged FGNGN to accelerate their work by collecting more participants by relaxing the operating procedures so that it became more flexible than a rigid study group system. Based on the significant output of FGNGN, SG13 examined the major fundamental documents carefully. After the NGN Global Standards Initiative (NGN GSI) meeting in Kobe, Japan, in April 2006, SG13 finally initiated the approval process at its July 2006 meeting. Thirteen Recommendations including two supplementary documents, are shown in Table 1, some of which have already been officially approved.

Before FGNGN was established, ITU-T Recommendations Y.2001 [3] and Y.2011 [4] specified a general reference model that assumed decoupling of services and transport. In accordance with this decoupling principle, FGNGN and subsequent SG13 activities further defined the NGN architecture in terms of multiple functional groups. For session-based services, one of the key approaches to implementation is to use the IP Multimedia Subsystem (IMS) [5]. IMS was examined to ensure that its features meet both fixed and mobile network requirements. Another key component in the NGN is Resource and Admission Control Functions (RACF) [6], which provide end-to-end quality of service (QoS). Along with these key components, the generic functional architecture, described in Y.2012 (code name: Y.NGN-FRA) [7], shows the overall structure of the NGN and gives clear guidelines for designing the associated signaling protocols.

By reviewing the NGN structure in detail, this paper describes what NGN looks like and how it provides services and capabilities. It begins by describing the target NGN services, they focus on session-based telephony and multimedia communication. It then turns to the high-level architecture consisting of several functional entities. These include

Manuscript received September 4, 2006.
Manuscript revised December 20, 2006.
[†]The authors are with NTT Corporation, Musashino-shi, 180-8585 Japan.
a) E-mail: morita.naotaka@lab.ntt.co.jp
DOI: 10.1093/ietcom/e90–b.5.1022

Table 1 NGN-related Recommendations.

Next Generation Networks	Ref.	Date	Note 1	Note 2	Brief description
Y.2000–Y.2099 - Frameworks and functional architecture models					
Y.2001	[3]	2004/12			General overview of NGN
Y.2011	[4]	2004/10			General principles and general reference model for Next Generation Networks
Y.2012	[7]	2006/09		Note 2	Functional requirements and architecture of the NGN
Y.2012 Supplement 1		2006/07		Note 2	Session/border control (S/BC) functions
Y.2013	[19]	2006/12			Converged Services Framework Functional Requirements and Architecture
Y.2021		2006/09		Note 2	IMS for Next Generation Networks
Y.2031		2006/09		Note 2	PSTN/ISDN emulation architecture
Y.2091		2007/03		Note 2	Terms and Definitions for Next Generation Networks
Y.2100–Y.2199 - Quality of Service and performance					
Y.2111	[6]	2006/09		Note 2	Resource and admission control functions in Next Generation Networks
Y.2171		2006/09		Note 2	Admission control priority levels in Next Generation Networks
Y.2200–Y.2249 - Service aspects: Service capabilities and service architecture					
Y.2201	[9]	2007/04	Note 1	Note 2	NGN Release 1 requirements
Y.2250–Y.2299 - Service aspects: Interoperability of services and networks in N GN					
Y.2261		2006/09		Note 2	PSTN/ISDN evolution to NGN
Y.2262		2006/12			PSTN/ISDN emulation and simulation
Y.2271		2006/09		Note 2	Call server based PSTN/ISDN emulation
Y.2300–Y.2399 - Numbering, naming and addressing					
Y.2400–Y.2499 - Network management					
Y.2401/M.3060	[10]	2006/03			Principles for the Management of the Next Generation Networks
Y.2500–Y.2599 - Network control architectures and protocols (SG11)					
Y.2600–Y.2699 - Future packet based networks					
Y.2601		2006/12			Fundamental characteristics and requirements of future packet based networks
Y.2611		2006/12			High level architecture of future packet based networks
Y.2700–Y.2799 - Security					
Y.2701		2007/04	Note 1	Note 2	Security Requirements for NGN Release 1
Y.2800–Y.2899 - Generalized mobility					
Y.2801/Q.1706		2006/11		Note 2	Mobility management requirements for NGN
Y.2900–Y.2999 - Carrier grade open environment					
Y.2901		2006/12			The carrier grade open environment reference model
Y.2902		2006/12			Carrier grade open environment components
Others					
Y.2000SerSup1	[8]	2006/07		Note 2	NGN release 1 scope

Note 1: Document is available in April 2007.
Note 2: Included in the thirteen documents completed in July 2006 meeting.

the session-related control functional entities, which provide roaming over the fixed network. These entities will support multiple application platforms that will provide a wide variety of services ranging from emulation of legacy Intelligent Network (IN) services to new multimedia converged applications. In the transport stratum, multiple gateway functions will be used to interwork with existing networks as well as to protect the NGN against malicious attack or unexpected behavior by customers or other networks. Typical interactions between the functional entities are briefly shown. The paper concludes with the new items required to extend the NGN architecture to support new capabilities toward Release 2.

2. What is the NGN?

The NGN currently referred to in ITU-T and other regional standard bodies is an IP-based carrier-grade network that is reliable enough to provide telephony services that have usually been available over the existing public switched telephone networks (PSTN). It should also be flexible enough to provide a wide variety of services, even unknown future ones, including fixed-mobile convergence, telecommunication and broadcast convergence, and private and enterprise communication convergence. ITU-T started its NGN study by defining it as: *a packet-based network able to provide telecommunication services and able to make use of multiple broadband, QoS-enabled transport technologies and in which service-related functions are independent from underlying transport-related technologies. It enables unfettered access by users to networks and to competing service providers and/or services of their choice. It supports generalized mobility which will allow consistent and ubiquitous provision of services to users* [3]. The first sentence in the above definition is intentionally general so as not to put any restriction on the packet technology used. There is no doubt, however, that the Internet protocol (IP) is the most promising packet technology for NGN. It is important to note that this does not mean that NGN is merely an enhanced version of the Internet. The Internet has been constructed on the *open and autonomous* concept in that it is terminal-centric and network-transparent. So far, the open and autonomous concept has brought about global acceptance of the Internet, but some serious concerns have been recognized, in particular regarding security and privacy. Who authenticates a terminal and authorizes the terminal's access to the network and to another terminal? Quality of service is also hard to guarantee under the autonomous concept. Who arbitrates the conflict of network usage between terminals? NGN aims to support terminals by putting sufficient capabilities on the network side. This is to ensure that a wide variety of terminals can be handled by the trustworthy network entities and to coordinate terminals if some conflict occurs such as traffic congestion. Both the Internet and NGN may use the same protocol, but the way they use it and the concept behind its use are fundamentally different.

The next step after the definition was to set up solid principles to be followed when implementing of the network design. One well-known approach to network design is provided by the seven-layer model in the Open Systems Interconnection (OSI). While the seven layers were originally used to realize advanced modularity, the model is now seen as too rigid to accommodate the very flexible layering that is currently needed. For example, NGN may require one protocol data unit, which is traditionally used in a lower layer, to be used in a higher layer and encapsulated by another protocol that was previously used in a higher layer. This contradicts the layer ordering rule of OSI. OSI also does not describe the same protocol as being used twice for different purposes such as IP-in-IP encapsulation, an NGN requirement. Another example is that the definitions of a termination point and the portion between the points are also fluid. One link, which previously meant a limited portion between nodes, may extend end to end. Moreover, one end, originally a termination point, may become a new relay point. Accordingly, the simple principle or reference model needed for the NGN era segregates the seven layers into two strata: one purely dedicated to delivering IP packets, which may include QoS guarantee, IP-level mobility, and security, and the other providing service-related control, which is required for web browsing, email exchange, IP telephony, video-conferencing, and so on. These two strata are called the transport stratum and the service stratum [4].

3. Services Examples and Assumed Communication Environment for NGN Release 1

Requirements were collected in terms of the services expected by customers as well as communication environments assumed by operators. Among the numerous services optimistically listed in the standard [8], the following two are recognized as being the most important ones under the study scope of Release 1.

- Multimedia services (including PSTN/ISDN simulation services providing PSTN/ISDN-like service capabilities with session control over IP interfaces and infrastructure, messaging services, push-to-talk over NGN, and point-to-point interactive multimedia services)
- PSTN/ISDN emulation services providing PSTN/ISDN service capabilities and interfaces using adaptation to the IP infrastructure

In addition, public interest aspects (such as emergency communications, support for the disabled, and lawful interception) should be taken into account because of the significance of NGN to daily life. E-mail and web browsing were not the focal points of Release 1. With regard to the communication environment, the following goals were identified for Release 1:

- a common platform for packet transport by IP technology with enhanced security and reliability;
- a common platform for service control through the use of IMS technology;
- extensive terminal support including multiple broadband access and fixed mobile convergence;
- media processing inside the network to allow the addition of content-level value by the network;
- extensive interworking with other networks including existing networks;
- smooth interaction with the applications provided by the information communication technology (ICT) industry.

According to the expected service examples and the assumed communication environment described above, Recommendation Y.2201 [9] lists high-level requirements and related capabilities to support the service objectives of NGN Release 1.

4. NGN Overview

The functional NGN architecture [7] was designed to meet the identified requirements [9] and to support the examples of expected services [8]. An overview of the NGN architecture is shown in Fig. 1. The NGN functions are divided into service stratum functions and transport stratum functions according to the NGN principle identified above [4].

Three major reference points are identified: User Network Interface (UNI), Network Network Interface (NNI), and Application Network Interface (ANI). UNI is for customers, while NNI is for other networks.

This clear distinction between UNI and NNI means that the interfaces to customers and to other networks are likely to be different. This differentiation is an indicator of the shift away from the open and autonomous concept of the Internet. The fundamental assumption of the Internet is that every component should be treated as an identical building block. That concept allows any combination of any components to grow. The NGN, however, is different. At the very least, UNI and NNI are different in the sense that the peer entity at the customer side and that at the network side are assumed to be different in terms of capability, scale, role, and responsibility. Another consideration behind the emphasis on external interfaces such as UNI and NNI is that the internal interfaces inside the NGN may actually be different from UNI and NNI. This distinction allows the network operator to construct its network in a flexible manner while ensuring that the key external interfaces conform. At the UNI and NNI boundaries, the NGN has an opportunity to protect itself against malicious attack or unexpected behavior by customers or other networks, while still maintaining services to ordinary customers.

ANI suggests a clear entry channel for creating services and applications over NGN. This will be useful in attracting new service developers with new ideas and new technologies. It is also expected that the ANI will draw the attention of new service providers who will promote new usages of the NGN in conjunction with the NGN operator.

Inside the NGN, the transport stratum delivers IP packets while enabling the Network Attachment Control Functions (NACF) to perform terminal management, and the Resource and Admission Control Functions (RACF) to realize IP resource management. The service stratum provides service control. For Release 1, session-related service control was investigated in detail to provide PSTN/ISDN emulation/simulation services and interactive multimedia conversational services. This part makes full use of the IP Multimedia Subsystem (IMS), originally developed by the 3rd Generation Partnership Project (3GPP) [5]. The separation of the service and transport strata allows flexibility in various aspects. One is installation independency. The equipment used on stratum is independent of the equipment used on the other stratum, allowing flexible deployment scenarios to meet the capacity requirements of each component. New service capabilities can be realized by introducing new servers while transport equipment remains unchanged. Even if the new service fails to become popular, the service-independent transport stratum can continue

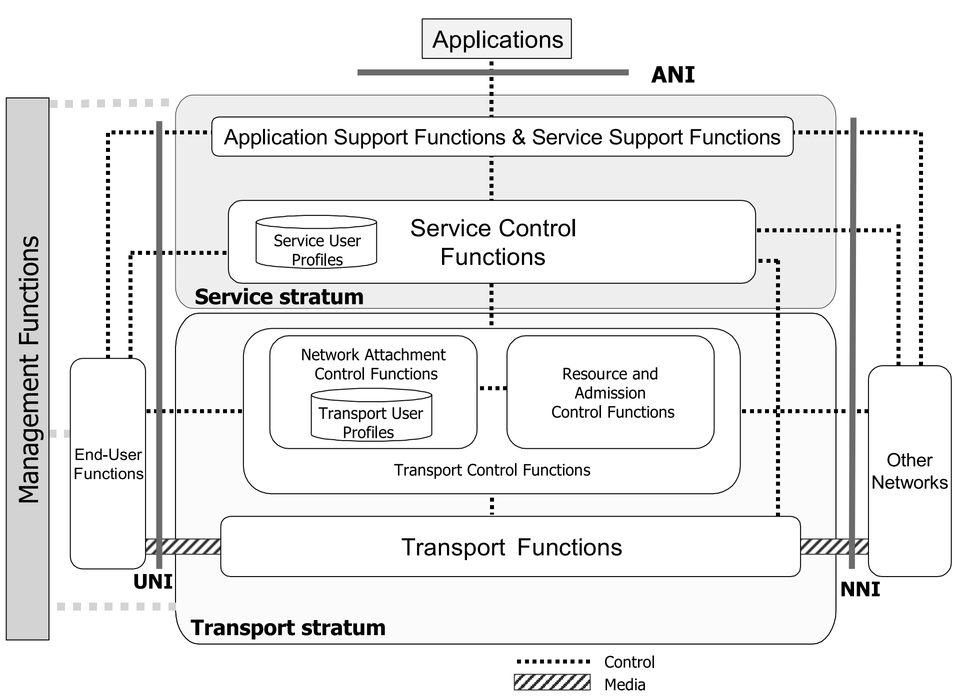


Fig. 1 NGN architecture overview.

to be used for other services. Another aspect is migration independency. Transport facilities can be upgraded or replaced with new technologies without changing service provisioning facilities. In an extreme case, a common transport stratum may be used by different retail sections of the same provider group. Such separation or modularity is a unique feature of the NGN architecture.

5. Detailed Functional Architecture

5.1 Methodology of Detailing Functional Architecture

Any study of the functional architecture of NGN should not only yield possible network capabilities as abstract concepts, but should also lead to protocol designs that have a real impact on equipment production. After requirements have been identified, the next step is rough categorization of capabilities, as described in the previous section, which should be followed by more fine grain categorization. In ITU-T standards, the concept of a Functional Entity (FE) has been used to make more concrete the functional blocks that are closer to actual devices. Functional entity is defined as follows [7]: *An entity that comprises an indivisible set of specific functions. Functional entities are logical concepts, while groupings of functional entities are used to describe practical, physical implementations.* An FE still has an abstract or logical nature and allows flexibility, but it represents a very useful level of functional description that enables engineers to easily imagine tangible equipment. Minor adjustment, such as the merger or separation of a limited number of particular FEs, will immediately provide the basis for protocol design. According to this FE-based convention or notation, all functional blocks in the detailed architecture are appended with FE, such as CSC-FE and AMG-FE. The strict FE-based convention used in ITU-T does cause some very minor misalignment of functional block names between ITU-T and other standards bodies (i.e., 3GPP and The European Telecommunications Standards Institute, The Telecoms & Internet converged Services & Protocols for Advanced Networks (ETSI TISPAN)), but the names are similar enough for people to easily associate them.

A unique situation that we faced during NGN architecture design was that it involved a kind of reverse engineering. The discussions showed us that IMS is the most promising candidate for the NGN control functions in the service stratum, while another approach, the so-called Softswitch- or call-server-based approach seems reasonable as a short-term but certainly reasonable solution. To avoid meaningless debate, the detailed functional architecture is designed to allow both IMS-based and call-server-based approaches. In this sense, the detailed functional architecture still means a generic one that can be instantiated in various ways.

The generic functional architecture of NGN Release 1 is shown in Fig. 2. Although the architecture was designed around Release 1 requirements, we are very confident that it is flexible enough to support new requirements beyond Release 1 without any dramatic changes to the whole structure.

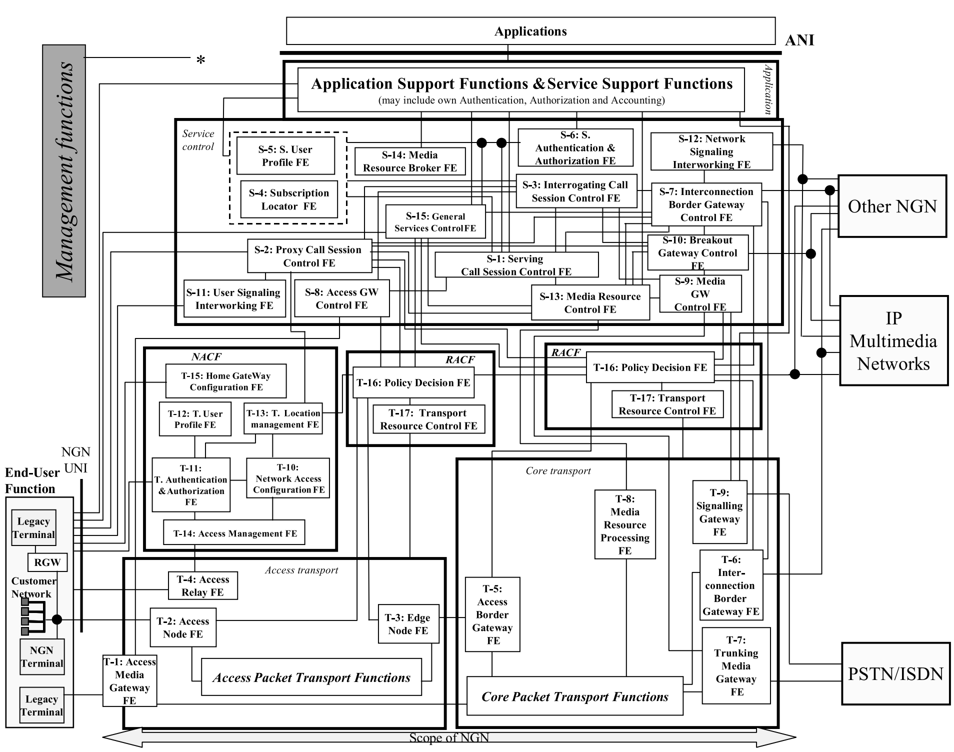


Fig. 2 NGN generalized functional architecture.

5.2 Call Session Control FEs

The most important FEs in the NGN are the three Call Session Control (CSC) FEs indicated by S-1, S-2, and S-3 in Fig. 2. They are a kind of session initiation protocol (SIP) server and are directly imported from 3GPP IMS.

The Serving CSC-FE (S-CSC-FE) indicated by S-1 in Fig. 2 is the ordinary SIP server for a user terminal. It is equivalent to a local switch in PSTN. It manages user terminals that subscribe to S-CSC-FE. It handles registration from the terminal and outgoing and incoming session requests from and towards the terminal, and it interacts with applications to provide value-added services. The difference between S-CSC-FE and ordinary SIP servers is that subscriber-related information that is to be managed by S-CSC-FE is dynamically downloaded from the Service User Profile Functional Entity (SUP-FE) for each registration. In other words, the S-CSC-FE responsible for a particular user terminal may change with the registration. SUP-FE is equivalent to the Home Subscriber Server (HSS) in 3GPP and the User Profile Server Function (UPSF) in TISPAN. At the registration phase, an incoming registration message from a terminal is first forwarded to SUP-FE. An S-CSC-FE responsible for the following process is then selected. This loose coupling between S-CSC-FE and the user profile makes it easy to shift one S-CSC-FE to another in the case of S-CSC-FE failure or an increase in the number of S-CSC-FEs necessitates an increase in processing capacity.

The Proxy CSC-FE (P-CSC-FE), the second CSC-FE indicated by S-2 in Fig. 2, is the first contact point for the user terminal. Every message from the user terminal should go through P-CSC-FE. The P-CSC-FE, which is separated from S-CSC-FE, is an extension of S-CSC-FE. It is assumed to be located in the visited network in the case of mobile network operation, which usually provides roaming service. The roaming service can be provided even in a fixed network since NGN adopts the P-CSC-FE concept. Since P-CSC-FE is assumed to be in the visited network, it controls media-related FEs in terms of QoS control and network address translation and firewall (NAT/FW) control. This is because S-CSC-FE has full responsibility for providing session-level services, but the responsibility is limited to the session level and so many have nothing to do with media-level services. Actual media packets, which follow signaling packets, do not have to enter the service domain where the S-CSC-FE resides.

The Interrogating CSC-FE (I-CSC-FE) is the third CSC-FE and is indicated by S-3 in Fig. 2. It is located at the boundary of the SIP service domain and distributes incoming SIP requests to an appropriate S-CSC-FE that is inside the service domain. This "message distributor" helps to discover the S-CSC-FE currently responsible for a particular user terminal.

In summary, P-CSC-FE enables a roaming configuration and I-CSC-FE enables more flexible and dynamic assignment of S-CSC-FEs.

The Service User Profile Functional Entity (SUP-FE), indicated by S-5 in Service Control in Fig. 2, maintains user profile information at the service-stratum level and can be instantiated by multiple entities. A Subscription Locator Functional Entity (SL-FE), indicated by S-4 in Service Control in Fig. 2, is the message distributor for multiple SUP-FEs.

The Service Authentication and Authorization Functional Entity (SAA-FE), indicated by S-6 in Service Control in Fig. 2, performs authentication and authorization at the service-stratum level.

5.3 Interworking FEs

Since many kinds of networks have already been deployed, the NGN should interwork with various other networks including the existing TDM-based network (TDM: time division multiplexing). Such an interworking scenario requires media packet processing in addition to signaling packet processing. It should be noted that a media packet refers to a packet that carries content information that users primarily want to exchange (e.g., voice, video, and text), while a signaling packet refers to a packet that carries control information to set up and release the arrangement needed by the media packets. NGN strictly applies the modeling principle of service and transport decoupling, and every media processing FE is controlled by a separate controller FE. Three interworking scenarios are considered.

- Interworking with another packet-based network: An Interconnection Border Gateway Control Functional Entity (IBC-FE) controls Interconnection Border Gateway Functional Entities (IBG-FEs), indicated by S-7 and T-6 in Fig. 2.
- Interworking with other PSTN/ISDN networks (i.e., trunking interworking): A Media Gateway Control Functional Entity (MGC-FE) controls Trunking Media Gateway Functional Entities (TMG-FEs), indicated by S-9 and T-7 in Fig. 2.
- Accommodation of PSTN/ISDN terminals (i.e., access interworking): An Access Gateway Control Functional Entity (AGC-FE) controls Access Media Gateway Functional Entities (AMG-FEs), indicated by S-8 and T-1 in Fig. 2.

The first of these interworking scenarios may involve additional FEs: Breakout Gateway Control Functional Entities (BGC-FEs), indicated by S-10 in Fig. 2, for intelligent selection of the PSTN/ISDN breakout point. A User Signaling Interworking Functional Entity (USIW-FE) and Network Signaling Interworking Functional Entity (NSIW-FE), indicated by S-11 and S-12 in Fig. 2, perform signaling protocol conversion such as that between SIP and H.323. Since the target signaling protocols cover a very wide range, the exact role of signaling interworking needs further study.

For trunking interworking, a Signaling Gateway Functional Entity (SG-FE), indicated by T-9 in Fig. 2, is used for lower-layer protocol conversion between NGN and PSTN, so it resides in the transport stratum.

5.4 FEs Related to Media Resource Processing

Media content can be directly handled by NGN; it is carried as the payload of a media IP packet. One simple but important need is transcoding when two terminals fail to share a common codec. This is very common when fixed and mobile terminals communicate with each other. Since the codecs were developed for much earlier bandwidths, no single codec satisfies both fixed and mobile access requirements, so the intermediate NGN should intervene to convert payloads. A conference bridge is another example where media processing is needed. In accordance with the decoupling principle, the Media Resource Control Functional Entity (MRC-FE) controls multiple Media Resource Processing Functional Entities (MRP-FEs), indicated by S-13 and T-8 in Fig. 2, to perform media processing. This decoupling principle refers to the separation into strata, described in Sect. 4. The NGN standard requires that the decoupling principle be followed. However, due to the lack of detailed specifications, in particular from operational and maintenance aspects, it is possible that MRC-FE might not work with MRP-FE even though a protocol was been adopted. To cope with this situation, another higher-level controller, the Media Resource Broker Functional Entity (MRB-FE), indicated by S-14 in Fig. 2, is introduced to select the MRC-FE and associated MRP-FEs.

Media resource processing needs further study because it involves interaction with other FEs. It is important to design this interaction carefully because it may impact the QoS, such as session setup time. Media resource processing is the most expensive facility in the network, so careful implementation is essential. This was one of the controversial issues raised during discussions of the architecture. After real implementations have been reviewed, MRP-FE may be integrated with other FEs, in particular, those related to interworking with other networks.

5.5 Transport FEs

Separation between the access and core transport functions has been introduced as in existing networks. The criterion adopted to distinguish core functions from access ones is whether or not switching occurs. In the access transport functions, the Access Node Functional Entity (AN-FE) and Edge Node Functional Entity (EN-FE), indicated by T-2 and T-3 in Fig. 2, constitute a pipe or tunnel that only aggregates upstream traffic towards the core network and distributes downstream traffic towards user terminals. Switching, which identifies the ultimate destination of IP packets and distributes them to the ultimate destination, is assumed to be not a concern of access. It is noted that AN-FE and EN-FE are allowed to be IP-aware to perform policy enforcement on a per-IP flow basis, IP address translation, and firewall control per IP flow. In the NGN, QoS, NAT, and FW controls are managed by RACF, so AN-FE and EN-FE are controlled by RACF.

At the entrance of the core network, an Access Border Gateway Functional Entity (ABG-FE), indicated by T-5 in Fig. 2, is positioned as the first switching point of IP packets.

5.6 Network Attachment Control Functions (NCAF)

User terminals are referred to as network attachments. Terminal configuration and maintenance are supported by Network Attachment Control Functions (NACF), which provide IP address assignment, identification and its assurance of the location associated with the assigned IP address, and home gateway management.

5.7 Resource and Admission Control Functions (RACF)

The most complicated issue is how to guarantee end-to-end QoS in NGN. If end-to-end really implies from one end terminal to the other end terminal across customer networks that are out of the control of the network, the network should rely on terminal capabilities and control customer networks much more, which is beyond the current study scope. For UNI-to-UNI, at least, RACF is tasked with achieving QoS assurance. NGN aims to introduce admission control over the packet infrastructure. Resources include not only bandwidth, but also address space, and these should be treated together. Triggered by service control functions such as P-CSC-FE, RACF decides whether a new session can be established without affecting already established sessions. Network conditions may be collected to assist in the making of such a decision. RACF includes the IP address and port number reservation and allocation for NAPT, as well as the opening and closing of a pin hole, which is the gate for IP flow. Access resources can be managed in Release 1, while core resources need further study. How to handle interaction among multiple RACF sets is an important issue for scaling RACF as part of Release 2.

5.8 General Services Control Functional Entity (GSC-FE)

Besides the important CSC-FEs mentioned above, another service control has been identified. The General Services Control Functional Entity (GSC-FE), indicated by S-15 in Fig. 2, is intended to model an RTSP (realtime streaming protocol) proxy, which is widely used in current commercial networks providing IP streaming service. It was hard to see a clear need for a new FE in addition to CSC-FE without looking into specific protocols, so GSC-FE is a place holder for future extension, in particular for IPTV support.

5.9 Service Support and Application Support Functions

Further details of service- and application-related FEs in and over NGN are shown in Fig. 3. The expected service environment includes Open Service Architecture (OSA) and Parlay type application servers, SIP application servers, and

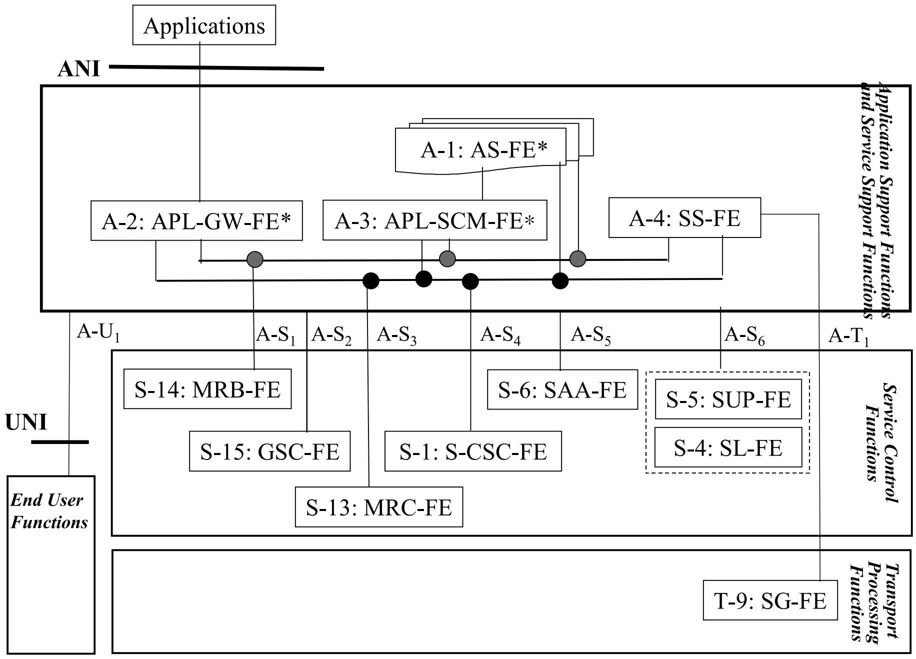


Fig. 3 Application/service support functions.

legacy intelligent network type applications. Detailed specifications are being studied outside ITU-T and more outside input is expected, such as from Open Mobile Alliance (OMA) and the Parlay Group.

5.10 Management Functions

Operation and management functions will play a key role in NGN. The number of functionalities to be covered is increasing, the modularity of equipment is becoming finer, the numbers of manufacturers and business players involved are increasing, and the service menu for customers is becoming more varied. However, convergence or integration of operations and maintenance is required for many reasons. Reductions in the operational cost will determine the success of NGN. Operational aspects are being studied in ITU-T SG4. The first framework Recommendation M.3060/Y.2401 [10] has been published. M.3060/Y.2401 presents the management requirements, general principles, and architectural requirements for managing NGN to support business processes to plan, provision, install, maintain, operate, and administer NGN resources and services. It also defines the concepts of the NGN management architecture (i.e., its business process view, functional view, information view, and physical view) and their fundamental elements, It also describes the relationships among the architectural views and provides a framework for describing the requirements for the specifications of the management physical view from the viewpoints of management functions and information. A logical reference model for partitioning the management function, the Logical Layered Architecture (LLA), is also provided in Release 1.

6. Open Issues for Future Enhancement

In this section, we consider how to extend the NGN functional architecture described above.

6.1 Support for IPTV

Release 1 focuses on realtime conversational services such as IP telephony and video chatting, for which SIP is the most suitable control protocol. Communication is established between two peers using the one-to-one information transmission mode. The information that is exchanged is usually conversation between two persons, and so does not require any rights management or value chain control. Realtime broadcasting over IP, which requires multicast capability to conserve bandwidth, has not been studied in depth. Two items that have not been studied are i) digital rights management for protecting the content holder's rights and ii) the handling of metadata. Metadata could be embedded along with the content and could point to associated content attributes and so create an additional value chain. These capabilities, not included in Release 1, should be part of Release 2.

With FGNGN having drawn much attention, another focus group, the focus group for IPTV (FG IPTV), was launched in July 2006. Although it needs some time to understand the different requirements from different players such as the broadcasting industry, video codec vendors, cable TV providers, and NGN operators, the most important topic is how to support IPTV over NGN. The most fundamental question is to what extent NGN functional components should be used ? Should RACF be used for providing QoS? Should the IMS platform be used? Are terminals for IPTV moving like mobile phones and portable computers, or are they stationary like TV sets in the home?

The IMS platform, on which NGN is based, is very powerful for providing roaming service for portable (handheld) television sets, but it might be too complicated if we are supporting static television sets that do not require a roaming service. The adoption of IMS may depend on the assumed service scenario.

The QoS provided by RACF seems attractive, but RACF has been developed so far without consideration of multicast capability and so will need enhancement.

6.2 Support for Full Mobility

The current IMS provides roaming capability that allows a terminal to use different network access points by identifying the location, access point, and address of the terminal; the terminal is associated with its profile information stored in the home network. Although NGN adopts the IMS platform, which originated with mobile network technologies, the current IMS does not support handover. Here, handover means a way to provide service continuity (i.e., continuity of media communication without interruption caused by packet loss or delay) while the terminal is moving. Service continuity is not included in Release 1, but should be part of Release 2.

Service continuity needs extra capabilities to be provided by the underlying IP layer rather than employing SIP control in the service stratum. Even for the IP layer, several ideas for achieving IP layer mobility have been proposed. These include mobile IPv6 extension for local mobility management [11], [12] and fast handover support [13], additional consideration of a proactive handover procedure [14], cross-layer interaction between Layer 2 and IP to make movement detection efficient [15], and IP address translation instead of IP-in-IP encapsulation, which is almost always assumed for IP mobility [16], [17]. During handover, QoS should be maintained [18]. To what degree are we going to rely on the end-to-end principle associated with mobile IP [17]? As noted in the previous section, how much commonality between mobile- and fixed-oriented terminals should we seek for a particular function? Once IP mobility is being provided successfully, the roaming capability provided by IMS may be redundant because more convenient and comprehensive mobility is provided by IP mobility. We need to clarify the roles of each mobility function provided by IMS and mobile IP.

Mobility between IP and non-IP access technologies is another issue to be solved in order to achieve smooth migration to an all-IP environment.

6.3 Converged Services

Draft Recommendation Y.2013 [19], which was produced just after the initial thirteen documents listed in Table 1, deals with the interaction between different service components in the service stratum. This allows converged services such as those between multimedia services triggered by PSTN/ISDN emulation service and those between IN-like services and multimedia services. Y.2013 is expected to provide coordination functions that are beyond the scope of the generalized functional architecture in Y.2012 [7]. Though the model adopted in Y.2013 has not yet been well proven and needs further study, it stimulates us by showing attractive new service instances based on the convergence of different service components. This is another important topic within the scope of Release 2.

6.4 Support for Radio Frequency Identification

Radio frequency identification (RFID) [20] is referred to as an ID-based application in ITU-T to focus on its networking aspects. Although we need further clarification of the requirements, one possibility is to use a very simple tag that is read by a tag reader. The tag reader should interact with servers to collect the tag's profile, discover the appropriate server to process the tag-related information, and so on. In this scenario, is the session provided by IMS applicable and useful? Can IMS solve the issues associated with RFID, such as privacy and confidentiality? What traffic characteristics are indicated by this scenario? What QoS is required? How many tags and how much generated traffic should we assume? What capability can we assume for the small RFID chips and the readers? The application of RFID is not limited to just a single use case. We have to start by categorizing RFID use cases and examine the requirements for individual categories. Support for RFID, or transaction service in other words, is not included in Release 1, but should be part of Release 2.

6.5 Performance Consideration

It has been a couple of years since the idea of IMS was proposed and its initial specifications were completed. This technology and its implementation, however, are not yet proven because the originally anticipated users, i.e., mobile network operators, are still thinking about when and how to introduce IMS widely in their networks. Indeed, fixed network operators may deploy IMS first. As described in the section on CSC-FEs, many FEs will need to interact to set up and manage a session. This may create a burden for CSC-FEs. In addition, we have to think about the volume of signaling messages to be handled. Since SIP has the potential to convey messages other than those for session control, we should implement CSC-FE very carefully. One example of such an extra message is the presence signal. Presence may require periodic message exchange to quickly detect a lack of presence or availability. This scenario generates more signaling message exchanges than an ordinary session setup. Another example is the messages generated by an application supporting text-chatting between customers. SIP can convey messaging information for text-based chatting whose traffic volume is also unknown. The current architecture has not yet been challenged in terms of processing capacity for signaling messages including extra messages. Performance verification and feedback from commercial use will strengthen and advance the IMS and thus NGN.

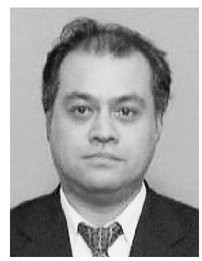
7. Conclusion

This paper reviewed the NGN architecture in conjunction with Y.2012 [4]. NGN deployment is just beginning and some feedback from commercial networks is expected to improve performance and capability. Release 1 is literally the first stage, though the NGN architecture is flexible enough to evolve to meet future demands.

References

[1] M.H. Reeve, C. Bilton, P.E. Holmes, and M. Bross, "21CN," Communications Engineer, vol.3, no.5, pp.25–29, Oct.-Nov. 2005.
[2] http://www.ntt.co.jp/ir/calendar_e/index.html
[3] ITU-T Recommendation Y.2001, "General overview of NGN," Dec. 2004.
[4] ITU-T Recommendation Y.2011, "General principles and general reference model for next generation networks," Oct. 2004.
[5] ITU-T Recommendation Q.1741.4, "IMT-2000 references to release 6 of GSM evolved UMTS core network," Oct. 2005.
[6] ITU-T Recommendation Y.2111, "Resource and admission control functions in next generation networks," Sept. 2006.
[7] ITU-T Recommendation Y.2012, "Functional requirements and architecture of the NGN," Sept. 2006.
[8] ITU-T Supplement 1 to Y.2000 series, "NGN release 1 scope," July 2006.
[9] ITU-T Recommendation Y.2201, "NGN release 1 requirements," to be approved in April 2007.
[10] ITU-T Recommendation M.3060/Y.2401, "Principles for the management of next generation networks," April 2006.
[11] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in IPv6," IETF RFC3775, June 2004.
[12] H. Soliman, C. Castellucia, K. El Malki, and L. Bellier, "Hierarchical mobile IPv6 mobility management (HMIPv6)," IETF RFC 4140, Aug. 2005.
[13] R. Koodi, ed., "Fast handovers for mobile IPv6," IETF RFC 4068, July 2005.
[14] C. Vogt and M. Zitterbart, "Efficient and scalable, end-to-end mobility support for reactive and proactive handoffs in IPv6," IEEE Commun. Mag., vol.44, no.6, pp.74–82, June 2006.
[15] Draft IEEE Std. P802.21/D00.05, "Draft IEEE standard for local and metropolitan area networks: Media independent handover service," Jan. 2006.
[16] P.K. Best and R. Pendse, "Quantitative analysis of enhanced mobile IP," IEEE Commun. Mag., vol.44, no.6, pp.66–72, June 2006.
[17] M. Yabusaki, T. Okagawa, and K. Imai, "Mobility management in All-IP mobile network: End-to-end intelligence or network intelligence?," IEEE Commun. Mag., vol.43, no.12, no., pp.16–24, Dec. 2005.
[18] R.L. Aguiar, S. Sargento, A. Banchs, C.J. Bernardos, M. Calderon, I. Soto, M. Liebsch, T. Melia, and P. Pacyna, "Scalable QoS-aware mobility for future mobile operators," IEEE Commun. Mag., vol.44, no.6, pp.95–102, June 2006.
[19] ITU-T draft Recommendation Y.2013, "Converged services framework functional requirements and architecture," to be approved in April 2007.
[20] R. Weinstein, "RFID: A technical overview and its application to the enterprise," IT Professional, vol.7, no.3, pp.27–33, May-June 2005.

Naotaka Morita received his B.E. and M.E. degrees from Nagoya University, Aichi, Japan, in 1985 and 1987, respectively. In 1987, he joined NTT Laboratories, where he engaged in research on ATM systems. Since 2000, he has been studying VoIP and interactive multimedia technologies. Since October 2004, he has been a Vice Chair of SG13 and Chairman of its Working Party 3, which is the lead study group for NGN, in ITU-T. He was a co-leader of Working Group 2, the Functional Architecture and Mobility Group in FGNGN, and an editor of ITU-T Recommendation Y.2012 "Functional Requirements and Architecture of the NGN."



Hideo Imanaka received the B.E., M.E. and Ph.D. degrees in electrical engineering from Mie University in 1985, 1987, and 2001, respectively. After joining NTT Telecommunication Network Laboratories in 1987, he engaged in research on a fiber optic access network architecture and network operation process reengineering methods. From 1996 to 2003, he was engaged in ERP system integration SE in the Solutions Division of NTT Communications. His current work includes the standardization of NGN. Dr. Imanaka is a member of SICE.