논문 2004-41TC-4-4

UMTS 패킷 코아 망에서 신호 및 데이터 트래픽 성능 분석

(Performance Evaluation of Signaling and Data Traffic in UMTS Packet Core Networks)

김 병 철*, 이 재 용*

(Byung-Chul Kim and Jae-Yong Lee)

요 약

GSM에서 발전한 UMTS 망의 경우 SGSN(Serving GPRS Support Node)과 GGSN(Gateway GPRS Support Node)을 중심으로 한 패킷 핵심 망을 포함하고 있다. 이러한 UMTS 망에서는 가입자 이동성, 위치 등록, 가입자 분포 등을 고려하여 망을 설계하고 장비의 제원을 정하여야 한다. 또한 향후 IMT-2000의 주요 트래픽이 데이터이기 때문에 인터넷 데이터 트래픽의 특성인 burstiness를 반영한 SGSN과 GGSN의 설계 기준이 마련되어야 하고 제약적인 시스템 조건 하에서 성능의 변화를 살펴보아야 한다. 이에 본 논문에서는 먼저 패킷 호 가입자의 신호 트래픽을 분석하며 이에 따른 망 성능 변화를 살펴보고, burstiness 특성을 가지는 패킷 데이터가 패킷 코아 망에 미치는 영향을 다양한 환경 하에서 시뮬레이션으로 분석한 뒤 UMTS 망에 새로이 추가하는 두개의 주요 망 요소인 SGSN과 GGSN의 요구 성능을 구하여 그 결과를 바탕으로 새로운 패킷 모드 핵심 망의 설계 기준을 제시하도록 한다.

Abstract

UMTS network, evolved from GSM, includes packet core network that consists of SGSNs and GGSNs. Service providers should consider subscriber mobility, location registration, and subscriber distribution when designing packet core networks and network elements. Since one of the major traffic sources for IMT-2000 will be data which has bursty characteristics, new design guidelines for dimensioning of SGSN and GGSN should be proposed under various constraints of system parameters. In this paper, we first evaluate the performance of signaling traffic for packet call subscribers. After that, we also obtain the impact of bursty data traffic on the SGSN and GGSN by simulation, and suggest new dimensioning guidelines for packet core network of UMTS under various environments.

Keywords: UMTS data traffic, UMTS signaling traffic, UMTS packet core networks

Ⅰ. Introduction

Recently, there have been lots of active researches in the area of mobile Internet, since the Internet and mobile communication systems are widely deployed and demand for mobile multimedia data has been increased. This means the service paradigm has been changed from simple voice to multimedia services including voice, data, and video. Therefore, the design of efficient packet core network becomes very important in the future mobile networks, especially in IMT-2000 network [1].

Packet-mode core network processes data packets based on packet switching method unlike the previous mobile circuit switching communication with the VLR(Visitor Location Register), HLR(Home Location Register), and MSC(Mobile Switching Center). In 3GPP, packet core networks consist of SGSNs

---

* 정회원, 충남대학교 정보통신공학부
(Dept. of Computer & Communication Engineering Chungnam National University)

※ This work was supported by grant No. R01-2003-000-10628-0 from Korea Science & Engineering Foundation
접수일자: 2004년1월31일, 수정완료일: 2004년4월6일

(Serving GPRS Support Node) and GGSNs (Gateway GPRS Support Node), that provide location management, access control management, and gate-way function to external Internet network [2][3].

To deploy UMTS networks efficiently, the guidelines of packet core network design are required at the first stage. But most researches about UMTS packet data have been concentrated on the radio channel performance, not on packet core networks. Thus, in this paper, we evaluate the performance of packet core networks in the aspects of both signaling and data traffic considering various mobility pattern of subscribers, packet arrival rate, service distribution of typical data user, and the number of SGSNs and GGSNs. Also, we get lots of useful performance results by changing the packet service rate and buffer size of SGSN and GGSN. All the performance results are obtained using the event-driven simulation by modeling packet core networks as a queueing network. These results can be used in the design of 3G packet core network elements and dimensioning.

In Section 2, we introduce the UMTS network architecture and signaling protocol first. After that, we propose simulation model for UMTS signaling traffic and some simulation results are shown. In Section 3, we show data traffic performance assuming self-similar Internet traffic model. Finally, we conclude the paper in Section 4.

II. Modeling and Performance Evaluation of UMTS Signaling Traffic

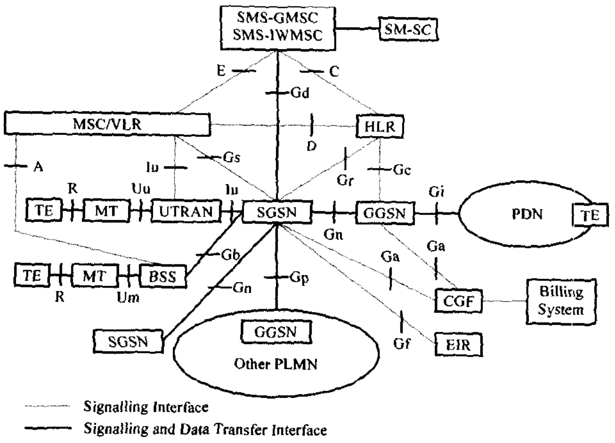


그림 1. UMTS 네트워크 구조
Fig. 1. UMTS network architecture.

UMTS networks have similar structure as GSM circuit switching network which consist of UTRAN (Universal Terrestrial Radio Access Network), RNC, MSC, and HLR/VLR for location management in the case of voice service. However, in order to process data packets efficiently, SGSN and GGSN are added to GSM and some packet control functions are augmented to RNC in the UMTS networks as shown in Fig. 1. The SGSN performs location management, securityand access control functions for each subscriber. It also performs mobility management by creating Mobility Management Context when a mobile subscriber wants to move to another network, and session management by creating PDP(Packet Data Protocol) Context when a user wants to make a session. The GGSN takes the role of connection management to external network and provides data routing to mobile user.

In this Section, we show signaling procedures in UMTS networks, propose its simulation model, and evaluate the signaling traffic performance of packet core networks.

2.1 UMTS signaling protocol

UMTS signaling messages are mainly used for Attach/Detach, PDP Context Activation/Deactivation and mobility management procedures. Attach/Detach procedure performs registration to network for packet data service using mobile terminal[12]. PDP Context procedure requests the service activation to core network for session creation. Then an IP address is assigned to the mobile terminal and resources are allocated to the session by negotiation between the mobile terminal and core network.

The types of signaling message and its flow can be represented as a queueing network model as shown in Fig. 2. For example, if a mobile terminal wants to perform PDP Context Activation, "Activate PDP Context Request" message(②) is sent to SGSN first. After that, "Create PDP Context Request" message(③) is sent to GGSN and corresponding response(④) would be returned to mobile terminal via SGSN after appropriate processing of "Create PDP Context Request"message at GGSN. Fig. 2 shows the procedure of signaling message exchange by representing each message as a circled number. Table 1 lists the generation rate of each message. For example, the generation rate of Active PDP Context message is λ_{ms}.

For the performance evaluation, we assume there are M GGSNs and N SGSNs in the system. All the SGSNs are connected to each GGSN. Also, the generation of PDP Context Activation Request from all the customers through a specified SGSN is assumed to be a Poisson process with average rate λ_{ms} per sec. Unlike the data traffic, it is reasonable to assume exponentially distributed inter-arrival times for signaling traffic when there are many users and random call trials [11]. Also the service time of signaling packet at SGSN and GGSN is assumed to have an exponential distribution with mean $1/\mu_1$ and $1/\mu_2$ respectively.

Moreover, service time between PDP context activation and deactivation is assumed to be an exponential distribution with mean T. Focusing on only SGSN and GGSN performance, we assume there's no queueing delay at HLR, RNC and MSC. Signaling message processing times at above elements are assumed to be exponentially distributed with mean 20msec, 20msec, and 40msec, respectively. That means HLR can process and send corresponding

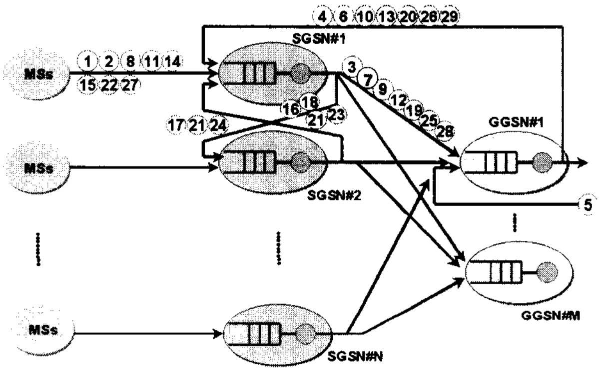


그림 2. 신호 메시지 유형과 플로우
Fig. 2. Signaling message type and flow.

표 1. 신호 메시지 생성율
Table 1. Signaling message generation rate.

NO	Message	Rate
1.	GPRS attach	$\frac{\lambda_{ms}}{n_{call}}$
2.	Activate PDP Context Request	λ_{ms}
3.	Create PDP Context Request	λ_{ms}
4.	Create PDP Context Response	λ_{ms}
5.	First PDU from network	$\lambda_{net}\times N$
6.	PDU notification Request	λ_{net}
7.	PDU notification Response	λ_{net}
8.	Activate PDP Context Request	λ_{net}
9.	Create PDP Context Request	λ_{net}
10.	Create PDP Context Response	λ_{net}
11.	GPRS detach Request	$\frac{\lambda_{ms}+\lambda_{net}}{n_{call}}$
12.	Delete PDP Context Request	$\frac{\lambda_{ms}+\lambda_{net}}{n_{call}}$
13.	Delete PDP Context Response	$\frac{\lambda_{ms}+\lambda_{net}}{n_{call}}$
14.	RA update Request	n
15.	Relocation Required	β
16.	Forward Relocation Request	$\frac{\beta}{N}$
17.	Forward Relocation Response	$\frac{\beta}{N}$
18.	Relocation Detect	$\frac{\beta}{N}$
19.	Update PDP Context Request	β
20.	Update PDP Context Response	β
21.	Forward Relocation complete	β
22.	RA update Request	β
23.	SGSN Context Request	$\frac{\beta}{N}$
24.	SGSN Context Response	$\frac{\beta}{N}$
25.	Update PDP Context Request	β
26.	Update PDP Context Response	β
27.	Relocation Required	γ
28.	Update PDP Context Required	γ
29.	Update PDP Context Response	γ

response message after average 20msec time without queueing delay, if the SGSN queries the HLR. All the generation rates of each signaling message should be decided based on the mobile user's mobility pattern and network topology, and we derive details in the next subsection.

2.2 Simulation Modeling

We used sim++[4][5] for system performance evaluation, which is a software package programmed in C++ and adequate for evaluating a queueing network. Because of the C++ property, developers can easily code a simulation programming using object-oriented model. Sim++ can perform and manage the overall simulation with event scheduling method. Also, it has lots of useful functions, so that traffic modeling can be done easily and statistics about message service time, queue usage time, and average queue size can be easily obtained.

Overall network structure for the simulation of UMTS signaling traffic is shown in Fig. 3. In this structure, the number of cells per RNC is l_1 and the number of RNCs that are connected to a SGSN is l_2.

For modeling the mobile user's mobility, density of users is assumed to have a uniform distribution with and the perimeter of RA(Routing Area) is assumed to be a L. A mobile user is assumed to move with average speed v in an arbitrary direction. With these assumptions, R, crossing rate of RA is given by $R = (\rho * v * L) / \pi$ according to Jain[6].

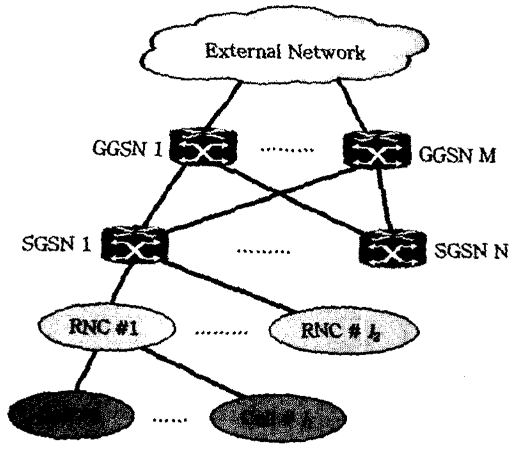


그림 3. 시스템 분석을 위한 망 구조
Fig. 3. Network structure for evaluation.

The network can be simplified as in Fig. 4. For simulation considering the packet service time distribution, the number of simultaneous active mobile users, N, can be expressed as $N = \lambda_{ms} * T$ by Little's law.

For network structure shown in Fig. 4, Inter-RNC (Intra-SGSN) crossing rate, β, can be derived as

$$\beta = (4 * \lambda ms * T * v) / (3600 * \pi * \sqrt{l_1}) \tag{1}$$

Inter-SGSN crossing rate, γ, can be given as

$$\gamma = \beta / \sqrt{l_2} \tag{2}$$

Also, if we assume periodic RA update interval is T_{update}, then RA update ratio of whole activated users becomes.

$$\eta = (\lambda ms * T) / T_{update} \tag{3}$$

2.3 Performance evaluation of the signaling traffic

To get the performance of SGSN, we assume average packet session duration is 3000 seconds and $l_1 = 16$, $l_2 = 4$, $T_{update} = 1$ hour. And, we vary SGSN average service time, $1/\mu_1$, as 1msec ~ 20msec. Then, we can get a SGSN average waiting time(W_S) vs. λ_{ms} as shown in Fig. 5 by simulation. The SGSN average service time includes both the waiting time at the SGSN buffer and service time at SGSN until the signaling message can be sent to another SGSN or GGSN.

In order to evaluate simulation results in other

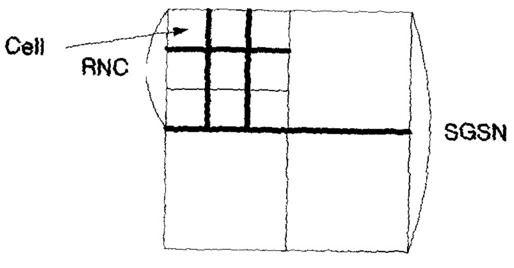


그림 4. 이동성 모델링을 위한 망 구조
Fig. 4. Network structure for mobility modeling.

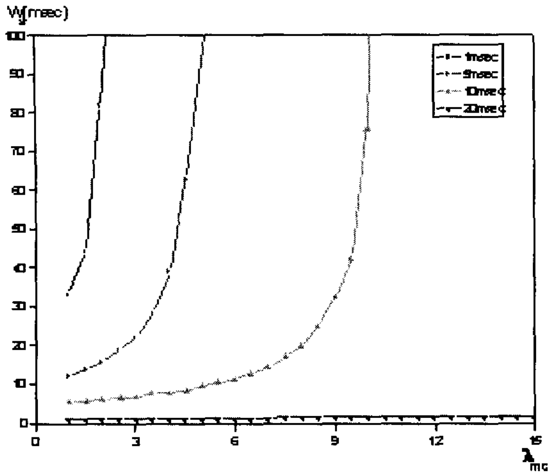


그림 5. 평균 SGSN 대기 시간 대 평균 도착율

Fig. 5. Average SGSN waiting time vs. average arrival rate.

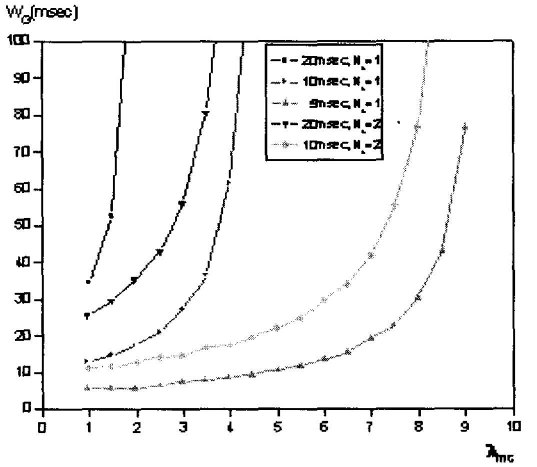


그림 6. 평균 GGSN 대기 시간 대 평균 도착율

Fig. 6. Average GGSN waiting time vs. average arrival rate.

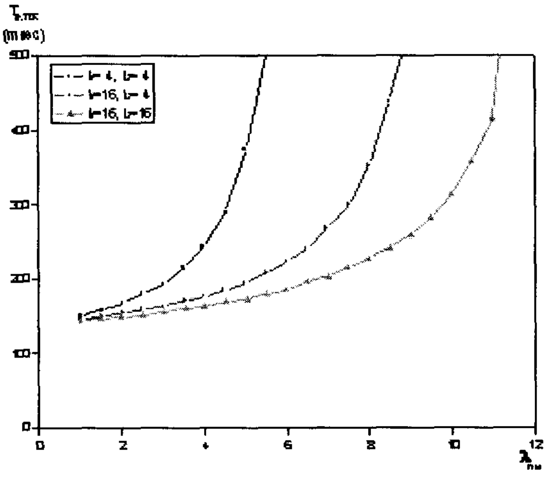


그림 7. RA와 SGSN 영역에 따른 SGSN내 relocation 시간

Fig. 7. Intra-SGSN relocation time for RA and SGSN area.

aspect, we can model the SGSN as an independent M/M/1 queue by using Jackson's results[7]. Then, we can get an average waiting time theoretically [7], and it is given as

$$Ws = (1/\mu_1)/(1-\rho_A) \qquad (4)$$

where, $\rho_A = \lambda_{ms} * F_S / \mu_1$.

F_S represents aggregate arrival factor, which means the average number of incoming signaling messages to a SGSN resulting from single PDP context activation request from a mobile user. This represents the sum of signaling messages(for example, RA update, mobility management, etc.) while a PDP Context exists. In Fig. 5, W_S of the SGSN rapidly increases at one of λ_{ms} values, which occurs at the point of $1-\rho_A = 0$. For example, in the case of $1/\mu_1$ = 20msec, it occurs at λ_{ms} = 2.67, and if $1/\mu_1$ = 5msec, λ_{ms} = 10.97.

In the case of GGSN, the number of connected SGSNs affects the performance. Fig. 6 shows average waiting time of the GGSN for various $1/\mu_2$ and N_g while increasing λ_{ms} value. N_g represents the number of GGSNs. As shown in Fig. 6, we can reduce load and get a required performance using multiple GGSNs because of load balancing effect.

Fig. 7 shows a performance result of mobility management for various l_1 and l_2. l_1 represents the number of cells forming a RA and l_2 represents the number of the RNCs that are connected to a SGSN. The Inter-SGSN handoff time, *TINTER*, increases when l_1 and l_2 decrease since small l_1 and l_2 values increase the load and average waiting time at SGSN. If l_1 and l_2 decrease, RA and SGSN coverage areas also decrease, and therefore the number of RAs and SGSN coverage areas increases. This results in the increase of generation rate of the Inter-RA and Inter-SGSN relocation for a mobile user. This means the increasing traffic load at SGSN and traffic waiting time. So, performance degradation occurs at the SGSN.

III. Modeling and Performance Evaluation of UMTS Data Traffic

In this Section, we simulate the data traffic characteristics and evaluate the packet loss probability and latency time. Using these results, we propose a guideline of packet core network dimensioning.

3.1 Packet Length Distribution of UMTS Data

UMTS networks have lots of short data packets, some middle sized packets and few long sized packets as suggested in[8]. Short-sized packets follow Mobtex distribution model, which has uniform distribution between 15~45 bytes on the uplink and 58~172 bytes on the downlink. Middle-sized packets follow Railway model, which has truncated exponential distribution as follows.

$$X = -170\log(U),\quad 0 < X < 1000\,bytes \tag{5}$$

Here, U is a uniform random variable that has values between 0 and 1. Long-sized packets follow FUNET model, which has truncated Cauchy distribution as follows.

$$X = \tan(\pi\,(U-0.5)) + 0.8 \quad (kbytes),\qquad 0 < X < 10\ kbytes \tag{6}$$

3.2 Self-similar Traffic Model

Leland[9] found that Internet traffic has self-similar characteristics through lots of measurements for Internet traffic for many years. According to these measurements, traffic is very bursty and has very similar characteristics over wide time ranges. For example, both of traffic measured at 0.01msec periods and 100s periods have very similar bursty characteristics[10]. So, in this paper, we use self-similar traffic generator for performance evaluation of UMTS data traffic. Self-similar traffic can be generated using Pareto random variable with heavy-tail distribution. Pareto distribution has the following probability distribution function.

$$F(x) = \Pr[Z \le x] = 1 - \left(\frac{a}{x}\right)^{c} \tag{7}$$

$$\overline{F}(x) = 1 - F(x) \tag{8}$$

Self-similar traffic used for simulation of data traffic performance is obtained by generating bandwidth according to Pareto distribution at every t period. If we assume that generated bandwidth at any instant is BW, then total traffic amount B is

$$B = BW \times \Delta t \tag{9}$$

Packet traffic is generated by dividing B by previously described packet length distribution. In

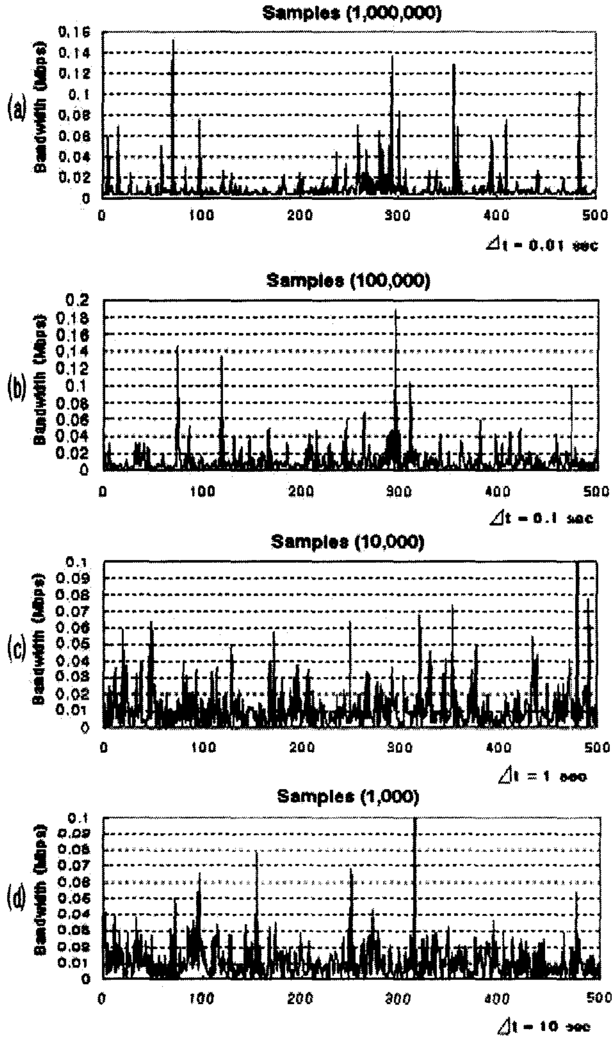


그림 8. 자기 유사 트래픽 생성의 예(c=1.5)

Fig. 8. Self similar traffic generation example (c=1.5).
(a) t = 0.01 sec (b) t = 0.1 sec (c) t = 1.0 sec
(d) t = 10 sec

Fig. 8, we can see the generated traffic has self-similar characteristics by grouping 10 sample and obtaining an average bandwidth at c = 1.5. One can notice that traffic has very similar bursty characteristics even if time scale becomes 1,000 times larger than the original.

3.3 Queueing network model of packet core networks

We have used the queueing network model for simulation network topology shown in Fig. 9.

Incoming traffic at 5 GGSNs is distributed into 2 SGSNs equally. The important system parameters we have to consider are GGSN buffer size, GGSN link capacity, SGSN buffer size, SGSN link capacity, average arrival bandwidth and c value. In this paper, we concentrate on the packet loss probability and latency time of SGSN buffers because the SGSN is a bottleneck point. And, we only evaluate the downlink traffic since downlink traffic is much heavier than the uplink traffic.

3.4 Performance evaluation of data traffic

Fig. 10 and Fig. 11 show the loss probability and latency time at the SGSN queue when SGSN buffer size varies, link bandwidth between SGSN and GGSN is 10Mbps and burstiness parameter c is 1.7. If the average arrival bandwidth for GGSN is 1 Mbps, then effective average arrival bandwidth at a SGSN becomes 5/2Mbps=2.5Mbps, since there are 5 GGSNs. From the figures, one can see that increasing the buffer size is not a good approach for reducing loss probability, since it results in increased latency time. So, good approach is to increase the link bandwidth.

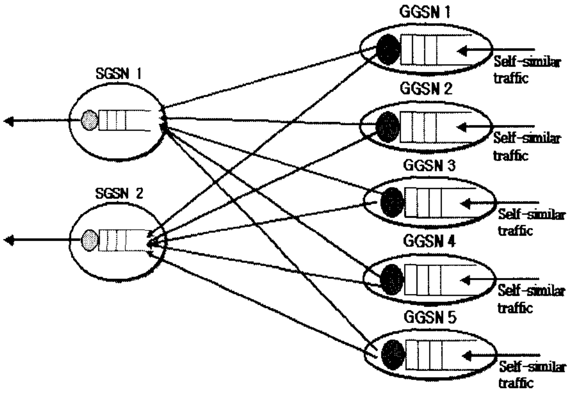


그림 9. 데이터 트래픽을 위한 UMTS 망의 큐잉 네트워크 모델
Fig. 9. Queueing network model of UMTS network for data traffic.

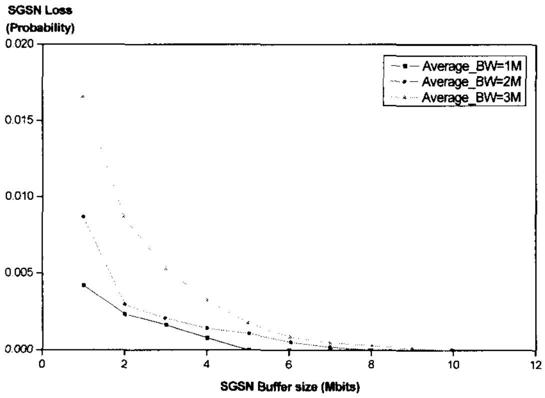


그림 10. SGSN 손실 확률 대 버퍼 크기
Fig. 10. SGSN loss probability vs. SGSN buffer size.

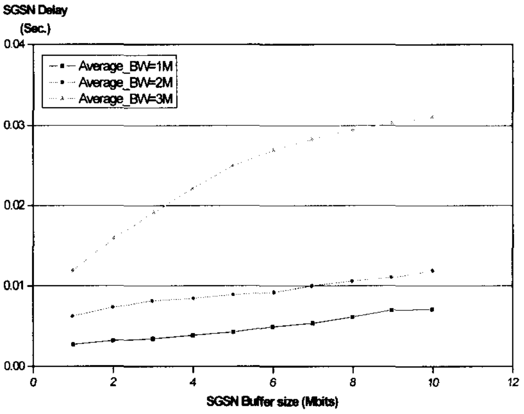


그림 11. SGSN 지연 시간 대 버퍼 크기
Fig. 11. SGSN latency time vs. SGSN buffer size.

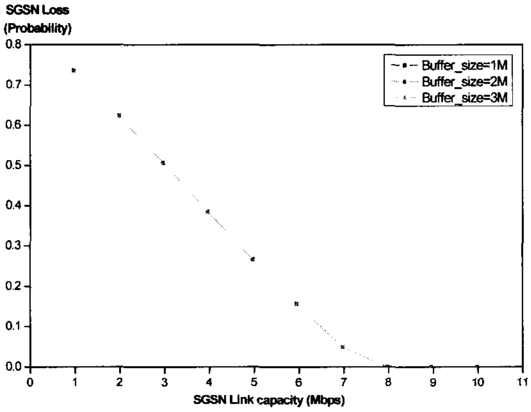


그림 12. 다양한 링크 성능에 따른 SGSN 손실 확률
Fig. 12. SGSN loss probability for various link capacities.

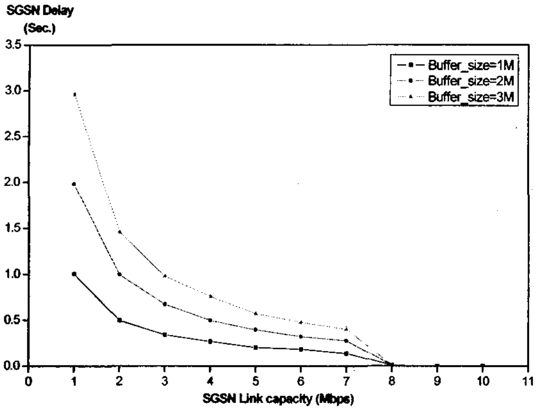


그림 13. 다양한 링크 성능에 따른 SGSN 지연시간
Fig. 13. SGSN latency time for various link capacities.

Fig. 12 and Fig. 13 show the loss probability and latency time at the SGSN queue for various SGSN link capacity when the average arrival bandwidth at GGSN is 3Mbps and c equals 1.7. If we increase the buffer size, the loss probability is not reduced effectively, but latency time increases. Therefore, if we want to enhance the performance, it is required to increase link capacity while keeping limited latency time.

IV. Conclusion

Mobile communications such as IMT-2000 need lots of resources because there are many signaling traffic such as packet call processing and mobility management of mobile users. Furthermore, network must be designed to consider bursty traffic characteristics as the packet data traffic increases. The 3GPP tries to designs UMTS packet core networks consisting of SGSN and GGSN while considering these characteristics and plans to do a trial service.

In this paper, we set up queueing network model using sim++ simulator to evaluate the system performance and obtain useful performance results for SGSNs and GGSNs. For the performance of the signaling traffic, we evaluate average waiting time of SGSN(GGSN) for various service time of each element, user distribution, and average call usage pattern. According to these results, the required processing capability of SGSN should be much higher than the processing capacity of basic packet call arrivals in our suggested structure, since complex signaling procedures are required. Also, SGSN load increases when the number of cells in a RA and the number of RNCs that are connected to a SGSN decrease, because mobility related signaling traffic increases.

We also evaluate the data traffic performance when there are 2 SGSNs and 5 GGSNs in UMTS networks. We use self-similar traffic model because previous Markovian model is not adequate for bursty traffic modeling over the Internet. Performance results show that increasing the link capacity is more effective than increasing the buffer size to satisfy the constraints of packet loss probability and latency time. For more accurate results, it is important to determine c value by actual measurement because traffic burstiness affects the performance evaluation results. We believe that these results can be utilized in the dimensioning of the packet core networks such as link capacity and buffer size when designing UMTS networks.

참고문헌

[1] Lim KeunHee, Youn HyeonSoo, "The prospect and trend of wireless Internet technologies", *Information Communication Technology*, vol. 1, No. 14, pp. 3-15, 1999.
[2] TS23.101, "General UMTS architecture"
[3] TS29.060, "General packet radio service (GPRS); GPRS tunnelling protocol (GTP) across the Gn and Gp interface"
[4] Robert M. Cubert, Paul Fishwick, "Sim++ version1.0 manual", University of Florida 1995.
[5] http://www.cise.ufl.edu/~fishwick/simpack/simpack .html
[6] S. Mohan and R. Jain, "Two user location strategies for personal communications services", *IEEE Personal communications*, pp. 42-50, 1994.
[7] L. Kleinrock, '*Queueing systems, volume 1: Theory*', John Wiley & Sons, Inc., 1975.
[8] GSM 03.60, V5.2.0, "Digital cellular telecommunication system (Phase 2+): general packet radio service (GPRS); Service description (Stage

2)", Dec. 1997.

[9] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic", *ACM Sigcomm '93*, SanFrancisco, CA, pp. 183-193, 1993.

[10] K. H. Park and W. Willinger, '*Self-similar network traffic and performance evaluation*', John Wiley & Sons, Inc., 2000.

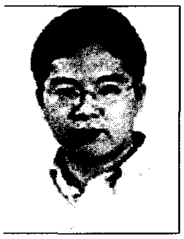
[11] Sheldon M. Ross, '*Introduction to probability models, 6th edition*', Academic Press, 1997.

[12] 3GPP TS 24.008 v3.414.0, "Technical specification group core network; mobile radio interface layer 3 specification; core network protocols Stage 3", 2002-12.

[13] 3GPP TS 25.401 v3.10.0, "Technical specification group radio access network; UTRAN overall description", 2002-06.

[14] Alexander Klemm, Christoph Lindermann, and Marco Lohmann, "Traffic modeling and characterization for UMTS networks", *IEEE Globecom*, 2001.

저 자 소 개



김 병 철(정회원)
1988년 서울대학교 전자공학과 (공학사)
1990년 한국과학기술원 전기 및 전자공학과 (공학석사)
1996년 한국과학기술원 전기 및 전자공학과 (공학박사)
1993년~1999년 삼성전자 CDMA 개발팀
1999년~현재 충남대학교 정보통신공학부 조교수
<주관심분야 : 이동인터넷, 이동통신 네트워크, 데이터통신>

이 재 용(정회원)
1988년 서울대학교 전자공학과 (공학사)
1990년 한국과학기술원 전기 및 전자공학과 (공학석사)
1995년 : 한국과학기술원 전기 및 전자공학과 (공학박사)
1990년~1995년 디지콤 정보통신연구소 선임연구원
1995년~현재 충남대학교 정보통신공학부 부교수
<주관심분야 : 초고속통신, 인터넷, 네트워크 성능분석>