

# Round-Trip Time 측정을 통한 인터넷 트래픽의 자기 유사성 분석

황인수\*, 송기평\*\*, 이동철\*\*, 박기식\*\*, 김창호\*, 김동일\*, 최삼길\*

동의대학교 전자통신공학과\*

한국전자통신연구원\*\*

Analysis of internet self-similar traffic through the round-trip time measurement

In-soo Hwang\*, Ki-pyong Song\*\*, Dong-chul Lee\*\*, Ki-sik Park\*\*

Chang-ho Kim\*, Dong-il Kim\*, Sam-gil Choi\*

Donggeui University\*, ETRI\*\*

E-mail : hwangis@hyomin.donggeui.ac.kr

## 요 약

인터넷의 민감한 트래픽 특성으로 인해 기존의 트래픽 모델링 분석법으로 최적화된 네트워크 환경을 구성하기에는 부족한 점이 많다. 본 논문에서는 트래픽의 버스트 특성을 정확히 예측하고 모델링 하기 위한 방법으로 자기 유사 특성에 대해 분석하고자 한다. 실제 인터넷 네트워크에서의 RTT(Round-Trip Time)를 측정함으로써 계층, 거리별 링크간의 LRD(Long-Range Dependence) 와 노드 큐의 특성, 자기유사성 정도를 구하고 측정된 데이터의 확률 분포를 통해 실제 트래픽의 특성에 대해 분석하였다.

## 1. 서 론

인터넷 네트워크의 불규칙적인 트래픽에서 규칙성의 패턴을 추출하고 새로운 트래픽 모델을 만들어 보다 합리적이고, 최적화 된 네트워크의 설계 및 구현을 위해 기존의 포아송(Poisson) 모델이나 마르코브(Markov)모델로 트래픽을 장기간 관찰하는 경우에 나타나는 버스트 특성의 소멸은 트래픽 모델로서 적절치 못하다는 것이 최근의 실험적 논문들에서 밝혀졌다. 이러한 비정상 과정의 트래픽 모델을 설명하는 자기 유사특성의 분석을 통해 실제 트래픽의 정확한 모델링 또는 점근적 근사 모델을 구성할 수 있다. 본 논문에서는 이러한 실제 트래픽의 자기 유사성을 관찰하기 위해 다섯 곳의 Round-Trip Delay Time(이하 RTD)을 측정함으로 LAN과 WAN 양쪽 어디에서나 일어나는 공통적인 트래픽 특성인 Long-Range Dependence(이하 LRD) 특성을 증명하고자 한다.

## II. 본 론

### 1. 자기 유사 트래픽 정의

### 1.1 연속시간 정의

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

이들은 다음의 관계를 가진다.

1.  $E[x(t)] = E[x(at)]$                       평균
2.  $Var[x(t)] = \frac{Var[x(at)]}{a^{2H}}$                       분산
3.  $R_x(t, s) = \frac{R_x(at, as)}{a^{2H}}$                       자기 상관

Hurst 파라미터 H는 얼마나 자기 유사성의 정도를 나타내는 척도이다. H는 통계적인 현상의 지속성과 확률과정의 장기간 의존성을 나타낸다. H=0.5는 기존의 트래픽 모델에서 고려된 자기 유사성의 부재를 나타내고, 그 값이 1에 가까울수록, 지속성의 정도 또는 장기간 의존성은 커진다.

### 1.2 이산 시간 정의

M-aggregated 시계열  $x$ 에 대해,  $x^{(m)} = \{x_k^{(m)}, k=0,1,2,\dots\}$  일 때

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

가 되고, aggregation 과정 이후에도 통계적 특성이 보존된다면, 자기 유사과정이다.  $x^{(m)}$ 의 에르고딕 과정은 시간평균이 조화평균과 같고, 시간평균의 분산은 조화평균의 분산과 동일하다. 시간평균의 편차  $m$ 이 커짐에 따라 분산은 0으로 빨리 접근하게 된다. 그러나 자기 유사특성은 정상 에르고딕 과정이 아니라 비정상 과정이므로 정상과정의 에르고딕 과정 보다 매우 천천히 0으로 수렴하게 된다. 모든  $m=1,2,\dots$ 에 대한 확률 과정  $x$ 는 파라미터  $\beta$  ( $0 < \beta < 1$ )에 대하여 다음을 만족한다면, 자기 유사하다고 한다.

$$\begin{aligned} Var(x_{(m)}) &= \frac{Var(x)}{m^\beta} && \text{분산} \\ R_x^{(m)}(k) &= R_x(k) && \text{자기상관} \end{aligned}$$

$\beta$ 는 앞에 정의했던 Hurst 파라미터  $H = (1 - \beta/2)$ 와 관련이 있고, 정상과정 및 에르고딕 과정에서는  $\beta=1$ 이고, 시간 평균의 분산은  $1/m$  비율로 감소하게 되지만, 자기유사 과정에서 시간평균의 분산은 더욱 천천히  $1/m^\beta$  비율로 감소하게 된다.

확률과정  $x$ 가 충분히 큰 모든  $k$ 에 대해 다음과 같다면 근사적으로 자기 유사하다고 한다.  
 $R_x^{(m)}(k) \rightarrow R_x(k), \text{ as } m \rightarrow \infty$  자기 상관

$R^{(m)}(\tau) \rightarrow 0, \text{ as } m \rightarrow \infty$  일 때는 일반적인 정상 과정의 특성을 나타내고, 백색 잡음 특성을 보인다.

위의 결과를 볼 때 자기 유사의 M-aggregate 과정에서  $m$ 이 무한대의 값을 가지더라도 자기상관은 일정한 값을 가지게 되고, 타임 스케일의 확대에 상관없이 트래픽의 버스트 특성을 유지하게 된다.

### 1.3 장기간 의존성(Long-Range Dependence)

기존의 트래픽 모델의 형태는 시간  $k$ 에 따라 자기 공분산 ( $C(k)$ )이 지수 적으로 급격히 감소하는 단기간 의존성(Short-range dependence)만을 고려하였고, 이 경우  $\sum_k C(k)$ 는 유한한 값을 가지게 된다.

$$C(k) \sim a^{k^\alpha}, |k| \rightarrow \infty, 0 < \alpha < 1 \quad (2)$$

이와는 반대로, 장기간 의존성은  $\sum_k C(k)$ 의 값이 무한대의 값을 가지게 되며 자기 공분산의 그래프는 쌍곡선의 형태로 천천히 감소하게 된다.

$$C(k) \sim |k|^{-\beta}, |k| \rightarrow \infty, 0 < \beta < 1 \quad (3)$$

여기서  $\beta$ 는 Hurst 파라미터( $H=1-\beta/2$ )에 정의되었고, 따라서 장기간 의존성은 자기 유사성을 나타내고, 얼마나 지속성을 유지하는가를 나타낸다.

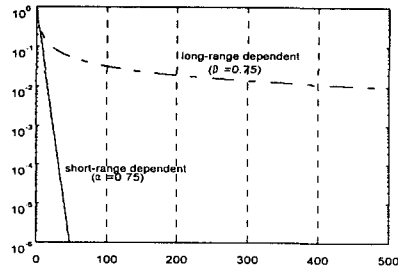


그림 1. 장기간 의존성과 단기간 의존성의 비교

### 1.4 Power Spectral Density(PSD)

장기간 의존성의 정도를 나타내는 다른 방법은 주파수 영역에서 psd를 구하는 것이다. psd의 특성은 다음과 같다.

$$\begin{aligned} S(\omega) &\sim \frac{1}{|\omega|^\gamma} \text{ as } \omega \rightarrow 0, 0 < \gamma < 1 \quad (4) \\ \gamma &= 1 - \beta = 2H - 1 \end{aligned}$$

장기간 의존성은  $\omega$ 가 0으로 가까이 갈 때 무한대의 값을 가지게 되나, 단기간 의존성은( $H=0.5, \gamma=0$ ) 전 대역에 걸쳐서 골고루 분포된 백색잡음과 같은 특성을 가진다.

$$s(\omega) = \sum_{k=-\infty}^{\infty} R(k) e^{-ik\omega}, S(0) = \sum_{k=-\infty}^{\infty} R(k) \quad (5)$$

$S(0) = \infty$ 가 되는 것은  $k$ 가 커짐에 따라 자기상관 함수  $R(k)$ 가 급격하게 감소하지 않고 매우 천천히 감소하는 특성을 가지기 때문이다.

### 1.5 Heavy-tailed 분포

heavy-tailed 분포는 패킷의 도착(interarrival) 시간 및 버스트 길이와 같은 트래픽 과정을 설명하는 확률의 특성을 나타내기 위해 사용되어 질 수 있다. 랜덤변수  $X$ 의 분포가 다음과 같다면 heavy-tailed하다고 한다.

$$1 - F(x) = \Pr[X > x] \sim \frac{1}{x^\alpha} \text{ as } x \rightarrow \infty, 0 < \alpha \quad (6)$$

이러한 분포는 분산의 정도가 심하고, 무한대의 값을 가지기도 한다. heavy-tailed 분포의 가장 간단한 형태는 파라미터  $k$ 와  $\alpha$  ( $k, \alpha > 0$ )를 갖는 Pareto 분포이다. 이것의 밀도함수와 분포함수는 다음과 같다.

$$f(x) = F(x) = 0 \quad (x \leq k) \quad (7)$$

$$f(x) = \frac{\alpha}{k} \left(\frac{k}{x}\right)^{\alpha+1}, F(x) = 1 - \left(\frac{k}{x}\right)^\alpha \quad (x > k, \alpha > 0) \quad (8)$$

그리고 기대값은 다음과 같다.

$$E[x] = \frac{k}{\alpha - 1} \quad (\alpha > 1) \quad (9)$$

지수밀도 PDF(Probability Density Function)는 거의 직선의 형태로 나타나고, Pareto PDF는 지수함수보다 서서히 감소한다.

### 1.6 측정된 트래픽의 자기유사성 분석

측정된 트래픽이 얼마나 자기 유사성을 가지는가를 나타내는 방법에는 Variance-time plot, R/S plot, Periodogram, Whittle's Estimator 등이 있다. Variance-time plot과 R/S plot은 트래픽의 자기 유사성을 분석하고, H의 근사값을 추정하는데 사용된다. 또한 Whittle's Estimator법은 어떤 확률 과정이 자기 유사하다고 가정하고, 신뢰구간을 가진 H값을 구하는데 사용한다.

본 논문에서는 Variance-time plot 방법을 사용하여 측정된 RTD 시간의 Hurst 파라미터를 구하였다.

자기 유사한 시계열의 경우  $m$ 이 매우 클 때의 분산은 다음 식을 따른다.

$$\text{var}(x^{(m)}) = \frac{\text{var}(x)}{m^\beta} \quad (10)$$

위 식을 고쳐 쓰면,

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

이 되고, aggregation 레벨  $m$ 에 대한  $\text{var}(x^{(m)})$ 을 log-log plot에서 그리게 되면  $-\beta$ 의 기울기를 가진 직선의 그래프를 얻게 된다. 이 그래프의 기울기가 0에 가까울 수록 자기 유사함을 나타내게 된다.

## 2. 패킷 RTD 과정의 정의와 측정

### 2.1 패킷 RTD 정의

트래픽의 특성을 정확히 관찰하기 위해서는 장기간 안정된 상태에서 가능한 여러 경로의 구간에서 측정된 결과 값이 요구된다.

이러한 패킷 RTD의 측정과 그것의 장기간 의존성을 분석함으로써 네트워크 성능지수에 영향을 미치는 지연시간과 노드의 상태를 알 수 있고,

정확한 네트워크 트래픽 설계를 가능케 하는 중요한 척도가 된다. 최근의 연구결과 LRD 특성은 기존의 지수 분포 트래픽 모델의 자기 상관함수보다 천천히 감소하여 네트워크 내에서의 평균적인 지연보다 훨씬 더 긴 지연을 발생시키고, 큐에서 트래픽의 버스트 함을 적절하게 처리(쓰고/지우는)하지 못하게 되어 전체 네트워크의 성능에 영향을 미치는 현상을 초래하게 된다. 이와 같은 지연 분포의 heavy-tailed 현상은 패킷의 도착 시간, 버스트 byte 크기, www에서의 파일사이즈, 그리고 유닉스 파일 시스템과 같은 네트워크 트래픽에 연관된 곳곳에서 발견된다.

따라서, 패킷 RTD 시간을 측정하여 LRD 특성을 분석함으로써 다음의 결과를 알 수 있게 된다.

- 네트워크 처리율
- 패킷 손실비
- 응답시간
- 버퍼 점유 시간

최근의 멀티미디어 데이터 및 어플리케이션 들은 인터넷을 경유하게 되고, 인터넷 내에서의 패킷 지연에 민감하다. 인터넷 프로토콜 중 가장 널리 사용하고 있는 TCP/IP에서 TCP는 신뢰성 있는 전송을 위해 주의 깊게 조정해야 하는 timeout 매개변수를 가지고 있고, 그중 패킷 RTD에 의해 동적으로 결정되는 RTO(Re-Transmission Timeout)가 가장 민감하다. 네트워크 내에 패킷 손실이 발생했을 때 RTO는 재전송을 지시하게 되고, 그 결과 end-to-end 간에 TCP/IP 프로토콜의 행동에 영향을 미친다.

본 논문에서는 네트워크 시간 프로토콜(NTP : Network Time Protocol) 제어 메시지 패킷에 의해 timestamp(위치/구간)에서 수집된 지연시간을 구하고, 네트워크 상태를 관찰하여 트래픽의 LRD 특성을 분석하고자 한다.

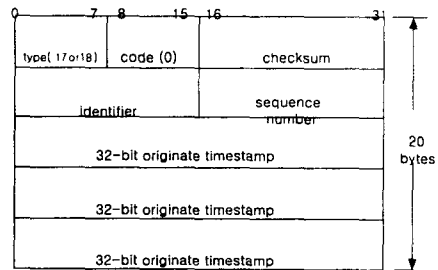


그림 2. ICMP timestamp request and reply message [1]

### 2.2 RTD 측정

RTD 시간을 측정하기 위해 학교 망의 한

Subnet에 위치한 Linux Platform (203.241.199.47) 으로부터 Yahoo Korea까지의 경로 상에 다섯 stamp를 두고, 1초당 한번씩 총 11만 번(패킷 손실을 고려하여)의 Sample을 Ping 프로그램을 사용하여 데이터를 얻었다.

다음은 측정된 다섯 곳의 IP 이다.

- 203.241.199.254 (자연과학대 gateway)
- 134.75.226.2 (동의대 DNS 서버)
- 203.241.174.25 (동명정보대)
- 134.75.3.2 (Seoul-Tageon-T3.Kreonet.re.kr)
- 164.124.96.131 (www.yahoo.co.kr)

### 3. 통계적 분석

#### 3.1 패킷 RTD의 LRD 특성 관찰

시계열에서 Long-range 의존성의 정도를 측정하는 방법중 간단하고, 사용이 많은 방법인 Variance-Time Plot을 통해 다섯 곳의 stamp에서 Hurst 파라미터를 구하고, Variance-Time Plot을 그렸다.

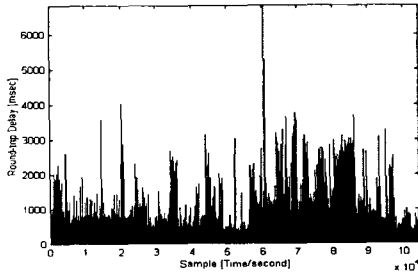


그림 3. Yahoo Korea에서의 RTD 시간 Trace

그림3은 Yahoo Korea로 패킷 발송을 했을 때 경로 상의 지연시간에 대한 응답결과를 나타낸다. 위 그림은 데이터 트래픽의 일반적인 형태와 유사한 버스트 특성을 가지고 있다.

각 측정 stamp의 RTD 시간 데이터는 평균적으로 약 4%의 패킷 손실율을 보이고, 위치가 가까운 곳에 있는 gateway와 DNS 는 예측 가능한 결과(약 1%의 패킷 손실율)를 가졌다. 측정 시점에서 자연대 Subnet의 네트워크 평균 이용율은 3~5%이었고, 트래픽 측정에 영향을 끼칠 수 있는 Linux Platform의 다운이나 네트워크의 과부하 현상은 발생하지 않아서 측정 데이터는 신뢰성을 가진다고 할 수 있다.[5]

그림5는 측정된 5개 경로의 데이터에 대한 Variance-Time Plot을 그린 것이다. 각 그래프는

비교를 위해 -1 기울기를 갖는 참조선(접선)을 가지고 있고, 측정된 데이터의 기울기는 근사공식(Least Square Line방법)으로 fitting하여 쉽게 H 값을 알 수 있다.

	Yahoo	134.75.3.2	동명정보대	DNS	Gateway
$\beta$	0.0841	0.1517	0.1567	0.5225	0.9666
H	0.9579	0.9242	0.9216	0.7387	0.5167

표 1. 각 Stamp의  $\beta$  와 H

그림 5로부터 거리가 먼 곳일수록 (지연시간이 클수록) 높은 H를 가지고, 가까운 거리일수록(지연시간이 작을수록) 트래픽의 버스트함(변이성)이 적어지는 결과가 나왔다.

또한 그림5를 통해 Yahoo Korea와 서울-대전 간 T3 라인의 라우터에서 높은 자기유사성을 가지고, 가까운 거리의 Gateway나 DNS 같은 경우는 낮은 자기 유사성을 가진다.

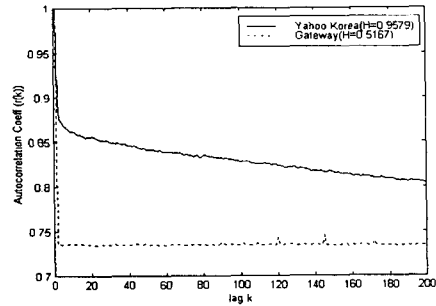


그림 4. Yahoo Korea와 Gateway에 대한 자기 상관 함수의 비교

그림 4는 시차  $k$ 에 따른 Yahoo Korea와 Gateway의 자기 상관함수를 나타내고, 그 결과, 자기 유사성이 클수록 높은 상관성(LRD)을 가지고, 자기 유사성의 정도가 낮을수록 낮은 상관성(SRD) 특성을 가지는 것을 알 수 있다.

인터넷에서 패킷 지연은 처리 지연, 큐잉 지연, 전송 지연, 전달 지연의 4가지 요소로 구성된다.[7] 이러한 패킷 RTD 지연 과정은 대부분 라우터 내부의 큐 성능에 의해 결정되고, 각 링크 노드의 버퍼 용량은 한계를 가지고 있기 때문에 Short-range 기반의 트래픽 모델로써 큐의 크기를 설계한 네트워크는 어떠한 시간에 큐의 정체 또는 폭주가 발생하게 된다. 이러한 버스트 특성을 반영하기 위해 자기 상관 계수가 천천히 감소하는 현상(LRD 특성)을 네트워크 설계시 고려하여

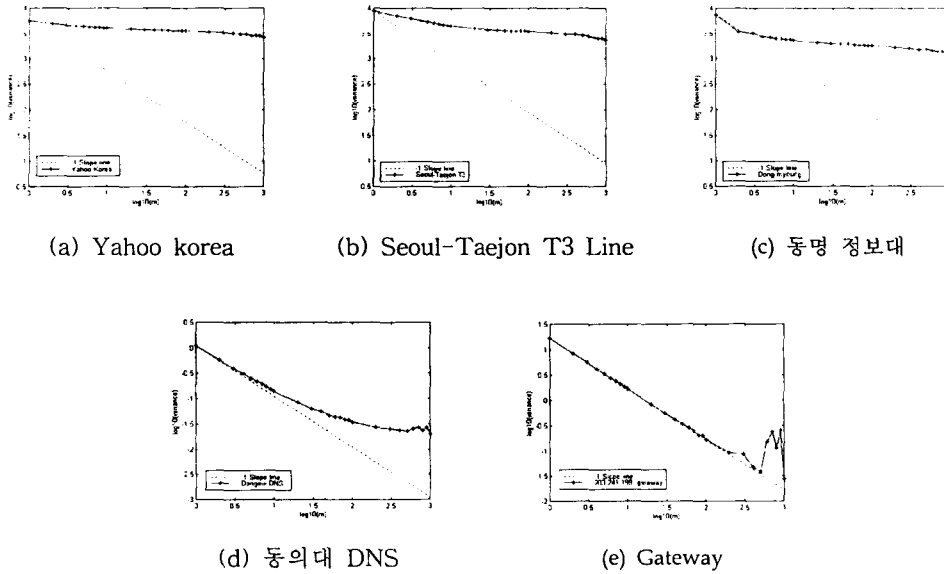


그림 5. 각 Stamp에서의 V-T plot

야 한다.

### III. 결론

인터넷은 다양한 형태의 트래픽이 결합된 거대한 네트워크이다. 대부분의 응용 프로그램이나 멀티미디어 데이터들(음성, 정지영상, 동영상)은 QoS(Quality of Service)를 요구하는 높은 트래픽을 발생함으로써 인터넷 성능에 영향을 미친다.

본 논문에서는 이러한 네트워크의 지연 특성을 RTD 시간 데이터로 측정하였고, 측정된 데이터의 통계적 분석을 통해 가까운 거리의 패킷RTD 시간을 제외한 3곳의 RTD 시간은 모두 자기 유사성을 가지고, 장기간 의존 특성(LRD)이 있음을 알 수 있었다. 최근 LAN 과 WAN 트래픽에서 발견된 자기 유사성과 LRD 특성은 네트워크 엔지니어가 성능 평가나 망 설계 시 고려해야 하는 중요한 요소가 될 것이다.

차후 과제로는 이러한 인터넷의 지연특성 및 큐에서의 지연특성을 정확히 모델링 할 수 있는 방법에 대한 연구가 요구된다.

### 참고 문헌

[1] W. Richard Stevens. "TCP/IP Illustrated,

volume1", Addison Wesley, 1994  
 [2] William Stallings, "High-Speed Networks", Prentice Hall, pp181-208  
 [3] Peyton Z. Peebles, Jr. Probability, "Random Variables, and Random Signal Principles", McGraw-Hill, pp134-198, 1993  
 [4] 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  
 [5] 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  
 [6] <http://www.ots.utexas.edu/ethernet/>, "Quick Reference Guide to 10Mbps Ethernet".  
 [7] 이병현, "네트워크의 성능 및 관리", 주간기술동향, pp769,772-773호 ETRI.  
 [8] 김창호, 김동일 외, '데이터 트래픽에서의 Self-similar 특성', 춘계해양정보통신학회, pp146-151,1999  
 [9] 김창호, 김동일 외, '트래픽에서의 장기간 의존성 및 Self-similar 특성', 하계통신학회, pp463-467, 1999