Network에서 트래픽의 self-similar 특성 분석

황인수* , 이동철** , 박기식** , 최삼길* , 김동일*
동의대학교 정보통신공학과*
한국전자통신연구원**

Analysis of self-similar characteristics in the networks

In-soo Hwang* Dong-chul Lee** Ki-shik Park** Sam-gil Choi* Dong-il Kim*
Dongeui University* ETRI**
E-mail : hwangis@hyomin.dongeui.ac.kr

요 약

기존의 트래픽 분석은 포아손 분포나 마르코프 모델로 트래픽을 모델링하고, 패킷의 큐 도착을 지수분포로 가정하였으나, 지난 몇 년간의 연구결과 네트워크 트래픽은 비주기성 및 버스트 특성을 가짐을 알게 되었다. 이러한 트래픽의 특성은 자기유사 특성을 가진 새로운 트래픽 모델로 분석 하므로써, 네트워크의 확장성, QoS, 그리고 최적화된 설계를 가능하게 한다. 본 논문에서는 시뮬레이션 망을 통한 small-scale 에서의 혼재된 트래픽 특성과 실제 WAN의 도착 지연시간, 그리고 전체망의 이용률에 대한 각각의 자기유사특성을 분석 한다.

ABSTRACT

Traffic analysis during past years used the Poisson distribution or Markov model, assuming an exponential distribution of packet queue arrival. Recent studies, however, have shown aperiodic and burst characteristics of network traffics. Such characteristics of data traffic enable the scalability of network, QoS, optimized design, when we analyze new traffic model having a self-similar characteristic. This paper analyzes the self-similar characteristics of a small-scale mixed traffic in a network simulation, the real WAN delay time, TCP packet size, and the total network usage.

Ⅰ. 서 론

오랜기간 네트워크의 용량 설계 및 성능예측에 포아손(Poisson) 패턴에 의한 큐잉분석이 수행되어 왔으나, 인터넷 사용과 그로 인한 멀티미디어 데이터 서비스의 급증은 네트워크 망과 전송장비 발전의 속도를 앞서는 새로운 사용자 트래픽 모델에 대한 분석을 요구한다. 이전까지 network 트래픽의 대부분은 지역(Local)내에서 외부로 전이 되지 않아 전체적인 트래픽율에 영향을 미치 않았으나 트래픽 부하를 외부(WAN)로 전달하므로 전체적인 network의 처리 및 이용율에 영향을 미치게 된다. 이러한 network 환경의 변화는 보다 안정적이고, 서비스의 질을 요구하면서도, 확장성을 만족해야 하는 어려움을 가진다. 따라서 네트워크 설계자나 시스템 분석가는 이러한 트래픽에서의 새로운 패턴을 추출해서, 실제 트래픽 모델과의 근사적인 분포모델을 구성하여 큐 설계나 망의 구성에 반영해야 한다. 본 논문에서는 이러한 비결정적이고, 불규칙적인 트래픽의 버스트 특성에서 규칙성의 패턴을 추출하기 위해 시계열의 확대나 이동에 상관없이 통계적 특성을 유지하는 자기유사성(Self-similarity)을 정의하고, small-scale network 시뮬레이션에서 혼재된 트래픽을 관찰하고, 실제의 WAN 및 LAN에서의 자기유사 특성을 분석 하고자 한다.

Ⅱ. 본 론

2-1. 연속시간 정의

시간에 대한 어떠한 확률 $x(t)$가 있을 때, $a^{-H}x(at)$와 $x(t)$가 통계적으로 같은 성질을 가지고 있다면, $x(t)$는 자기 유사 특성을 가진다. $H(0.5 \le H \le 1)$

이들은 다음의 관계를 가진다.

$$1.\ E[x(t)] = E[x(at)] \qquad \text{Mean}$$

$$2.\ Var[x(t)] = \frac{Var[x(at)]}{a^{2H}} \qquad \text{Variance}$$

$$3.\ R_x(t,s) = \frac{R_x(at,as)}{a^{2H}} \qquad \text{Autocorrelation}$$

Hurst 파라미터 H는 자기 유사성의 특성을 나타내는 척도이다. H의 값을 통해 통계적인 현상의 지속성 및 장기간 의존성(LRD)을 나타낸다. H=0.5는 비 자기 유사 특성을 가지는 기존의 트래픽 모델을 나타내고, 그 값이 1에 가까울수록, 통계적 특성의 지속성 정도 또는 장기간 의존성은 커진다.

2-2. 이산 시간 정의

시계열 x에 대해, m-aggregation 식은 $x^{(m)} = \{x_k^{(m)},\ k=0,1,2,...\}$ 일 때

$$x_k^{(m)} = \frac{1}{m}\sum_{i=km-(m-1)}^{km} x_i \qquad (1)$$

(1)식과 같이 되고, aggregation 과정 이후에도 통계적 특성이 보존된다면, 자기 유사과정이다.

시간평균은 편차 m이 커짐에 따라 자기공분산은 0으로 급격히 감소하게 된다. 그러나 비정상 과정인 자기 유사특성은 정상과정의 에르고딕 보다 천천히 0으로 감소하게 된다.

m=1,2,...에 대한 확률 과정 x는 파라미터 β $(0<\beta<1)$에 대하여 자기 유사특성을 나타낸다.

$$Var(x_{(m)}) = \frac{Var(x)}{m^{\beta}} \qquad \text{Variance}$$

$$R_{x^{(m)}}(k) = R_x(k) \qquad \text{Autocorrelation}$$

β는 Hurst 파라미터 $H=(1-\beta/2)$의 변수이고, 정상과정의 에르고딕 과정에서는 β=1이 된다.

시간 평균의 분산은 1/m 비율로 감소하게 되지만, 자기유사 과정에서 시간평균의 분산은 더욱 천천히 감소($1/m^{\beta}$)하게된다.

확률과정 x에서 충분히 큰 모든 k에 대해 다음과 같다면 근사적으로 자기 유사라고 한다.

$$R_{x^{(m)}}(k) \to R_x(k),\ m \to \infty \qquad \text{Autocorrelation (2)}$$

일반적인 랜덤 과정은 $R^{(m)}(\tau) \to 0,\ as\ m \to \infty$ 일 때 잡음 특성(1/f noise 특성)을 보인다.

위의 결과 m-aggregate 과정에서 m이 무한대의 값을 가지더라도 자기 상관은 일정한 값을 가지게 되고, 타임 스케일의 확대에 상관없이 트래픽의 버스트 특성을 유지하게 된다.

2-3. 장기간 의존성(Long-Range Dependence)

기존의 트래픽 모델의 형태는 시간 t에 따라 자기 공분산($C(k)$)이 지수 적으로 급격히 감소하는 단기간 의존성(Short-range dependence)만을 고려하였고, 상관성은 장기간의 짧은 타임 스케일 동안만 유지된다. $\sum_k C(k)$는 유한한 값을 가지게 된다.

$$C(k) \sim a^{|k|} \quad ,|k| \to \infty,\ 0<a<1 \qquad (3)$$

그러나 장기간 의존성은 $\sum_k C(k)$의 값이 무한대의 값을 가지게 되어 자기 공분산의 그래프는 천천히 감소하게 된다.

$$C(k) \sim |k|^{-\beta} \quad ,\ |k| \to \infty\ ,\ 0<\beta<1 \qquad (4)$$

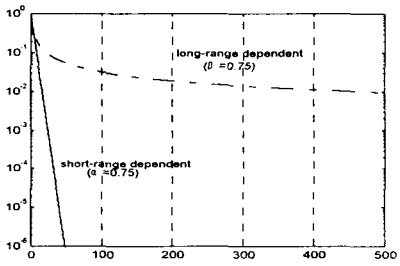


그림 1. 장기간 의존성과 단기간 의존성의 비교

2-4. Heavy-tailed 분포

Heavy-tailed 분포는 패킷의 도착 시간 및 버스트 길이와 같은 트래픽 과정을 설명하는 확률의 특성을 나타내기 위해 사용되어 진다. 랜덤변수 X의 분포가 다음과 같다면 heavy-tailed하다고 한다.

$$1-F(x) = \Pr[X>x] \sim \frac{1}{x^{\alpha}} \quad x \to \infty,\ 0<\alpha \qquad (5)$$

이러한 분포는 분산의 정도가 심하고, 무한대의 값을 가지기도 한다. heavy-tailed 분포의 가장 간단한 형태는 파라미터 k와 α("shape" parameter , $\alpha>0$)의 Pareto 분포이다. 밀도함수 와 분포함수는 다음과 같다.

$$f(x) = \quad F(x) = 0 \quad (x \le k) \qquad (6)$$

$$f(x) = \frac{\alpha}{k}\left(\frac{k}{x}\right)^{\alpha+1}, F(x) = 1-\left(\frac{k}{x}\right)\ (x>k, \alpha>0) \qquad (7)$$

따라서 기대값은 다음과 같다.

$$E[x]=\frac{\alpha}{\alpha-1}k \quad (\alpha>1) \qquad (8)$$

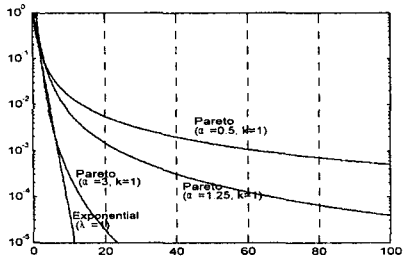


그림 2. Pareto 및 지수 확률 밀도 함수

지수 확률밀도 함수는 거의 직선의 형태로 감소하고, Pareto 확률밀도 함수는 α 값에 따라 지수함수보다 서서히 감소한다.

2-5. 자기유사성의 정도 추정

트래픽이 자기 유사성 정도를 나타내는 방법으로는 Variance-time plot, R/S plot, Periodogram, Whittle's Estimator등이 있다.

Variance-time plot과 R/S plot은 트래픽의 자기 유사성을 분석하고, H의 근사값을 추정하는데 사용된다. Whittle's Estimator법은 어떤 프로세스가 자기 유사하다고 가정하고, 추정구간을 가진 H 값을 구하는데 사용한다.

본 논문에서는 Variance-time plot 방법을 사용하여 β 와 H 값을 구한다.

2-6. Variance-time Plot

매우 큰 시계열 m의 분산식은 다음과 같고,

$$var(x^{(m)}) = \frac{var(x)}{m^{\beta}} \qquad (9)$$

(9)식을 고쳐쓰면, 다음(10)식과 같이 된다.

$$\log[var(x^{(m)})] \sim \log[var(x)-\beta\log(m)] \quad (10)$$

$var(x^{(m)})$을 log-log plot에서 그리게 되면 $-\beta$ 의 기울기를 가진 직선의 그래프를 얻게 된다. -1 보다는 크고 0보다는 적은 기울기의 범위에서 기울기 0에 가까운 값을 갖는 그래프가 자기 유사함을 나타낸다.

III. 트래픽의 측정과 자기유사성 분석

데이터 트래픽의 실제 측정과 시뮬레이션 네트워크에서의 트래픽 분석을 통해서 통계적인 특성과 장시간 시계열에서 장기간 의존성을 갖는 자기 유사특성의 존재를 밝히고자 한다.

3-1. Simulation network Traffic 특성

가상의 시뮬레이션 네트워크를 구성해서 세 곳의 노드중 한곳에서는 heavy-tailed 특성을 가지는 대표적인 트래픽 분포 모델(Out2)인 Pareto 트래픽을 발생시키고, 다른 두 곳의 노드에서는 Exponential 분포(Out0,1)를 가지는 트래픽을 발생 시켰을 때의 시간별 트래픽의 상태를 나타낸다. 시뮬레이션 결과 Pareto 트래픽 모델은 α ($\alpha>0$)인 "Shape" 변수를 가지고 있고, 그 값을 달리 함에 따라, $0<\alpha<1$ 초기 구간에서 급격한 상태 변이가 발생하며, 만일 α가 0.5보다 적은 조건에서는 주어진 시간 조건에서 무한 감소 값을 갖게 되어 시뮬레이션 결과는 무한 루프를 돌게 된다. 그러나 ($\alpha>1$)구간에서는 CBR 트래픽의 초기 천이상태가 안정적이며, 신뢰성 있는 Pareto 분포 모델을 갖는 시뮬레이션을 수행할 수 있다. 아래 그림은 가로축이 시간(60초)이고, 세로축은 전송량 이고, $\alpha=0.9$ 일때의 그래프이다.

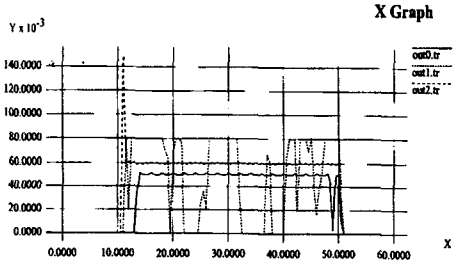


그림 3. Pareto , Exponential Traffic

3-2. Ethernet Traffic

실제의 Network 트래픽을 분석하기 위해 동의대학교 교내에서 4월 15일부터 23일까지 9일간 초 당 한번의 Sample을 측정했다. 대학 전체가 시험기간인 이유로 전체 네트워크 평균이용률은 0.8801%로 낮았으나 버스트 특성을 가지고 있고, 최대/최소 카운트 노드 수는 157/14개이었다

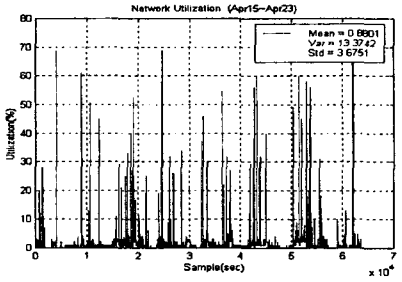


그림 4. Network Utilization Trace

(그림5)는 Ethernet Network traffic 에서의

bytes/frame의 Variance-Time plot이다. V-T plot 방법은 m-aggregation 레벨마다의 분산값을 0과 규준화한 -1 기울기 그래프 상에 나타내고, 그 값들을 Line fitting(Least square Line) 방법에 의하여 음의 기울기인 β 값을 구한 후, Hurst Parameter 값을 계산한다. 그림에서 -1 기울기의 그래프는 H=0.5를 갖는 Poisson 가정의 분포를 나타내고, 기울기 0 ($H=1$)에 가까울수록 실제 트래픽의 특성을 나타낸다. 아래 그림의 H는 0.9711의 장기간 의존성(LRD)을 갖는 Self-similar 특성을 보이고 있다.

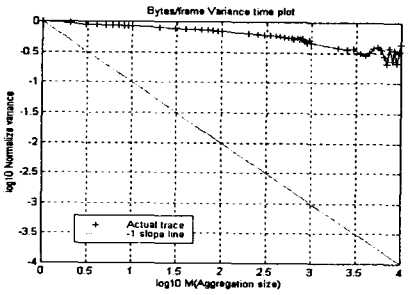


그림 5. Bytes/frame 의 V-T plot

3-3. WAN Traffic

WAN Data traffic을 정확히 측정하기 위해서 WAN과 접속되어 있는 Gateway (SAP : Service Access Point)의 DSU/CSU에서 직접적인 Traffic monitoring이 필요하나, RTT(Round Trip Time)를 측정할 수 있는 Ping 프로그램을 사용해서 구간(Stamp)별 지연 변이와 Drop 율에 의해 간접적인 네트워크의 사용율과 트래픽 상태를 추정할수 있다.

RTT 시간을 측정하기 위해 학교 망의 Subnet에 있는 Linux Platform (203.241.205.114)에서 *www.Altavista.com*까지 경로상의 네곳에 초당 한번씩 약 3만6천 개의 sample data를 Ping 프로그램을 통해서 얻었다.

다음은 RTT를 측정한 네 곳의 주소이다.

- 134.75.225.2 (Dongeui Univ. Gateway)
- 168.126.16.55 (Apagate-gsrl-gel.kornet.net)
- 209.249.192.4 (Main3-249-192.sjc.above.net)
- 204.152.190.11 (www.altavista.com)

첫 번째의 동의대 Gateway는 가장 가까운 2번의 hop을 거친 후, 평균 2ms 이내의 일정한 지연 현상을 가졌고, 이러한 지연 변이의 버스트 특성 감소로 인해 V-T plot은 SRD(Short-range Dependence) 특성을 가지므로 기존의 Poisson 트래픽 모델로 망의 분석이 가능하다.

거리(hop)의 증가에 따른 자기 유사 특성은 더욱 뚜렷이 나타남을 알 수 있다.(그림7 (d),(c))

	203.241.205.254	134.75.226.1	Yahoo	168.126.16.55	209.249.192.4	204.152.190.11
β	0.9666	0.3005	0.0841	0.1645	0.0079	0.008
H	0.5167	0.8498	0.9579	0.9178	0.10039	0.9846

표 1. 각 Stamp의 β와 H

데이터 트래픽의 전송거리(hop)의 증가는 새로운 연구 결과를 보이고 있는데 이전에 조사한 Yahoo(Korea)까지의 V-T plot은 자기유사성을 만족하는 안정된 그래프 형태를 보이나 국가간 Gateway를 벗어날 때 m-aggregation 레벨의 분산의 그래프는 기울기 0 ($H=1$)의 값을 초과하게 되는 결과를 보인다. 또한 전송 거리가 아주 가깝다면 그림(a) 처럼 -1의 기울기 ($H=0.5$)를 갖는 그래프에 거의 일치하게 되는 현상을 보인다. 이러한 전송 거리(hop 수)에 따른 V-T plot의 분석은 트래픽의 자기 유사성 특성의 적용을 제한하게 된다. 즉 너무 먼 거리까지(국가간의 링크)의 트래픽의 통계적 분석은 주된 원인 이외의 여러 가지의 변화 요인에 의해 분석의 가치를 낮게 할 수 있음을 알리는 것이다.

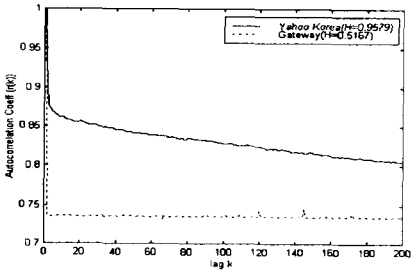


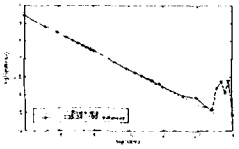
그림 6. Yahoo Korea와 Gateway 의 자기 상관 함수

IV. 결 론

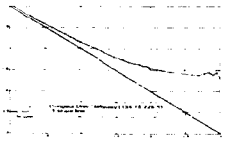
인터넷 트래픽의 증가로 인해 네트워크의 물리적인 구성과 트래픽 관리의 기술은 점차 서비스의 Quality와 확장성을 요구하는 거대한 하나의 단일 가상 네트워크를 구성하게 한다.

이러한 외적인 변화 요인은 네트워크의 성능평가 및 용량산출 또는 네트워크의 설계를 위해 기존의 트래픽 모델로써 네트워크의 통계적인 특성을 파악하는 것이 단편의 메시지나 음성과 영상이 혼재하는 멀티미디어 데이터를 전송하는 트래픽에서는 올바른 분석 모델이 아님을 알게 되었다.

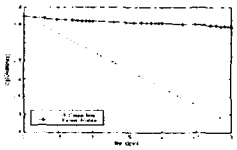
인터넷에서 지연은 처리, 큐잉, 전송, 전달지연의 4가지 요소로 구성된다. 이러한 지연 요인 중 대부분 라우터 내부의 큐 성능에 의해 결정되고, 각 링크 노드의 버퍼 용량은 한계를 가지고 있기



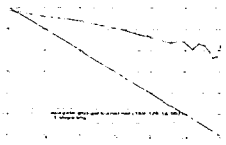
(a) 203.241.205.254



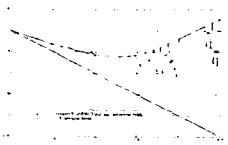
(b) 동의대 Gateway



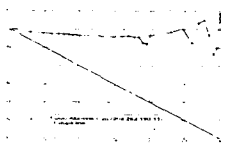
(c) Yahoo korea



(d) 168.126.16.55



(e) 209.249.192.4



(e) www.altavista.com

그림 7. 각 구간에서의 V-T Plot

때문에 SRD(단기간 의존성)기반의 트래픽 모델(Poisson, 지수분포)로써 평균 패킷 전송 지연이나 최적의 큐의 크기로 설계한 네트워크는 어떠한 시점에서 트래픽의 증가로 인한 큐의 정체 또는 폭주가 발생하게 되었을 때 LRD(장기간 의존성) 특성을 과소 평가한 큐에서는 과도한 트래픽을 처리하지 못하고, 입력 패킷의 폐기 또는 재전송 요구로 인해 결과적으로 망 전체의 트래픽을 다시 증가시키는 결과를 초래하게 된다.

이러한 시계열에서의 트래픽의 자기유사 특성과 LRD를 반영하기 위해 적절한 트래픽 패턴을 분석한 모델을 사용해서 네트워크의 용량을 산출하고, 최적의 큐 값을 구해야 한다.

본 논문의 연구 결과 전송 거리(hop)에 따른 트래픽의 특성은 각기 다른 트래픽 모델의 적용을 제안하고 있다. 차후 트래픽의 자기 유사성의 적용 범위에 대한 연구와 실제 트래픽에서의 자기 유사성 개념을 도입한 큐의 성능분석이 요구된다.

참고문헌

[1] W. Richard Stevens. "TCP/IP Illustrated, volume1", Addison Wesley, 1994

[2] William Stallings, "High-Speed Networks", Prentice Hall, pp181-208

[3] Comer,"Internetworking with TCP/IP", Volume I , Prentice Hall,pp517-556

[4] Peyton Z. Peebles, Jr. Probability, "Random Variables, and Random Signal Principles", McGraw-Hill, pp134-198, 1993

[5] Qiong Li, David L. Mills, "On the Long-range Dependence of Packet Round-trip Delays in the Internet. IEEE International Conference on Communications - Volume 2", pp1185-1191,1998

[6] Walter Willinger, Murrad S Taqqu, Robert sherman, Daniel V. Wilson, 'Self-similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level', ACM/Sigcomm'95

[7] Kevin Fall, 'Network Emulation in the Vint/NS Simulation', Proceedings of the 4th IEEE Symposium on Computers Communications 1999

[8] Kevin Fall, Kannan Varadhan,"ns Notes and Documentation[1]", The VINT Project,1999

[9] http://www.ots.utexas.edu/ethernet, "Quick Reference Guide to 10Mbps Ethernet".

[10] http://www-mash.cs.berkeley.edu/ns

[11] 이병헌, "네트워크의 성능 및 관리", 주간기술 동향, pp769,772-773호 ETRI.

[12] 강신각,최선완,오행석,"네트워크 시뮬레이터 NS-2", 주간기술동향, ETRI

[13] 이기주,"확률 및 통계",ohm,1990

[14] 황인수, 김동일 외,' Round-Trip Time 측정을 통한 인터넷 트래픽의 자기 유사성 분석', 추계해양정보통신학회,p326-330, 1999

[15] 김창호, 김동일 외, '트래픽에서의 장기간 의존성 및 Self-similar 특성', 하계통신학회, pp463-467, 1999