Next-Generation PONs: A Performance Investigation of Candidate Architectures for Next-Generation Access Stage 1

Jingjing Zhang and Nirwan Ansari, New Jersey Institute of Technology
Yuanqiu Luo, Frank Effenberger, and Fei Ye, Huawei Technologies

ABSTRACT

Driven by bandwidth-demanding video applications, next-generation access stage 1, proposed by FSAN as the first stage of next-generation passive optical networks, increases the subscriber bandwidth by upgrading the current GPON network. To reduce the upgrading cost, NGA1 conforms to the standardized GPON and the deployed optical distribution network. This article presents the investigation of five candidate NGA1 architectures from the perspective of the MAC-layer bandwidth allocation including the analysis of the traffic characteristics of the subscribers' applications and the criteria in mapping them into proper transmission containers. Extensive simulations were conducted to investigate and compare the performances of the five candidate NGA1 architectures.

INTRODUCTION

Several different types of bandwidth-hungry applications and services, including multimedia-oriented applications such as high-definition television (HDTV), are rapidly being deployed in the access network. Hence, telecommunication operators are driven to upgrade their access networks to provide broader bandwidth for their subscribers.

The Full Service Access Network (FSAN) [1], an affiliation of network operators and telecom vendors, has completed its mission on the gigabit-capable passive optical network (GPON). The International Telecommunication Union — Telecommunication Standardization Sector (ITU-T) G.984 series specifies various aspects of GPON, including the general architecture, the physical layer, the transmission convergence (TC) layer, and the GPON management and control [2]. A typical GPON system provides 2.488 Gb/s of downstream bandwidth and 1.244 Gb/s of upstream bandwidth. GPON possesses the following characteristics [3]:

- GPON directly reflects the requirements of network operators because the GPON standardization is driven by operators through FSAN.
- As with ITU-T G.983 broadband PON (BPON), GPON provides high product interoperability by standardizing a management interface, referred to as the optical-network-unit management and control interface (OMCI), between optical line terminals (OLTs) and optical network units/terminals (ONUs/ONTs). We use ONUs and ONTs interchangeably in this article.
- GPON accommodates three layer-2 networks: asynchronous transport mode (ATM) for voice, Ethernet for data, and proprietary encapsulation for video, thus enabling GPON with full-service support capability, including voice, time division multiplexing (TDM), Ethernet, ATM, leased lines, and wireless extension.
- GPON defines the GPON encapsulation method (GEM) to achieve efficient packaging of user traffic, with frame segmentation to better provide quality of service (QoS) for delay-sensitive traffic such as voice and video applications.
- GPON supports radio frequency (RF) video transmission in the waveband from 1550 nm to 1560 nm.

Having turned over the work of GPON standard maintenance to ITU, FSAN is now studying the next-generation access (NGA). The objective of NGA is to facilitate high bandwidth provision, large split ratio, and extended network reach. FSAN has planned two stages of NGA evolution: NGA1 and NGA2 [1]. NGA1 focuses on PON technologies that are compatible with GPON standards (ITU-T G.984 series) and compatible with the current optical distribution network (ODN) as well. In contrast, NGA2 is a long-term solution with an entirely new optical network type. The objective of NGA2 is to provision an independent PON scheme, without

0163-6804/09/$25.00 © 2009 IEEE

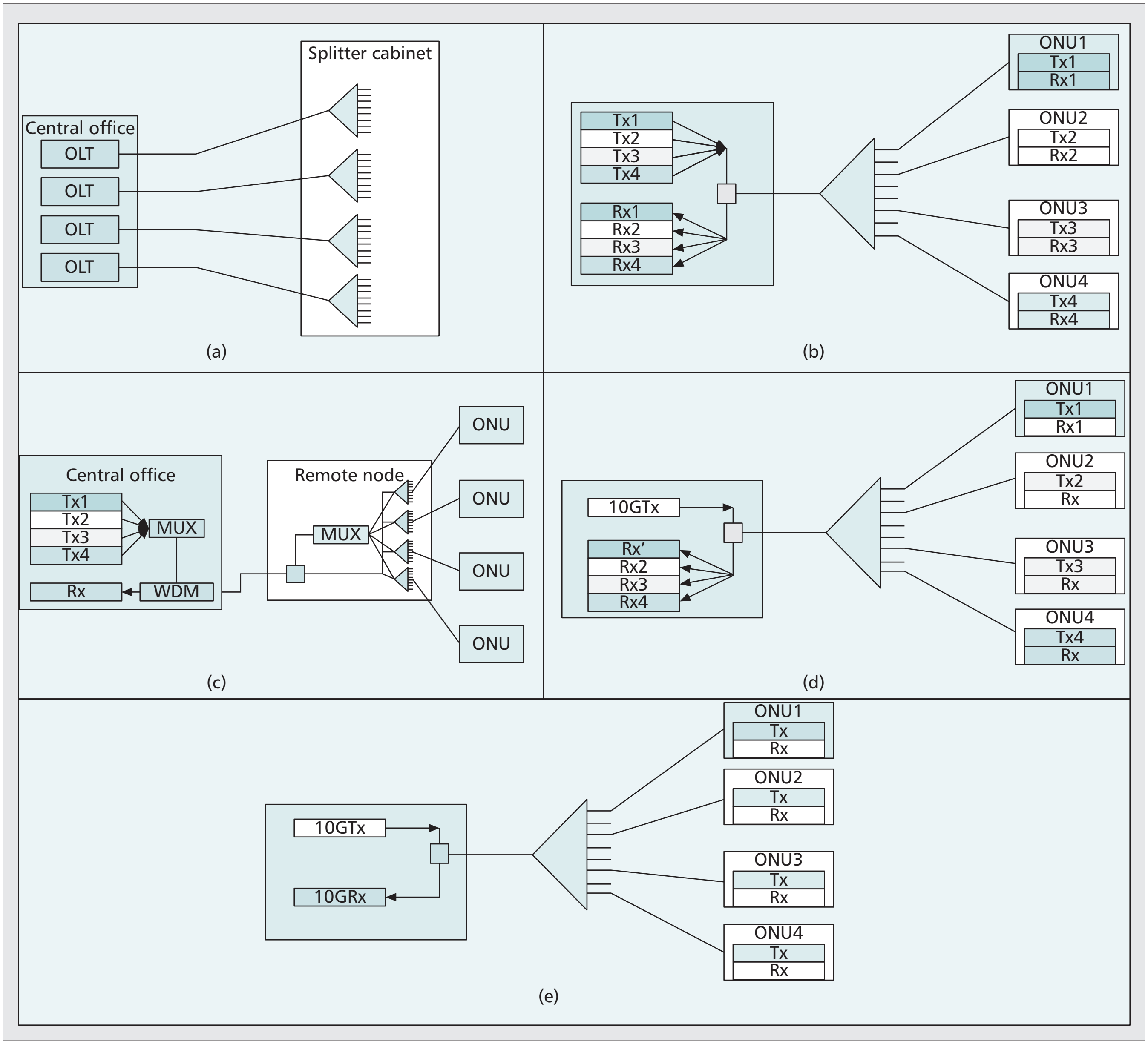


Figure 1. *NGA1 candidate architectures: a) GPON with bidirectional physical PON split reduction; b) GPON with bidirectional wavelength PON split reduction; c) GPON with downstream wavelength PON split reduction; d) XG-PON1 with 10G downstream, Nx2.5G upstream; e) XG-PON2 with 10G symmetric.*

being constrained by the GPON standards and the currently deployed outside plant. This article focuses on candidate NGA1 network architectures, as well as their performance comparison.

Five typical candidate network architectures have been proposed for NGA1 [1]. These five architectures employ different methods in achieving the increase of the upstream and downstream bandwidth per ONU. In this article, we analyze and compare the five candidate NGA1 network architectures in detail. The rest of the article is organized as follows. First, we describe in detail the five candidate architectures of NGA1, as well as the abstraction of these architectures from the perspective of medium access control (MAC)-layer resource allocation. Next, we present the resource allocation model stated in the GPON standard (ITU-T G.984.3). Then, we proceed to a general analysis of the access-network traffic. We then describe our simulation environment, present the simulation results, and make comparisons of the five candidate architectures under various traffic scenarios. Last, a conclusion is drawn.

NGA1 CANDIDATE ARCHITECTURES

On the one hand, given that NGA1 is compatible with the GPON standards and the currently deployed ODN, candidate NGA1 architectures try to leave the ODN part untouched but upgrade its devices at head ends including OLTs and ONUs. On the other hand, like GPON, NGA1 still employs the TDM principle in multiplexing and demultiplexing traffic onto a wavelength channel. The following describes five candidate architectures that upgrade the GPON network in different ways.

		NGA1-1	NGA1-2	NGA1-3	NGA1-4 (2.5G up)	NGA1-4 (2 x 2.5G up)	NGA1-5
Upstream	Number of virtual groups	4	4	1	1	2	1
	Number of ONUs per group	8	8	32	32	16	32
	Rate (Gb/s)	1.244	1.244	1.244	2.488		10
Downstream	Number of virtual groups	4	4	4	1		1
	Number of ONUs per group	8	8	8	32		32
	Rate (Gb/s)	2.488	2.488	2.488	10		10

Table 1. *Abstraction of NGA1.*

NGA1-1: G-PON with Bidirectional-Physical PON Split Reduction

The first scheme simply reduces the PON split ratio physically. It divides a typical PON with 32-way physical splits into four virtual groups with eight-way splits. Such an upgrade does not require any changes of ONUs, but does require three times more OLT equipment. Figure 1a illustrates the architecture for this scenario.

From the MAC-layer resource allocation point of view, the modified architecture can be regarded as four virtual groups with eight ONUs, respectively. The ONUs in each group share the link with the same rate as that of the GPON, namely, 2.488 Gb/s for downstream and 1.244 Gb/s for upstream.

For simplicity, we use NGA1-1 to denote this architecture in the rest of the article.

NGA1-2: G-PON with Bidirectional-Wavelength PON Split Reduction

Rather than reducing the physical split ratio in NGA1-1, this architecture uses multiple, independent wavelengths in each direction to increase the network capacity (Fig. 1b). It is suggested to employ coarse-wavelength division multiplexing (CWDM) technology in the upstream for low cost transmitters at the ONU side and dense-wavelength division multiplexing (DWDM) in the downstream because of the scarce frequency band. This architecture does not require any changes of ODN but requires modification of the transmitters and receivers at both the OLT and ONU side.

Similar to NGA1-1, from the MAC-layer resource allocation point of view, the typical 32 ONUs in this architecture are divided into virtual groups, each of which operates in an independent wavelength. We denote this architecture as NGA1-2 in the rest of the article.

NGA1-3: G-PON with Downstream-Wavelength PON Split Reduction

This architecture (Fig. 1c) utilizes wavelength-division multiplexing (WDM) technology to increase the downstream capacity only, but leaves the upstream intact. Typically, the downstream transmission selects four wavelengths within the GPON downstream band of 1480 nm to 1500 nm. In this case, the ONUs can receive the downstream signal without any modification. NGA1-3 thereby avoids the high cost of upgrading ONUs, as compared to NGA1-2. Similar to the former two architectures, for the downstream scenario, 32 ONUs can be regarded as belonging to four independent virtual groups, each of which operates at a 2.488 Gb/s-link rate and supports eight ONUs. For the upstream scenario, it remains the same as GPON, that is, the link rate is 1.244 Gb/s, and the number of ONUs is 32. In the rest of the article, we denote this architecture as NGA1-3.

NGA1-4: XG-PON1 with 10-G Downstream, Nx2.5-G Upstream

This architecture upgrades the downstream link capacity to 10 Gb/s. The difficulty with the architecture of 10 Gb/s is enabling the burst mode time-division multiple access (TDMA) operated at 10 Gb/s. Because of the limitation of available components and design practices, many simple circuit techniques become impractical when the rate goes beyond 5 Gb/s. Overcoming this limit requires specialized hardware and is thus costly. To minimize additional investment, an architecture was proposed to upgrade only the downstream to 10 Gb/s, but to use one or more 2.5-Gb/s wavelengths in the upstream as shown in Fig. 1d.

This architecture still can be considered as a TDM system both in the downstream and upstream. The downstream transmission is modeled as 32 ONUs sharing a 10-Gb/s link. Depending on the number of available upstream wavelengths, the ONUs in the upstream scenario are divided into a different number of groups operating at 2.5 Gb/s. If two wavelengths are adopted in the upstream, the ONUs in the upstream scenario are divided into two virtual groups, each of which has 16 ONUs sharing a 2.5-Gb/s upstream link. If one wavelength is adopted, it is abstracted as 32 ONUs sharing a 2.5-Gb/s upstream link. This architecture is referred to as NGA1-4.

NGA1-5: XG-PON2 with 10-G Bidirectional

When devices capable of a 10-Gb/s burst mode become commercially available, the architecture with both the downstream and upstream transmission being upgraded to 10 Gb/s can be realized (see Fig. 1e). In this case, the transmission in both upstream and downstream can be abstracted as 32 ONUs sharing a 10-Gb/s link. This architecture is referred to as NGA1-5.

When devices capable of a 10-Gb/s burst mode become commercially available, the architecture with both the downstream and upstream transmission being upgraded to 10 Gb/s can be realized. In this case, the transmission in both upstream and downstream can be abstracted as 32 ONUs sharing a 10-Gb/s link.

Among the five candidate architectures, NGA1-4 is a promising and economical architecture to meet the future bandwidth requirement in NGA1. First, the upstream and downstream bandwidths are increased to 2.5 Gb/s and 10 Gb/s, respectively. The increased bandwidths are potentially able to accommodate the future bandwidth-consuming applications in NGA1. Second, the 2.5-Gb/s burst mode receiver requires lower cost compared to the 10-Gb/s burst mode receiver in NGA1-5, making NGA1-4 a more economical solution for a GPON upgrade.

Essentially, the upgraded systems are still TDM systems. By abstraction, the typical 32 ONUs in GPON are divided into multiple virtual groups, where the ONUs in each group share a link in TDM fashion. Hence, the five candidate architectures can be regarded as TDM systems with different link rates and different numbers of shared ONUs. In this article, we assume that ONUs are evenly divided among the upstream or downstream channels. The upstream and downstream link rates of GPON are 1.244 Gb/s and 2.488 Gb/s, respectively. Table 1 summarizes the characteristics of the abstracted TDM systems in both directions for the five architectures.

For the upstream, NGA1-1 and NGA1-2 are abstracted as four virtual groups, each of which has eight ONUs sharing a 1.244-Gb/s link; NGA1-3 is abstracted as one virtual group with 32 ONUs sharing a 1.244-Gb/s link; NGA1-4 is abstracted as one virtual group with 32 ONUs sharing a 2.488 Gb/s-link if one 2.488 Gb/s link is used, and two virtual groups each of which with 16 ONUs sharing a 2.488-Gb/s link if two 2.488-Gb/s links are used; NGA1-5 is abstracted as one virtual group with 32 ONUs sharing a 10-Gb/s link. For the downstream, ONUs in NGA1-1, NGA1-2, and NGA1-3 are abstracted as four virtual groups, each of which is with eight ONUs sharing a 2.488-Gb/s link; ONUs NGA1-4 and NGA1-5 are abstracted as one virtual group with 32 ONUs sharing a 10-Gb/s link.

BANDWIDTH ALLOCATION IN GPON

NGA1 is designed to be compatible with the GPON standards. A good approach for designing bandwidth allocation in NGA1 would be following the scheme in GPON. This section describes the bandwidth allocation in GPON.

In GPON [2], the frame length is fixed as 125 µs. The basic control unit of bandwidth allocation is a transmission container (TCONT), indexed by allocation ID (alloc-ID). The process that OLT dynamically allocates upstream bandwidth to TCONTs, based on their traffic load, is referred to as dynamic bandwidth allocation (DBA). GPON supports two DBA methods to infer the buffer occupancy status of each TCONT: status-reporting DBA and traffic-monitoring DBA. In status-reporting DBA, ONUs report the TCONT buffer status to the OLT directly. In traffic-monitoring DBA, the OLT infers the buffer status of the TCONT based on the historical information of bandwidth utilization and assigned bandwidth amount.

GPON describes each alloc-ID by the following four-tuple $<R_F, R_A, R_M, X_{AB}>$:

R_F: The ***fixed bandwidth*** reserved in each frame, regardless of the real incoming traffic and network load.

R_A: The ***assured bandwidth*** for the TCONT when the TCONT has enough traffic to consume the assured bandwidth.

R_M: The ***maximum bandwidth*** assigned to the TCONT.

X_{AB}: An indication of non-guaranteed bandwidth. It can be referred to as either non-assured bandwidth or best-effort bandwidth. The assignment of the non-assured bandwidth and best-effort bandwidth adopts the following rules:

–For non-assured bandwidth, the OLT assigns bandwidth to the TCONT in proportion to the sum of the fixed bandwidth and assured bandwidth of that TCONT.

–For the best-effort bandwidth, the OLT assigns bandwidth to the TCONT in proportion to the value of the maximum bandwidth minus the guaranteed bandwidth of that TCONT (the sum of fixed bandwidth and assured bandwidth).

The DBA in GPON follows a strict priority hierarchy: fixed bandwidth, assured bandwidth, non-assured bandwidth, and best-effort bandwidth. First, the OLT assigns the upstream bandwidth to the fixed bandwidth requirement of each TCONT. Then, the OLT allocates to the assured bandwidth components of each TCONT, as long as the TCONT has enough traffic to consume the assured one. Third, the OLT satisfies the requirement of non-assured bandwidth. Last, the OLT allocates the remaining bandwidth to the best-effort bandwidth component.

GPON further defines five TCONT types with different combinations of traffic descriptors:

TCONT type 1: Fixed bandwidth component only, suitable for constant bit-rate (CBR) traffic that is delay and jitter sensitive

TCONT type 2: Assured bandwidth component only, suitable for traffic that does not have strict delay and jitter bounds

TCONT type 3: Assured bandwidth component plus non-assured bandwidth component, suitable for variable-rate, bursty traffic that requires an average rate guarantee

TCONT type 4: Best-effort bandwidth component only, suitable for non delay-sensitive bursty traffic

TCONT type 5: Any combination of traffic descriptors, suitable for most of the general traffic

TRAFFIC MODELING AND MAPPING

We next analyze the traffic characteristics of the subscribers' applications and then discuss the criteria in mapping them into proper TCONTs, which determine the QoS performance of the application.

TRAFFIC MODELING

The access network accommodates a variety of applications, such as Web surfing, video conferencing, standard definition television/video on demand (SDTV/VoD), HDTV/VoD, real-time TV, file sharing, video upload/download, multiplayer gaming, and telecommuting. Table 2 illustrates the typical upstream and downstream bandwidth consumptions of some applications [4].

Table 2 shows that applications are diversified with diverse bandwidth demands, and they will be even more diversified with technological advances. However, voice, video, and data still constitute the main components of many applications. For example, video conferencing, video on demand, video upload/download, and real-time TV can be considered as video traffic, but with different encoder modes and QoS requirements. In the following, we present the main characteristics of voice, video, and data traffic at the application layer.

Voice — There are typically two kinds of voice traffic: legacy telephone and voice over IP (VoIP). Legacy telephone traffic is 64 Kb/s CBR traffic and conveyed in TDM fashion with frame length equaled to 125 microseconds, whereas voice-over-Internet Protocol (VoIP) traffic generated by speech codec is commonly modeled with the on-off model [5]. The on period is referred to as the voice activity period, whereas the off period is the voice inactivity period. The voice traffic during the on period is encoded into bit streams. The coding rates of the common codec schemes, such as G.711, G.723, and G.729, range from 2.15 Kb/s to 64 Kb/s. Then, the bit stream undergoes the packetization process with periods ranging from 10 ms to 50 ms. It implies that the minimum interval of two packets for a single VoIP connection is 10 ms/125 ms = 80 GPON frames.

Video — A digital video consists of video frames displayed at a certain frame rate. Each individual video frame contains many picture elements (pixels). For the uncompressed video frame, typically each pixel is quantized by eight bits. To improve the encoding efficiency, many video codecs such as H.263, H.263+, MPEG2, and MPEG4, have been developed and deployed. Some of them use a non-scalable, single-layer encoding technique, whereas some adopt scalable multilayer encoding techniques. It was shown that encoded videos without rate control on the stream level have strong correlations over a relatively long time period due to correlations of the video contents. However, such correlations drop sharply for encodings with rate control, which keeps the bit rate close to a certain number [6].

For transmission over packet-switched networks, the video frames are then packaged into packets, each of which may contain multiple video frames, a single video frame, part of a video frame, or a frame in one layer. Then, the packets are transmitted at once or at a constant bit rate over a frame or multiple frames.

Application	Upstream (Mb/s)	Downstream (Mb/s)
Web surfing	< 1	1–2
Video conference, premises surveillance	2–3	2–3
SDTV VOD, telecommuting	2–3	3–4
File sharing and home video sharing	4–5	5
Real-time SDTV, network PVR	5	5
Multiplayer gaming, interactive distance learning	6	6
Premises Web hosting	6–7	6–7
Large file sharing	12	12
Video upload/download	13	13
HDTV VoD	1–2	15
Network hosted applications and storage	10	> 25
Next-generation 3D TV	1–2	> 25

■ **Table 2.** *Upstream and downstream traffic constituents [4].*

Data — File sharing and Web browsing are the main applications that generate data traffic. It was shown that traffic generated by world-wide Web (www) browsing on the Internet possesses a self-similarity property [7]. In the Gnutella P2P file-sharing system, the size of an audio file ranges from 1 Mbyte to 8 Mbytes, the size of software ranges from 10 Kbytes to 100 Mbytes, and the size of a video ranges from 1 Mbyte to 100 Mbytes, as reported in 2002. The file size is changing and should increase with time.

Other Traffic Types — There are many other applications. Their characteristics are determined by the specific properties of these applications. For example, it was shown that online gaming traffic with a fixed set of players is stable and predictable. The traffic contains periodic bursts of tiny packets. In addition, the session time of a player is not heavy-tailed but drops sharply as time goes by [8].

Different applications have different characteristics. The characteristics of an application must be taken into account when mapping requests of an application into the proper TCONTs and setting parameters for TCONTs.

TRAFFIC MAPPING

Because GPON adopts the strict priority hierarchy in bandwidth allocation, the received QoS of a request is determined by the TCONT type into which the request is mapped. The mapping must consider not only the QoS requirement of the application but also the traffic characteristics.

TCONT type 1 is designed to carry delay-sen-

Regarding the maximum bandwidth (MB), a value that is too small could degrade the performance when the traffic consists of large bursts, whereas a value that is too large could result in unfairness and easily could be exploited by malicious users to consume the bandwidth.

sitive CBR traffic. Such traffic includes legacy TDM voice traffic and rate-controlled broadcast TV and HDTV streams with CBR characteristics. Another example of an application that potentially is mapped into TCONT type 1 is generated by business subscribers, who are willing to spend more to acquire a guaranteed QoS for certain of their applications such as video conferencing. If these applications are mapped into TCONT type 1, these applications have a virtually independent channel irrespective of the bandwidth requests of other applications. Therefore, their QoS always can be guaranteed. This mapping also can be considered as a scheme for realizing a layer-1 virtual private network (VPN) for business subscribers.

VoIP traffic modeled with an on-off model is an example that constitutes TCONT type 2. The bandwidth during the *off* period is wasted if it is mapped into TCONT type 1. Thus, it is more bandwidth efficient to map VoIP traffic into TCONT type 2, during which bandwidth is allocated only when there is traffic.

For TCONT type 3, an example is the video-on-demand application with the variable bit-rate (VBR) characteristic. The assured bandwidth part provisions the application with an average rate guarantee. On the other hand, the VBR characteristic induces either excessive or inadequate bandwidth provision when the application is mapped into TCONT type 1. For the sake of bandwidth efficiency and QoS provisioning, it is proper to map such traffic into TCONT type 3.

For TCONT type 4, best-effort applications, for example, Web surfing and file sharing, are a proper fit.

For TCONT type 5, online gaming, which aggregates voice, video, and interactive applications, is a suitable mapping.

TCONT PARAMETERS SETTING

After determining the types of TCONT into which an application is mapped, the next task is to decide the specific parameters of that TCONT type, including fixed bandwidth, assured bandwidth, and maximum bandwidth. On the one hand, parameters should guarantee the QoS performance of the application. On the other hand, they should be bandwidth efficient.

For TCONT type 1, ideally, the fixed bandwidth in one GPON frame should be equal to the actual arrival traffic. For CBR traffic, the fixed bandwidth is equal to the traffic rate. For TCONT type 2 with on-off traffic, it is recommended that the assured bandwidth be set as the peak data rate during the *on* period for good delay performance. For TCONT 3, a small packet-loss ratio requires the assured bandwidth part to be larger than the average data rate. For the best-delay performance, the bandwidth allocated in one allocation frame should be greater than the arrived traffic during this frame. Hence, the assured bandwidth should be set as the peak data rate for the smallest delay. However, setting the assured bandwidth with the peak data rate reduces the number of admitted requests as compared to the scheme of setting it with the average data rate. Similar to TCONT type 3, the parameters of TCONT type 5 also depend on the requirements of the specific application. Regarding the maximum bandwidth (MB), a value that is too small could degrade the performance when the traffic consists of large bursts, whereas a value that is too large could result in unfairness and easily could be exploited by malicious users to consume the bandwidth.

THE SIMULATION SETUP AND ENVIRONMENT

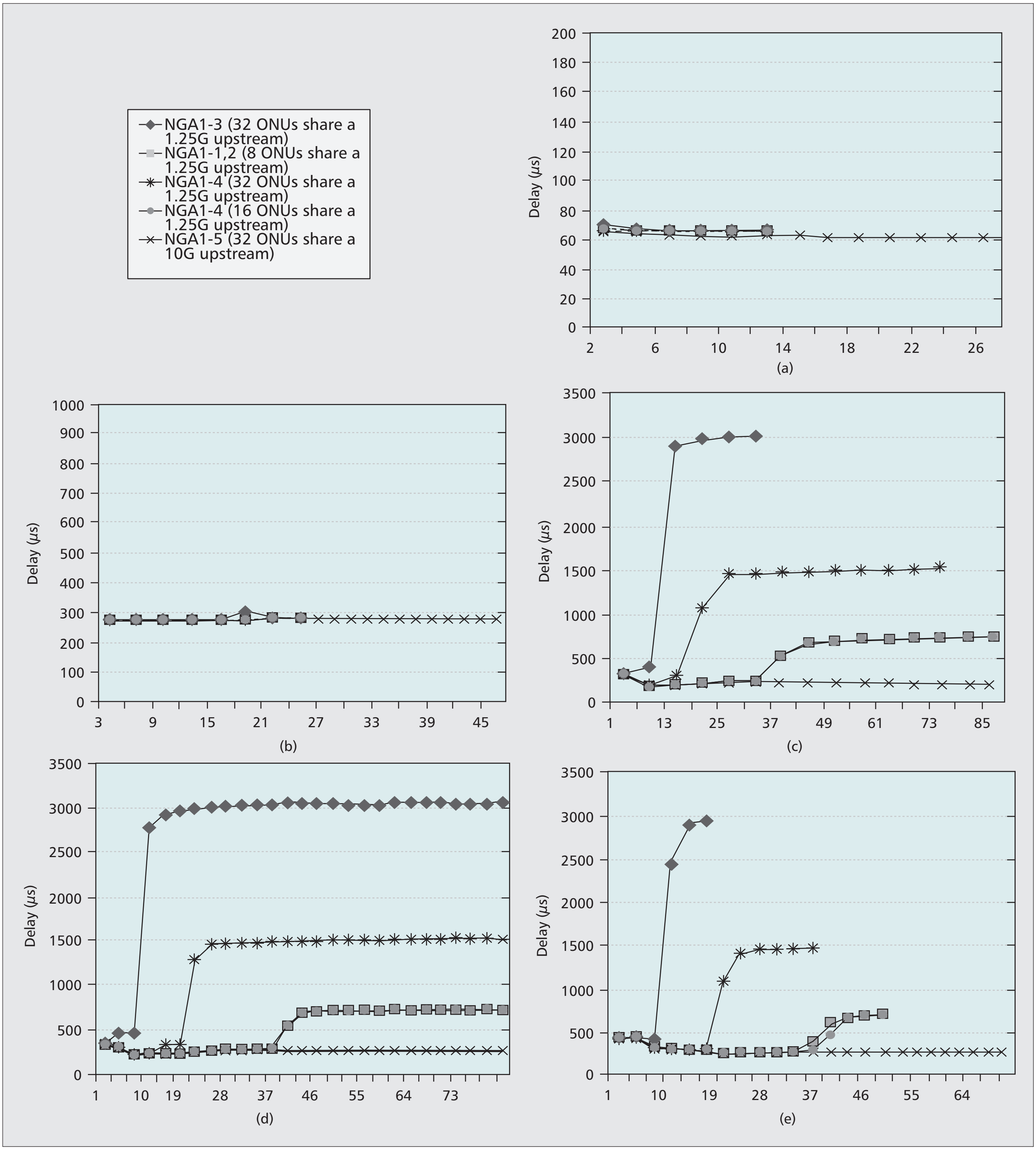
Because upstream-bandwidth allocation is more challenging than downstream-bandwidth allocation, we focus on the upstream-bandwidth allocation in this article. We consider status report DBA in the simulation, that is, TCONT piggyback reports of its queue information to the OLT immediately after data transmission in each frame. Then, the OLT collects the reports from all TCONTs and makes the bandwidth allocation decision.

Traffic Generation — As shown previously, different applications exhibit different characteristics. It is impractical to simulate the traffic of every application, especially as new applications are emerging rapidly. We use CBR traffic for TCONT type 1, on-off traffic for TCONT type 2, and self-similar traffic with burst characteristics for TCONT types 3, 4, and 5 to investigate the statistical gain. In addition, the real traffic constituents of subscribers vary from time to time and from subscriber to subscriber. It is also impractical to simulate every dynamically varied traffic constituent. In the simulation, we consider uniform traffic distribution among ONUs. We assume each ONU has eight TCONTs. All of them are of the same TCONT type.

The self-similar traffic is generated by aggregating a sum of on-off sources with *on* and *off* periods exhibiting Pareto distribution. The main parameters of the self-similar traffic consist of the peak rate, average rate, and Hurst parameters. Denote X as a Pareto-distributed random variable with the Pareto parameter, α. Then, $P(X > x) = (x/x_m)^{-\alpha}$, $1 < \alpha < 2$, $x > x_m$. The average duration of the *on* period is $x_m\alpha/(\alpha - 1)$. Let n be the number of aggregated on-off sources; the rate of the *on* period is r. Theoretically, the parameters of the generated traffic are as follows [8]:

Hurst parameter: $(3 - \alpha)/2$
Peak rate: $n*r$
Average rate: $1/2*n*r$

In the simulation, the same α and x_m are set for the *on* and *off* periods: α is set as 1.1 for the highly self-similar traffic; x_m is set as the time duration of transmitting 800 bytes in its corresponding uplink so as to generate large-size packets. Five on-off sources are aggregated to generate self-similar traffic. The peak rate of the aggregated traffic is set as 8 Mb/s. The self-similar traffic goes through the packetization process for every 1460 bytes or at the end of an *on* period. We assume that the application-layer packets go through Real-time Transport Protocol (RTP), User Datagram Protocol (UDP), IP, Ethernet, and GEM encapsulation before being transmitted, where RTP/UDP/IP is the popular protocol stack for real-time data transmission. Therefore, 8 (RTP) + 12 (UDP) + 20 (IP) + 18 (Ethernet) = 58 bytes is added into each packet before GEM encapsulation. The buffer size of each TCONT is chosen to be 15,180 bytes, capable of

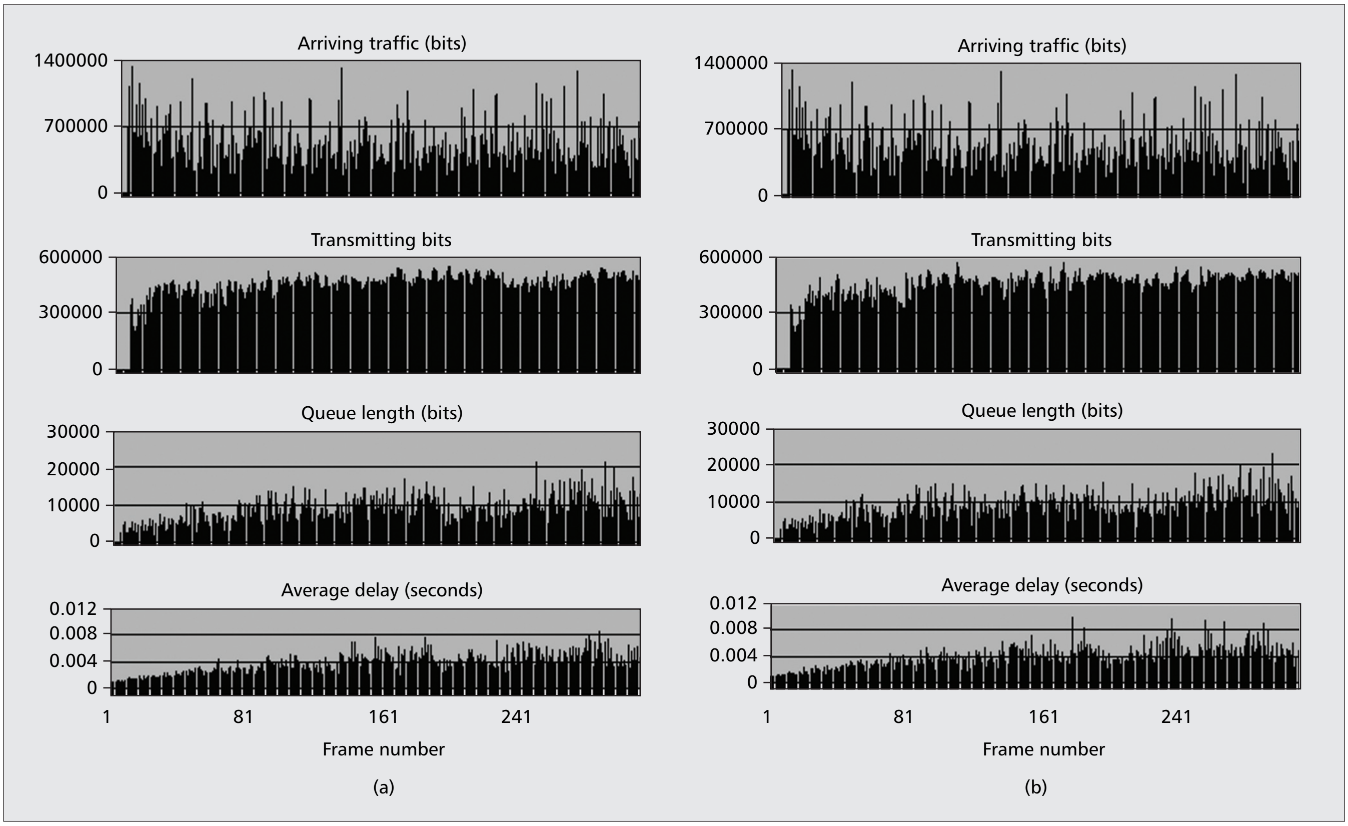


■ **Figure 2.** *Delay of five NGA1 candidate architectures vs. the number of self-similar traffic connections: a) TCONT type 1; b) TCONT type 2; c) TCONT type 3; d) TCONT type 4; e) TCONT type 5.*

accommodating ten packets of 1518 bytes. The propagation delay is set as 250 μs, which equals twice the frame length.

TCONT Parameter Setting — We generate CBR traffic with a rate of 8 Mb/s for TCONT type 1; on-off traffic with a rate of 8 Mb/s during the *on* period for TCONT type 2; self-similar traffic with Hurst parameter of 0.95; and an average rate of 8 Mb/s for TCONT types 3, 4, and 5. To reflect the incoming traffic amount, the fixed bandwidth and assured bandwidth are set to be linearly proportional to the number of connections:

TCONT type 1: The fixed bandwidth (FB) is set to be the CBR.
TCONT type 2: Assured bandwidth (AB) is set to be the data rate during the on period.
TCONT type 3: AB is set to be 1/2 of the average data rate.



(a)

(b)

■ **Figure 3.** *Comparison between a) NGA1-1/NGA1-2; b) NGA1-4 with two upstream wavelengths.*

- TCONT type 5: FB is set to be 1/4 of the average data rate at the application layer plus some bandwidth for lower layer headers.

The MB of all the TCONTs are set to be twice the peak data rate.

SIMULATION RESULTS AND ANALYSIS

In the simulation, we investigate the performance of each TCONT type independently. For a given TCONT type, we fix the traffic of other types of TCONTs but vary the number of connections of this particular TCONT type to obtain a different traffic load. We want to assess the performance trend with respect to the variation of the traffic load of a TCONT type. The performance results shown in all figures are on a per ONU basis. Each data point of the performance curve is averaged over 2000 frames.

Figure 2 presents the delay performance of the five types of TCONTs. For TCONT type 1, fixed bandwidth is reserved in every frame irrespective of the incoming traffic. The average delay is half of the duration of the allocated frame. The number of supported requests is proportional to the network bandwidth. Similar to TCONT type 1, TCONT type 2 does not share bandwidth with others either. Its delay performance does not change with the increase of admitted requests. Because the allocation of bandwidth for TCONT type 2 varies with the reported bandwidth requirements, the packet delay is restricted by the report-grant process and hence is greater than that of TCONT type 1. Simulations show that the average delay is about 280 μs. For TCONT types 3, 4, and 5, the average delay is restricted by the report/grant process and is around 250 ms when the network is lightly loaded (Figs. 2c–2e). When the number of connections is increased to a value that causes network overload, packets are backlogged in the queue and are lost when the queue becomes full. We assume the first come, first served (FCFS) principle, implying that once admitted, the packet cannot be transmitted until all of the packets that arrived prior to its arrival are transmitted. Consider NGA1-5 as an example; when the network is heavily loaded, every ONU is allocated with 10 Gb/s*125 μs/32/8 ≈ 4882 bytes per frame. For the buffer size of 15,180 bytes and the packet length of 1518 bytes, one packet has a chance to be transmitted after the earlier nine packets are transmitted. Its delay will be 15180/4882*125 ≈ 388(μs) (Fig. 2e). For the network with a small capacity, the allocated bandwidth per ONU is small, thus resulting in longer queuing time.

Next, we investigate the statistical gain of the network. We focus on two architectures: NGA1-1 and NGA1-2 with 32 ONUs being divided into four virtual groups, each of which shares a 1.25 Gb/s link; NGA1-4 with 32 ONUs being divided into two virtual groups, each of which shares 2.5 Gb/s. Under the scenario of non-uniform traffic distribution among ONUs, NGA1-4 should outperform NGA1-1/NGA1-2 because of the statistical gain. Here, we consider uniform traffic distribution among ONUs. All ONUs have eight TCONTs of type 4 with best-effort traffic.

Figure 3 shows the input traffic, transmitted traffic, average queue length, and average packet delay in each frame. The two architectures are fed with exactly the same traffic with bursty characteristics. The transmitted traffic becomes much smoother compared to the input traffic. Both the queue length and the delay fluctuates drastically across frames. There is no obvious difference in the queue length and delay performance between the two architectures.

CONCLUSION

With the rapid emergence of bandwidth-demanding applications, NGA1, proposed by FSAN, intends to increase the subscriber bandwidth while being compatible with the standardized GPON. In this article, we first analyzed five candidate NGA1 architectures from the viewpoint of MAC-layer bandwidth allocation, in which these architectures can be abstracted as TDMA systems with different link rates and numbers of ONUs. We also analyzed the traffic characteristics of current diversified applications. Last, simulations were conducted to assess the performances of five types of TCONTs for the five candidate architectures, respectively. Generally, when the traffic is uniform among all ONUs, the performance is determined by the average obtained bandwidth per ONU.

REFERENCES

[1] Full-Service Access Network (FSAN); http://www.fsanweb.org

[2] ITU-T Rec. G.984 series.

[3] F. Effenberger *et al.*, “An Introduction to PON Technologies,” *IEEE Commun. Mag.*, Mar. 2007, pp. S17–S25.

[4] W. Jiang and H. Schulzrinne, “Analysis of On-Off Patterns in VoIP and Their Effect on Voice Traffic Aggregation,” *Proc. ICCCN 2000*, Las Vegas, NV, Oct. 16–18, 2000, pp. 82–87.

[5] P. Seeling, M. Reisslein, and B. Kulapala, “Network Performance Evaluation Using Frame Size and Quality Traces of Single Layer and Two Layer Video: A Tutorial,” *IEEE Commun. Surveys & Tutorials*, vol. 6, no. 3, 3rd qtr. 2004, pp. 58–78.

[6] T. Karagiannis, M. Molle, and M. Faloutsos, “Long-Range Dependence: Ten Years of Internet Traffic Modeling,” *IEEE Internet Comp.*, vol. 8, no. 5, Sept. 2004, pp. 57–64.

[7] W.C. Feng *et al.*, “A Traffic Characterization of Popular On-Line Games,” *IEEE/ACM Trans. Net.*, vol. 13, no. 3, June 2005, pp. 488–500.

[8] Generator of Self-Similar Network Traffic v. 2; http://wwwcsif.cs.ucdavis.edu/~kramer/code/trf_gen2.html

ADDITIONAL READING

[1] M. Wegleitner, *DSL Forum Q3-07 Quarterly Meeting*, Nashville, TN, Aug. 28, 2007.

BIOGRAPHIES

JINGJING ZHANG (jz58@njit.edu) received a B.E. degree from Xi’an Institute of Posts and Telecommunications, China, in 2003, and an M.E. degree from Shanghai Jiao Tong University, China, in 2006, both in electrical engineering. She is working toward her Ph.D. degree in the Department of Electrical and Computer Engineering, New Jersey Institute of Technology (NJIT), Newark. Her research interests are in the areas of broadband access networks, planning of access networks, and optimization in network resource allocation.

NIRWAN ANSARI [S’78, M’83, SM’94, F’09] (nirwan.ansari@njit.edu) received his B.S.E.E. (summa cum laude) from NJIT in 1982, his M.S.E.E. degree from the University of Michigan, Ann Arbor, in 1983, and his Ph.D. degree from Purdue University, West Lafayette, Indiana, in 1988. He joined the Department of Electrical and Computer Engineering at NJIT as an assistant professor in 1988, and has been a full professor since 1997. He has also assumed various administrative positions at NJIT. His current research focuses on various aspects of broadband networks and multimedia communications. He authored *Computational Intelligence for Optimization* (Springer, 1997; translated into Chinese in 2000) with E. S. H. Hou and edited *Neural Networks in Telecommunications* (Springer, 1994) with B. Yuhas. He has also contributed over 300 technical papers, over one-third of which have been published in refereed journals/magazines. He is a senior technical editor for *IEEE Communications Magazine*, and also serves on the advisory and editorial boards of five other journals. He has served or serves the IEEE in various capacities such as Chair of IEEE North Jersey Communications Society Chapter, Chair of IEEE North Jersey Section, member of IEEE Region 1 Board of Governors, Cluster Chair of IEEE Communications Society Networking TC Cluster, Chair of IEEE Communications Society Technical Committee on Ad Hoc and Sensor Networks, and Chair/TPC Chair of several conferences and symposia. Some of his recent awards include the NJIT Excellence in Teaching in Outstanding Professional Development (2008), the IEEE Leadership Award (2007, from the Central Jersey/Princeton Section), the IEEE MGA Leadership Award (2008), and the NCE Excellence in Teaching Award (2009). His teaching was also recognized by his designation as an IEEE Communications Society Distinguished Lecturer.

YUANQIU LUO [M] (yluo@huawei.com) received her B.S. degree in electronics and information systems and her M.E. degree in electrical engineering from Shandong University, China. She received her Ph.D. degree in electrical engineering from NJIT. She is a senior research engineer in the advanced technology department of Huawei Technologies USA. Her research interests are in the areas of broadband access networks, next-generation optical access, network modeling, and integrated optical/wireless networks. She is actively involved in the standards for next-generation passive optical networks. Before joining Huawei Technologies USA, she was with NEC Laboratories America.

FRANK J. EFFENBERGER completed his doctoral work in 1995. In 2006 he became director of FTTx in the advanced technology department of Huawei Technologies USA. In 2000 he began working at Quantum Bridge Communications (now a part of Motorola), where he managed system engineering in its PON division. This work supported the development and standardization of advanced optical access systems based on BPON and GPON technologies. Previous to that role, he had a position with Bellcore (now Telcordia), where he analyzed all types of access network technologies, focusing on those that employed passive optical networks. He witnessed the early development of the FSAN initiative and the development of the APON standard. He remains heavily involved in standards work, and has been a leading contributor and editor of the major PON standards in the ITU. He is working on forward-looking fiber access technologies, including the 802.3av 10 G EPON and ITU NGA topics. In 2008 he became the chairman of ITU-T Q2/15, the group that creates standards for optical access systems.

FEI YE [M] (yephay@huawei.com) received both his B.S. degree in computer science and his M.E. degree in computer systems and architecture from Huazhong University of Science and Technology, China. He is a research engineer in the advanced technology department of Huawei Technologies Shenzhen. His research interests are in the areas of broadband access networks, next-generation optical access, and integrated optical/wireless networks.

> *The transmitted traffic becomes much smoother compared to the input traffic. Both the queue length and the delay fluctuates drastically across frames. There is no obvious difference in the queue length and delay performance between the two architectures.*