

# Network 최적 설계를 위한 네트워크 트래픽의 self-similar 특성 분석에 관한 연구

이동철\* , 김창호\*\* , 황인수\*\*\* , 김동일\*\*\*

한국전자통신연구원\*

LG전자 · 정보통신\*\*

동의대학교 정보통신공학과\*\*\*

## A study about analysis of self-similar characteristics for the optimized design networks

Dong-chul Lee\* , Chang-ho Kim\*\* , In-soo Hwang\*\*\* Dong-il Kim\*\*\*

ETRI.PEC\* , LG Electronics Corp.\*\* , Dongeui University\*\*\*

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

### 요 약

최근 인터넷 사용의 급증은 전체적인 망 이용율의 증가를 야기시켜 트래픽의 증가 원인이 된다. 이러한 트래픽 분석을 통해 통계적인 특성을 갖는 장치의 설계와 배치가 요구된다. 따라서, 이러한 인터넷 트래픽의 자기유사성을 분석하고, Simulation 과정을 통해 최적화된 설계요소를 찾아서 실제 네트워크의 성능향상을 연구하고자 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 network Traffic.

### 1. 서 론

지금까지 네트워크의 용량 설계 및 성능예측에 포아송(Poisson) 패턴에 의한 Exponential 도착 분포 모델을 따르는 큐잉분석이 수행되어 왔다. 그러나 인터넷 사용인구의 증가와 멀티미디어 데이터 서비스 현실적 요구는 네트워크 망 장비와 전송서비스의 발전 속도를 앞지르고 있다. 따라서, 새로운 네트워크 트래픽 환경에 적합한 트래픽 모델의 통계적 분석이 요구된다.

그 결과, 네트워크 설계자나 시스템 분석가는 이러한 트래픽에서의 새로운 도착 패턴의 통계적 특성을 추출해서, 실제 트래픽 모델과의 근사한 도착 분포모델에 따른 큐 설계나 망의 구성이 이루어져야 한다. 본 논문에서는 이러한 비결정적이고, 불규칙적인 트래픽의 버스트 특성에서 규칙성의 패턴을 추출하기 위해 시계열의 확대나 이동에 상관없이 통계적 특성을 유지하는 자기유사성(Self-similarity)의 이론적인 성격을 정의하고, 네트워크 시뮬레이션을 통해 Self-similar 특성을 갖는 트래픽 인가하여 실제 네트워크에서 일어나

는 현상을 관찰, 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.  $H(0.5 \leq H \leq 1)$

이들은 다음의 관계를 가진다.

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

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

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

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

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

(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.

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

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

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

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

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

확률과정  $x$ 에서 충분히 큰 모든  $k$ 에 대해 다음과 같다면 근사적으로 자기 유사라고 한다.

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

일반적인 랜덤 과정은  $R^{(m)}(\tau) \rightarrow 0$ , as  $m \rightarrow \infty$  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.

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

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

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

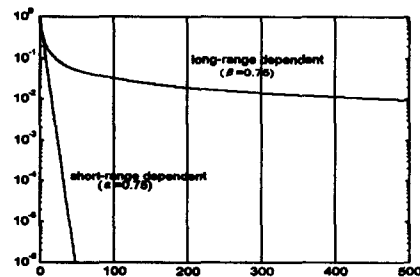


그림 1. 장기간 의존성과 단기간 의존성의 비교

### 2-4. Heavy-tailed 분포

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

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

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

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

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

따라서 기대값은 다음과 같다.

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

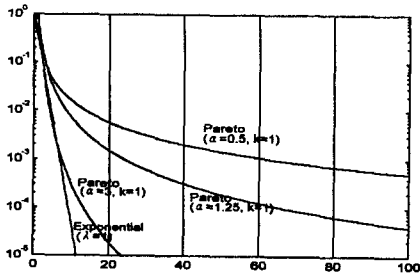


그림 2. Pareto 및 지수 확률 밀도 함수

지수 확률밀도 함수는 거의 직선의 형태로 감소하고, Pareto 확률밀도 함수는  $\alpha$  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 Estimator 등이 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  $\beta$  와 H 값을 구한다.

### 2-6. Variance-time Plot

매우 큰 시계열 m의 분산식은 다음과 같고,

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

(9)식을 고쳐쓰면, 다음(10)식과 같이 된다.

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

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

### III. 트래픽의 측정과 자기유사성 분석

데이터 트래픽의 실제 측정과 시뮬레이션 네트

워크에서의 트래픽 분석을 통해서 통계적인 특성과 장시간 시계열에서 장기간 의존성을 갖는 자기 유사특성의 존재를 밝히고자 한다.

### 3-1. Ethernet Traffic

실제의 Network 트래픽을 분석하기 위해 동의 대학교 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

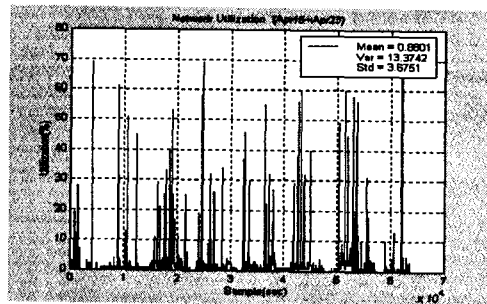


그림 3. Network Utilization Trace

(그림5)는 Ethernet Network traffic에서의 bytes/frame의 Variance-Time plot이다. V-T plot 방법은 m-aggregation 레벨마다의 분산 값을 0과 표준화한 -1 기울기 그래프 상에 나타내고, 그 값을 Line fitting(Least square Line) 방법에 의하여 음의 기울기인  $\beta$  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.

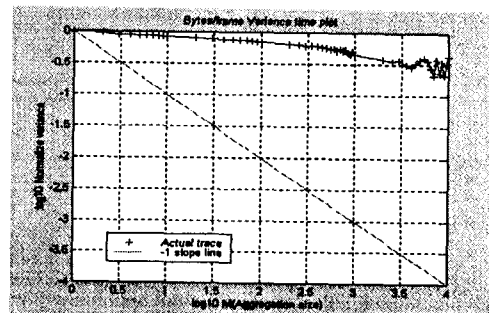


그림 4. Bytes/frame 의 V-T plot

### 3-2. Simulation network Traffic 분석

시뮬레이션 네트워크를 구성하고, 각 노드에 서 heavy-tailed 특성을 가지는 대표적인 트래픽

분포 모델인 Pareto 도착 분포를 갖는 트래픽을 발생시켰을 때, Shape 변수( $1 \leq \alpha \leq 2$ )의 값에 따른 시간별 트래픽의 상태와 Drop율, 네트워크 이용율, 그리고 스위치의 입력버퍼 변화를 분석한다.

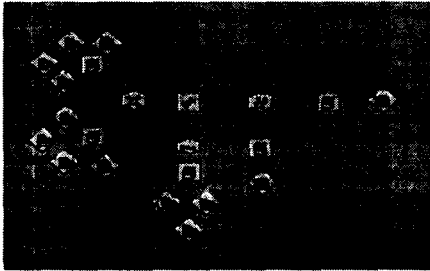


그림 5 시뮬레이션 네트워크 모델

우선, 각 노드에 입력되는 트래픽 소스의 큐 도착 분포변수를 Pareto 분포로 설정하고, Shape 변수( $1 \leq \alpha \leq 2$ )의 값을 변화시켰을 때, 스위치 노드에 입력되는 중첩된 트래픽들의 자기유사 정도가 입력변수인 Pareto의 Shape 변수를 따르는지 살펴보았다.

$$\alpha = 3 - 2H \quad (11)$$

식(11)에 의해,  $\alpha = 2$ 가 일 때  $H = 0.5$ 가 되고,  $\alpha = 1.1$  일 때  $H = 0.95$ 가 되어 비 자기유사성의 트래픽 모델과 자기유사성의 트래픽 모델을 만들게 되고, 스위치의 입력버퍼에서 도착하는 트래픽의 버퍼사용율 Trace(그림6, 그림7)를 관찰하여 자기유사성을 분석하면 다음과 같다.

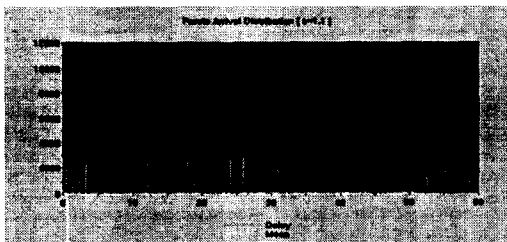


그림 6 Pareto Arrival Distribution (  $\alpha=1.1$  )

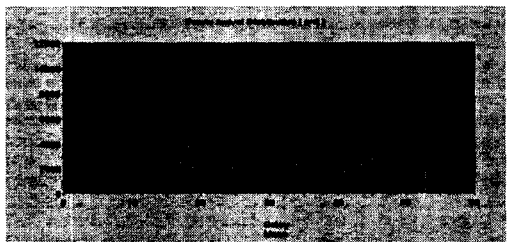


그림 7 Pareto Arrival Distribution (  $\alpha = 2$  )

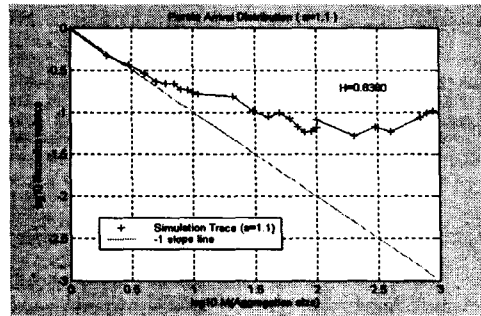


그림 8 Pareto 분포( $\alpha=1.1$ )의 VT-plot

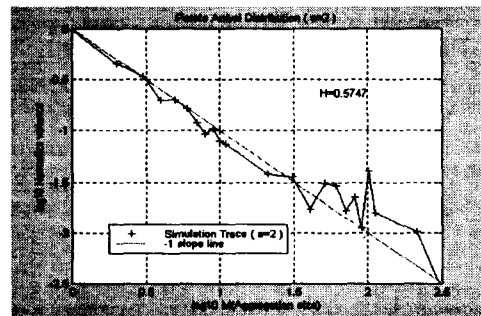


그림 9 Pareto 분포( $\alpha=2$ )의 VT-plot

소스 트래픽의 Pareto 변수에 따른 자기유사 정도를 관찰한 결과, Pareto Shape 변수  $\alpha$ 가 1.1 일 경우, 이론적인  $H$ 의 값은 0.95가 된다. 시뮬레이션 네트워크에서 다음과 같은 변수로 지정 후 트래픽을 인가한 결과  $H$ 는 0.8380의 값이 되어 이론적인 값보다는 작지만, Self-similar 트래픽 특성을 나타내고 있다.

다음은 이러한 트래픽 소스를 시뮬레이션 네트워크 모델에 인가하였을 때, 네트워크에 일어나는 통계적인 현상을 분석한다.

표-2. Drop율 비교

shape 변수( $\alpha$ )	1.1	1.3	1.5	2
Drop Packet(수)	9	7	21	24

첫 번째로, Drop율을 살펴보면 Pareto ( $\alpha=1$ )은 입력 큐에 도착하는 트래픽의 Heavy-tail 특성에 의해 버퍼 도착확률이 분산되는 반면, SRD 특성을 갖는 Pareto ( $\alpha=2$ )는 단기간 버퍼에 집중적으로 도착함으로, 중첩된 트래픽에서는 Drop의 확률이 높아지게 된다.

표 3. 지연을 비교

shape 변수( $\alpha$ )	1.1	1.3	1.5	2
지연을 평균(ms)	4.774	5.0046	4.951	5.2744

두 번째로, 시뮬레이션 네트워크에서 지연을 비교해 보면, 버퍼의 오버플로 확률이 높은 Pareto ( $\alpha=2$ )에 의해 버퍼에서 서비스를 위해 대기하는 시간이 길어지고, 따라서 전체 시스템의 지연에 영향을 미치게 된다.

표 4. 채널 이용률 비교

shape 변수( $\alpha$ )		1.1	1.3	1.5	2
Channel Utilization (%)	Ethernet	2.742	3.721	5.112	7.164
	Token Ring	0.824	0.915	1.562	1.972

마지막으로, 채널 이용률은 shape 변수의 증가에 따라, 각 링크의 이용률이 증가함을 보이고 있다.

이러한 Self-similar의 heavy-tailed 특성을 나타내는 Pareto 모델을 네트워크에 적용함으로써, 기존의 Poisson 도착패턴에 의한 Exponential 분포 모델의 트래픽 특성이 실제 트래픽의 통계적인 특성과는 명백한 차이가 있고, 따라서 Self-similar 특성을 고려한 네트워크 분석이 요구된다.

#### 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 음성 영상 이 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 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 논문에서는, Pareto 변수에 따른 Self-similar의 강도를 조사하고, 각기 다른 변수에 의해 네트워크에서 발생하는 통계적인 특성을 분석하였다.

차후 트래픽의 자기 유사성의 적용 범위에 대한 연구와 실제 트래픽 에서의 자기유사성 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, Murad 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", 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] 이병현, "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