

# UMTS의 Core Network에서 VC Holding Time Policy에 관한 연구

서준배<sup>0</sup> 광용원<sup>\*\*</sup> 김영진<sup>\*\*\*</sup> 이형우<sup>\*</sup> 조충호<sup>\*\*</sup>

<sup>0</sup>고려대학교 전자 및 정보공학부

<sup>\*\*</sup>고려대학교 전산학과

<sup>\*\*\*</sup>한국전자통신연구원

{jibseo.yongwon}@tigerking.korea.ac.kr yikim@amadeu.etri.re.kr {hwlee, chcho}@tiger.korea.ac.kr

## Virtual Circuit Holding Time Policies for UMTS Core Network

June-Bae Seo<sup>0</sup> Yong-Won Kwak<sup>\*\*</sup> Young-Jin Kim<sup>\*\*\*</sup> Hyoung-Woo Lee<sup>\*</sup> Chong-Ho Cho<sup>\*\*</sup>

<sup>0</sup>Dept. of Electronics and Information Engineering, Korea University

<sup>\*\*</sup>Dept. of Computer Science, Korea University

<sup>\*\*\*</sup>ETRI

### 요약

UMTS(Universal Mobile Telecommunication System)의 Core Network에서는 SGSN(Serving GPRS Support Node)과 GGSN(Gateway GPRS Support Node)사이의 전송계층을 IP-Over-ATM network을 기반으로 한다. 이는 비연결형(connectionless) IP 트래픽들을 연결지향형(connection-oriented) ATM 전송계층을 통해 전송함으로써 이때 효율적인 자원관리를 위해 적절한 VC(Virtual Circuit)의 접속과 해제를 수행해야 한다. 본 논문에서는 [1]에서 제안되는 GPRS(General Packet Radio Service)의 Web트래픽 모델에 대하여 기존의 VC의 Holding-time을 결정하는 기법인 Holding Cost Pricing Model(LRU, Mean-Variance, Adaptive policy)[2]을 GGSN과 SGSN사이의 ATM전송계층에 적용하였다. 각각의 기법들에서 VC의 이용률(utilization)과 설정율(setup rate)을 비교함으로써 Adaptive policy기법의 성능이 다른 기법에 비해 효율적임을 알 수 있다.

### 1. 서론

UMTS에는 두 가지 종류의 GSN(GPRS support node)인 GGSN과 SGSN이 있다. [3][4]에서는 GGSN과 SGSN사이의 전송계층을 ATM의 SVC(Switched Virtual Circuit)로써 연결하고 상위계층은 IP망을 기반으로 하고 있다. 따라서, GGSN과 SGSN사이에서 IP-over-ATM network이 형성되는데, 이때 IP트래픽 발생시 SVC설정이 요구되며 이는 processing overhead, setup latency, UNI/PNNI signaling을 유발시키게 된다. IP패킷이 도착하지 않는 동안, 사용하지 않는 SVC의 연결의 해제는 네트워크 자원의 낭비를 예방하고, 할당된 SVC의 이용률을 높일 수 있다.

본 연구에서는 UMTS에서의 Web 트래픽 모델에 대해 기존의 Holding Cost Pricing Model [2]을 적용하였다.

### 2. 관련연구(Holding Cost Pricing Model)

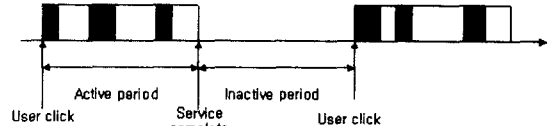
[5]에서는 각 IP 트래픽의 특성에 따라 PVC(Permanent Virtual Circuit), SVC와 Layer3-forwarding을 혼합하여 사용할 것을 제안하였다. ICMP, IGMP, HTTP-request, ACK 등과 같이 짧은 생명주기의 작은 트래픽량을 갖는 플로우(flow)에 대해서는 PVC를 사용하고 Video, Data등과 같이 긴 생명주기의 큰 트래픽량을 가지며 QoS가 요구되는 플로우에 대해서는 SVC를 사용한다.

[6][7]에서는 분석을 통하여 적절한 timeout을 제안하지만 IP계층에서의 스케줄링(scheduling)과 혼잡제어(congestion control)를 고려할 때 각 패킷의 도착간격의 프로세스가 Markovian으로 가정하기는 힘들다는 문제점이 있다.

본 연구에서 GGSN과 SGSN사이의 모든 경로는 SVC를 사용하는 것으로 가정한다[4]. 이는 모든 경로가 PVC로 구성되어 있다면 어떤 클래스(class)나 플로우 혹은 GTP(General Tunneling Protocol) 터널링에서 요구되는 동적인 대역을 할당할 수 없고, 다양한 QoS 요구 사항들을 충족시켜줄 수 없기 때문이다.

#### 2-1. WWW Traffic Model [1]

사용자가 URL 요청을 했을 때, WWW server는 사용자에게 응답을 발생시키는 client program을 실행시키고, URL에 관계된 모든 요청이 완료될 때, 사용자는 다음 요청을 시작하기 전에 정보를 읽는 시간이 필요하다. 따라서 WWW 트래픽을 active와 inactive period로 구성하여 모델링한다. WWW 트래픽의 도착 패턴은 다음 [그림 1]과 같다.



[그림 1]. WWW Traffic model

Active ON(active period)은 사용자의 요청시간부터 서비스가 완료되는 시간을 말하며, 각 파일들의 전송 시간과 두개의 파일을 수신할 때의 짧은 간격인 active OFF time으로 구성된다. Inactive OFF(inactive period)는 사용자가 수신한 정보를 읽는 시간(서비스 완료와 다음 서비스 요청의 시작 시간)을 말하는데, ON period는 downlink상에서 파일의 전송시간을 말하고, OFF period는 데이터 전송이 없는 시간을 말한다.

ON time에서의 파일 크기는 다음의 Pareto 분포를 따른다.

$$F_w(x_w) = 1 - \left(\frac{k_w}{x_w}\right)^a$$

여기서  $k_w$ 는 최소파일크기를 나타낸다.

파일전송 뒤에 따르는 Active OFF의 시간은 다음과 같이 Weibull 분포를 따른다.

$$\Gamma_w(t_w) = 1 - e^{-\left(\frac{t_w}{a}\right)^b}$$

여기서  $t_w$ 는 active off의 시간을 나타낸다.

Active ON period가 끝난 뒤의 Inactive OFF는 다음의 Pareto 분포를 따른다

$$\Gamma_w'(t'_w) = 1 - \left(\frac{k'_w}{t'_w}\right)^{a'}$$

여기서  $t'_w$ 는 Inactive off period의 시간을 말하며,  $k'_w$ 는 최소파일 크기이다.

Active ON에서 파일개수의 분포는 다음의 Pareto 분포를 따른다.

$$F_f(x_f) = 1 - \left(\frac{k_f}{x_f}\right)^{a_f}$$

$k_f$ 는 최소 파일의 개수이다. 따라서 Active ON period에서 각 파일이 개수에 해당하는 만큼 Active OFF가 존재한다. 다음의 [표1]은 본 논문에서 사용한 변수들이다.

[표1] 각 분포의 변수[1]

Active Period	각 파일 개수의 분포	Pareto 분포	$k_f=1$ $a_f=2.43$
	각 파일 크기의 분포	Pareto 분포	$k_w=2730.67$ (bytes) $a_w=1.2$
	Active OFF 분포	Weibull 분포	$a=0.382$ , $b=1.46$
Inactive Period	Inactive OFF 분포	Pareto 분포	$k'_f=1$ $a'_f=1.5$

2-2. Holding Cost Pricing Model [2]

하나의 SVC가 설정될 때마다  $O$  (Open Cost)가 형성되는데 이것은 SVC설정 비용과 사용자가 SVC설정을 기다리는 비용을 포함한다. 다양한  $O$ 의 값을 통하여 만족할만한 성능 대 비용의 tradeoff를 이룰 수 있다. 따라서 각각의 IP패킷이 SVC에 도착한 후에 그 SVC를 얼마나 오랫동안 열어둘 것인가를 결정해야 한다. 이 시간의 길이가 timeout이며, 만일 timeout전에 도착한 패킷이 없다면 그 SVC는 설정을 해제 시켜야 하고, 새로운 IP패킷이 도착될 때 다시 설정 되어야 한다. SVC의 설정을 오랫동안 유지하는 것은 Holding Cost(IP패킷이 없는 동안 SVC설정을 유지할 때의 비용)가 커지는 결과를 초래한다. 각각의 SVC에는 같은 timeout이 적용되며 각 SVC에 도착하는 패킷의 도착에 대한 상관관계(correlation)는 없는 것으로 가정한다.

1) LRU(Least Recently Used) policy: 만일 IP 패킷이  $t$ 시간동안 발생하지 않는다면 다음 IP 패킷의 도착시간은  $ct$  후에 발생할 것이라고 예측한다(여기서  $c$ 는 상수). 상대적으로 예측되는 도착시간은 LRU stack의 상대적인 깊이(depth)에 일치한다. 따라서 이 방식을 적용함으로써 미래의 도착시간을 예측할 수 있다. 만일  $O/c$  ( $O$ 는 time unit)동안 도착하는 패킷이 없다면 SVC 설정을 해제하고 모든 SVC에 같은 timeout인  $O/c$ 를 적용한다.

2) Mean Variance policy: 이 기법은 도착간격의 평균과 표준편차를 EWMA(Exponentially Weighted Moving Average) 방식을 이용하여 추정한다.

$$u_{k+1} = \alpha u_k + (1-\alpha)u_k$$

$$\sigma_{k+1} = \alpha|u_k - t_k| + (1-\alpha)\sigma_k$$

만일  $u-2\sigma > O$  이면 SVC의 설정을 즉각적으로 해제하고  $u+2\sigma < O$  이면 SVC설정을 유지한다. 만일  $u-2\sigma \leq O \leq u+2\sigma$  이고  $u > O$  이면 SVC설정을 해제하고 그렇지 않다면 timeout은  $\min(O, \max(u+2\sigma, C))$ 으로 설정하여 이 기간 만큼 기다린다. (본 논문에서는  $u_0=0, \sigma_0=0, C=3$ 을 사용함)

3) Adaptive policy: 이 기법은 한 IP트래픽 내부에서 다음 패킷의 도착까지의 도착간격은 지금까지 관찰한 도착간격의 분포와 같다고 가정한다. 따라서 비용(Cost)은 다음 패킷이 도착하기 전까지 설정을 유지한 시간과 다음 도착할 패킷을 위해 다시 설정되어야 하는  $O$ 를 더한다. 만일 timeout이  $t$ 로 설정되면 다음 패킷의 기대되는 비용은 아래와 같다.

$$C(t, D) = \int_0^t x f_D(x) dx + (t+O) \int_t^\infty f_D(x) dx$$

여기서  $D$ 는 도착간격의 분포이고  $f_D$ 는  $D$ 의 pdf이다.

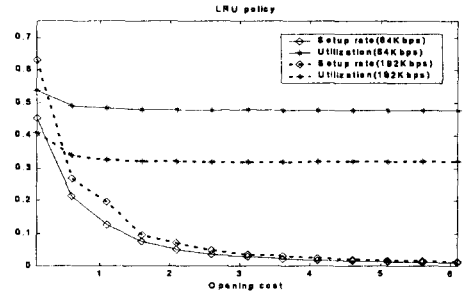
일반적으로  $D$ 를 알 수 없으므로, 도착간격의 관찰을 통하여 Histogram을 형성함으로써 이 Histogram이  $D$ 의 pdf로 근사 되어 질 수 있다. 따라서,  $i \in \{0, m-1\}$  ( $m$ 은 Histogram entries의 개수)에 대하여  $H(i)$ 가  $[i \times M/m, (i+1) \times M/m]$ 에서의 관찰된 도착간격의 수라고 정의하면, ( $M$ 은 최대 timeout)

$$E[C] = \sum_{0 \leq i \leq ((i+1) \times M/m) \leq t} \frac{H(i) \times i \times M}{n(H) \times m} + (t+O) \left( \sum_{i \leq m-1, ((i+1) \times M/m) > t} \frac{H(i)}{n(H)} \right)$$

으로 기대되어지는 Cost는 근사되어 질 수 있다. 여기서  $n(H)$ 는  $H(i)$ 의 총 entries의 합이다. (본 논문에서는  $M=3 \times O, m=30$ 을 사용함)

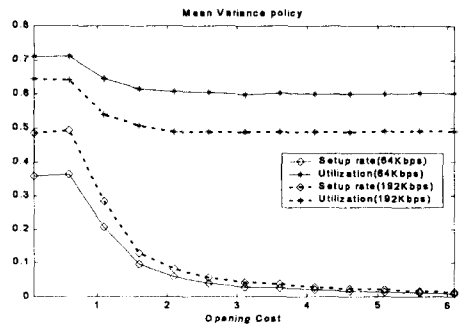
3. 실험결과 및 분석

[그림 2]는 LRU policy기법을 적용했을 때의 결과이다. 전송속도가 낮을 때(64Kbps)는 IP트래픽의 전송시간이 길어짐으로써 이용률이 높아짐을 볼 수 있다. 여러 개의 Web트래픽을 통합(aggregation)하여 전송시간을 증가시키고 여기에 알맞은 용량을 갖는 SVC를 할당하는 것이 SVC의 이용률을 높이고 설정을 낮출 수 있음을 예상할 수 있다.



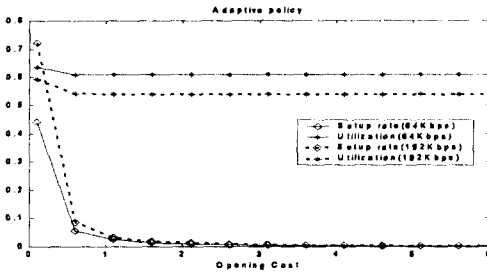
[그림 2]

[그림 3]은 Mean-Variance policy기법을 적용했을 때의 결과이다.  $O$ 가 작을 경우, 이용률이 높게 나오는 것은 예측한 도착간격이  $O$ 보다 크게 예측될 때, 전송이 끝난 시점에서 SVC를 곧바로 해제하기 때문이며, 예측된 도착간격이  $u-2\sigma \leq O \leq u+2\sigma$ 이고  $u < O$ 인 경우 timeout만큼 기다려야 하는 횟수가 적기 때문인데 이는 예측된 평균이 클수록 알 수 있다. 이는 평균과 분산이 큰 Pareto분포가 영향을 미침을 알 수 있다.

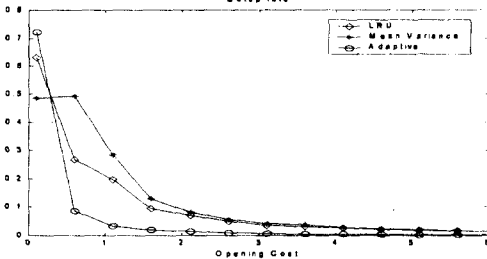


[그림 3]

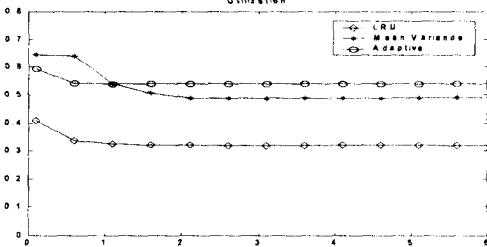
[그림 4]는 Adaptive policy기법에 해당하는 결과이다.  $O$ 가 0.6을 중심으로 설정율이 급격히 떨어짐을 볼 수 있는데 이는 Adaptive policy가 이용률보다  $O$  설정율을 반영하기 때문이다.



[그림4]



[그림5]



[그림6]

[그림 5]와 [그림 6]은 각각의 기법(전송속도:192Kbps)에 대해 설정율과 이용률을 비교한 것이다.  $O$ 가 0.6이하일 때, Mean-variance policy기법이 Adaptive policy기법보다 설정율이 낮고, 이용률이 높음을 볼 수 있는데 이는 Mean-variance policy기법은 전송이 끝난 직후 이전의 예측된 도착간격이  $O$ 보다 클 경우는 기다리지 않고 SVC 설정을 바로 해제하지만, Adaptive policy기법은 항상 histogram으로 예측된 timeout만큼 기다렸다가 SVC의 설정을 해제한다. 따라서 timeout이 다소 과소평가(under-estimate)되어 예측됨 알 수 있다.  $O$ 가 0.6이상일 경우, Adaptive policy기법에서는 설정율이 급격히 감소하는 반면, 이용률을 어느 정도 유지됨을 볼 수 있다.  $O$ 가 1이상인 경우는 Pareto분포의 영향이 크데, Mean-Variance policy기법의 성능의 저하를 볼 수 있다.

4.  $t_H$

패킷의 도착간격  $D$ 의 분포  $f_D$ 는 다음과 같다.

$$f_D(t) = \alpha \times \frac{\partial \Gamma_w(t_w)}{\partial t_w} + (1-\alpha) \times \frac{\partial \Gamma_w'(t_w)}{\partial t_w}$$

$$\alpha = \frac{\frac{a_f}{a_f-1} k_f \left(\frac{a}{b}\right) \Gamma_s(a^{-1})}{\frac{a_f}{a_f-1} k_f \left(\frac{a}{b}\right) \Gamma_s(a^{-1}) + \frac{\alpha'_w}{\alpha'_w-1} k'_w}$$

여기서  $\Gamma_w(\cdot)$ 는 Active OFF의 분포를 나타내는 Weibull

분포의 CDF 이고  $\Gamma_w'(\cdot)$ 는 Inactive OFF의 분포를 나타내는 Pareto 분포의 CDF,  $\Gamma_s(\cdot)$ 는 감마(gamma) 함수이다. 여기서  $\alpha$ 는 전체 Active OFF와 Inactive OFF에서 Active OFF가 차지하는 비중(weight)이 되는데, 파일의 개수의 분포를 나타내는 Pareto분포에서  $k_f$ (최소 전송 파일 개수)와  $a_f$ ( $>1$ , 평균과 분산을 결정하는 변수)에 민감하며 이에 따라, 도착간격의 분포  $f_D$ 는 H(Hurst-parameter)를 0.5에서 1사이의 값을 가진다.

따라서 기대비용을 최소화하는  $t_H$ 의 조건은 다음과 같다.

$$\frac{\partial C(t, D)}{\partial t_H} = 0 \text{ and } \frac{\partial^2 C(t, D)}{\partial t_H^2} > 0, \text{ therefore}$$

$$O \times f_D(t_H) = \int_{t_H}^{\infty} f_D(t) dt \text{ and } O \times f_D'(t_H) + f_D(t_H) > 0$$

5. 결론 및 향후 연구과제

본 연구에서는 UMTS에서의 Web트래픽 모델을 이용하여 기존의 Holding Cost Pricing Model을 적용하여 그 성능을 비교하였다. Mean Variance기법은 도착간격의 예측에 있어 EWMA방식에 기반을 둬므로써  $O$  값에 따라 이용률과 설정율을 적절히 만족시켜주지 못함을 알 수 있었다. 또한 Adaptive policy방식은 Cost함수를 만족시켜주나 각 SVC마다 SVC설정부터 해제까지 histogram을 기반으로 계산해야 함으로 높은 processing부담을 가지고 있다. 또한 LRU방식을 통하여 가급적이면 Web 트래픽들을 통합(aggregation)하여 SVC를 설정하는 것이 SVC의 이용률을 높임을 알 수 있었다. Processing부담을 덜 갖는 SVC설정과 해제에 대한 기법이 연구되어야 할 것이다. 또한  $n$ 개의 같은 용량의 SVC가 있고  $m$ 개의 Web트래픽( $n < m$ )이 있을 때 SVC를 할당하는 Paging Pricing Model[2]에 대해서도 실험되어져야 할 것이다.

6. 참고 문헌

- [1] J.HO, Y.Zhu, S.Madhavapeddy: " Throughput and Buffer Analysis for GSM Packet Radio Service", IEEE INFOCOM' 99
- [2] S.Keshav et al.: " An Empirical Evaluation of Virtual Circuit Holding Time Policies in IP-Over ATM Networks", IEEE JSAC, VOL13, NO., 9, OCTOBER 1995
- [3] TS 23.101 General UMTS Architecture
- [4] TS 23.925 UMTS Core network based ATM Transport [Online Available]www.3gpp.org
- [5] C.Y.METZ: " IP Switching: Protocols and architectures", McGraw-Hill.
- [6] M.HASSAN and M.ATIQUZZAMAN : " A Delayed Vacation Model of an M/G/1 Queue with Setup Time and Its Application to SVCC-Based ATM Networks", IEICE Trans. Comm., VOL E80-B, No2, FEBRUARY 1997
- [7] Z.NIU, Y. TAKAHASHI: " An Extended Queuing Model For SVC-Based IP-OVER-ATM Networks and ITS Analysis", IEEE INFOCOM' 98
- [8] V.Paxson, S.Floyd: " Wide Area Traffic: The Failure of Poisson Modeling", IEEE/ACM Trans., On Network., June, 1995
- [9] W.E.Leland, M.S.Taqqu, D.V.Wilson: " On the Self-Similar Nature of Ehternet Traffic (Extended Version)", IEEE/ACM Trans. On Network., Feb., 1994
- [10] Y.Shu Z.Jin, L.Zhang, L.Wang: " Traffic Prediction Using FARIMA Models", IEEE INFOCOM ' 99