

실시간 운영중인 Ethernet 트래픽의 장기간 의존성 및 Self-similar 특성에 대한 연구

김창호¹, 이동철², 박기식³, 류용희⁴, 최삼길¹, 김동일⁵
¹동의대학교, ²영진전문대학, ³한국전자통신연구원

A Study on the Long-range Dependence and Self-similar Characteristic of Real-time Ethernet Traffic Trace

Chang-ho Kim¹, Dong-chul Lee², Ki-shik Park³

Yong-hee Ryu⁴, Sam-gil Choi¹, Dong-il Kim⁵

¹Donggeui University, ²Youngjin Jr. College, ³ETRI

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

요 약

LAN, WAN 및 VBR 트래픽 특성에 관한 최근의 실험적 연구들은 기존의 Poisson 가정에 의한 모델들이 네트워크 트래픽의 장기간 의존성 및 self-similar 특성들을 과소평가 함으로써, 실제 트래픽의 특성을 제대로 나타낼 수 없다는 것을 지적해 왔다. 본 논문에서는 실시간 운영중인 Ethernet 트래픽을 측정하여, 통계학적인 접근법들을 통해 이러한 self-similar 특성들에 대해 평가하였다. 그리고 exactly self-similar 모델링인 Pareto-like ON/OFF 소스 모델링에 의한 트래픽과 기존의 Poisson 모델링을 비교 분석함으로써, self-similar 트래픽이 실제의 Ethernet 트래픽 특성을 잘 반영한다는 것을 보였다.

I. 서 론

네트워크 트래픽에 관한 최근의 논문들은 LAN, WAN, VBR 비디오 트래픽이 통계학적으로 self-similar 과정에 의해 더욱 더 잘 모델링 된다는 것을 설득력 있게 주장해왔다^{[1][5]}. 이것은 Poisson 과정에 비해 매우 다른 이론적 특성을 가진다. 즉, 광범위한 time scale 상에서 aggregation의 정도가 변하더라도 통계학적으로 동일한 특성을 지니게 된다. 이러한 특징들은 self-similarity, long-range dependence(Joseph Effect), infinite variance (Noah Effect), slowly decaying variance 등으로 대변된다^{[1][5]}.

본 논문에서는 실시간 운영중인 Ethernet LAN 으로부터 1초 간격으로 약 백만 개의 패킷 trace 를 측정하여, 여러 가지 self-similar 특성에 대해 분석하였다. 그리고 실제의 트래픽과 비교를 위하여 self-similar 트래픽 모델링에 의해 생성한 인위적인 트래픽과 기존의 Poisson 모델링에 의한

인위적인 트래픽과 비교하였다.

II. Self-similarity 트래픽의 특성

self-similar 확률과정은 기존의 논문들에서 여러 가지 방법으로 정의되어져 왔다. 본 논문에서는 먼저 연속시간 확률과정에 대해 살펴본 후, 데이터 트래픽과 관련된 이산시간 확률과정에 대해서 살펴본다^{[1][5]}.

2.1 연속시간 정의

어떠한 실수 $a > 0$ 에 대해, 확률과정 $a^{-H}x(at)$ 가 $x(t)$ 와 통계적으로 동일한 특성을 가진다면, 확률과정 $x(t)$ 는 파라미터 $H(0.5 \leq H \leq 1)$ 를 가지고 통계적으로 self-similar하다. 이러한 관계는 다음의 3가지 조건으로 표현된다^{[2][4][5][6]}.

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

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

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

Hurst 또는 self-similarity 파라미터 H는 self-similarity의 핵심척도이다. 다시 말하면, H는 통계적인 현상의 지속성(persistence)에 대한 척도이고 확률과정의 장기간 종속에 대한 척도이다. H=0.5의 값은 self-similarity의 부재를 나타내고, H가 1에 가까울수록, 지속성의 정도 또는 장기간의 종속의 정도는 더욱 커진다.

2.2 이산시간 정의

정상 시계열(stationary time series) x에 대해, m-aggregated 시계열 $x^{(m)} = \{x_k^{(m)}, k=0,1,2,\dots\}$ 를 인접한 m크기의 블록을 겹침 없이(nonoverlapping) 원래의 시계열을 합계함으로써 정의한다. 이것은 다음과 같이 표현된다.

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

만약 이것의 확률과정의 통계(mean, variance, correlation, etc)가 동일한 압축 사본을 간직하고 있다면, self-similar 과정으로 다룰 수 있다.

확률과정 x가 모든 $m=1,2,\dots$ 에 대해서 다음과 같다면 파라미터 β ($0 < \beta < 1$)에 대하여 정확하게(exactly) self-similar하다고 한다.

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

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

파라미터 β 는 앞에 정의했던 Hurst 파라미터 ($H=1-(\beta/2)$)와 관련이 있다. 정상과정 및 에르고딕 과정에서는 $\beta=1$ 인 경우로써, 시간 평균의 분산은 $1/m$ 비율로 감소하게 되지만, self-similar 과정에서는 시간평균의 분산은 더욱 천천히 ($1/m^\beta$) 감소하게 된다.

확률과정 x가 충분히 큰 모든 k에 대해 다음과 같다면 근사적으로(asymptotically) self-similar하다고 한다.

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

따라서, 이러한 self-similarity의 정의에 의해서, aggregated 과정의 자기상관은 원래의 확률과정과 동일한 형태를 가진다. 이것은 변이성 또는 버스트한 정도가 서로 다른 타임 스케일에서 동일하게 나타난다는 것을 의미한다.

2.3 장기간 의존성(Long-range Dependence: LRD)

self-similar 과정의 중요한 특징들 중의 하나는 장기간 의존성으로 나타난다. 일반적으로, 단기의 종과정은 자기공분산이 최소한 지수적으로 급격히 감소한다는 조건을 만족시킨다^{[5][7]}.

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

일반적으로 기존의 논문에서 고려된 데이터 트래픽 모델들의 형태는 단기간 의존 과정을 고려했다.

$$\sum_{k=0}^{\infty} x^k = \frac{1}{1-x}, \quad |x| < 1 \quad (3)$$

여기서 $\sum_k C(k)$ 가 유한하다는 것을 알 수 있다. 이에 반해, 장기간 의존과정은 hyperbolic하게 감소하는 자기공분산을 가진다.

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

여기서 β 는 앞에서 정의된 파라미터이고 $H=1-(\beta/2)$ 이기 때문에 Hurst 파라미터와 연관이 있다. 이 경우 $\sum_k C(k) = \infty$ 이 됨을 알 수 있다. 장기간 의존성은 self-similar 과정들에서의 지속적 현상을 반영한다. 즉, 모든 타임스케일에서의 군집 및 버스트한 특성의 존재를 나타낸다.

2.4 Heavy-tailed Distributions

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

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

가장 단순한 heavy-tailed 분포는 파라미터 k와 α ($k, \alpha > 0$)를 갖는 Pareto 분포이다. 밀도함수와 분포함수는 다음과 같다.

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

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

따라서 기대값은 (10)식과 같이 나타난다.

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

파라미터 k는 랜덤변수가 취할 수 있는 최소값을 나타내며, 파라미터 α ($H=(3-\alpha)/2$)는 랜덤

변수의 기대값 및 분산을 결정한다. 만약 $\alpha \leq 2$ 이면, 분포함수는 무한 분산을 가지고, $\alpha < 1$ 이라면, 무한한 기대값과 분산을 가지게 된다. 그림 1은 log-linear 스케일의 Pareto 및 지수밀도 함수를 비교한 것이다. 이 그림에서 지수밀도 함수는 거의 직선으로 나타나고, Pareto 분포의 tail은 지수 함수보다 매우 더 서서히 감소한다. 그러므로 'heavy tail' 분포함수라 한다.

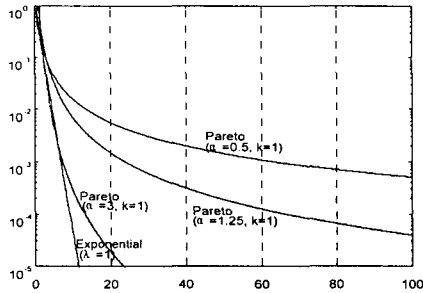


그림 1. Pareto 및 지수 확률 밀도 함수

III. 트래픽 측정 및 self-similar 트래픽 모델링 방법

3.1 트래픽 측정

Ethernet LAN 패킷 트래픽의 측정은 HP사의 Internet Advisor를 사용했고, 트래픽 샘플링 간격은 1초 단위로 약 12일 동안(99년8월2일-8월13일) 백만 개의 데이터를 수집했다. 이 기간동안은 방학기간이라 평균이용률은 0.5%이내였다. 트래픽 측정은 동의대학교 산학협력관의 한 서버넷(203.241.205.0)에서 수행하였다.

3.2 Self-similar 트래픽 모델링 방법

Self-similar 트래픽을 모델링하기 위해 다양한 접근 방법들이 취해져 왔다. 그 중에는 FGN (Fractional Gaussian noise), FBM (Fractional Brownian Motion), FARIMA(p,d,q)(Fractional ARIMA) 등 다양한 접근 방법들이 사용되어 왔었다^{[3][5][6]}. 또한 Heavy-tailed 분포를 갖는 수많은 ON/OFF 소스들의 중첩이 결과적으로 self-similar aggregate 트래픽의 결과를 낳는다는 연구결과도 있었다^[9].

본 논문에서는 self-similar 트래픽 모델링을 위해 100개의 독립적인 Pareto-like ON/OFF 소스 모델을 중첩(이하 "Pareto100" 발생기)시켜 발생시켰다^{[3][9]}. ON기간에서는 단위크기의 패킷이 CBR로 발생하고, OFF기간에서는 트래픽이 발생

하지 않는다고 가정했다. 그리고 Pareto 분포를 갖는 확률변수를 발생시키기 위해 역 변환법을 사용하였다^[10]. 이 과정의 블록 다이어그램은 그림 2와 같다. 그리고 결과적으로 aggregate된 표본경로(sample path)의 주변분포(Mean=0, Var=1로 정규화, 샘플갯수=백만개)는 FGN과 동일한 PDF (Probability Density Function)를 가짐을 알 수 있다 (그림 3)^[8].

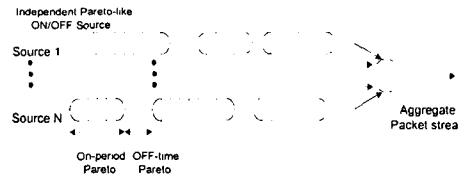


그림 2 "Pareto100" 발생기의 블록 다이어그램

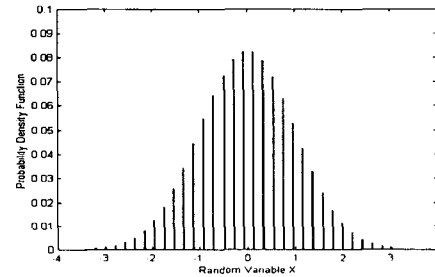


그림 3 정규화된 "Pareto100" 트래픽의 주변분포

IV. 패킷 트래픽 특성 분석

4.1 자기 발견적(Heuristic) 방법에 의한 분석

그림 4는 자기 발견적인 방법으로 간단하게 self-similar 특성을 살펴보는데 많이 사용된다^{[5][6]}. 그림의 왼쪽 열은 실측한 Ethernet 트래픽 trace를 서로 다른 aggregation 레벨에 따라 1000개의 샘플을 그린 것이다. 오른쪽 그림은 실측한 트래픽과 동일한 확률적 특성으로 상기의 "Pareto100" 발생기에 의해 인위적으로 발생시킨 것이다. 비교를 위해 가운데 그림은 기존의 Poisson 가정에 의해 동일한 확률적 특성을 지니도록 인위적으로 발생시켰다.

분포적인 관점에서 살펴보면 실제의 데이터(왼쪽)와 self-similar 모델링에 의해 발생시킨 트래픽은 동일한 특성을 지님을 알 수 있다. 즉, aggregation 레벨에 상관없이 버스트함이 유지됨을 알 수 있다. 이에 반해, 기존의 Poisson 과정에 의해 발생시킨 트래픽은 낮은 aggregation 레벨에서는 그 버스트함 볼 수 있으나 aggregation의 레벨이 높아짐에 따라 버스트함이 급격하게 사라지

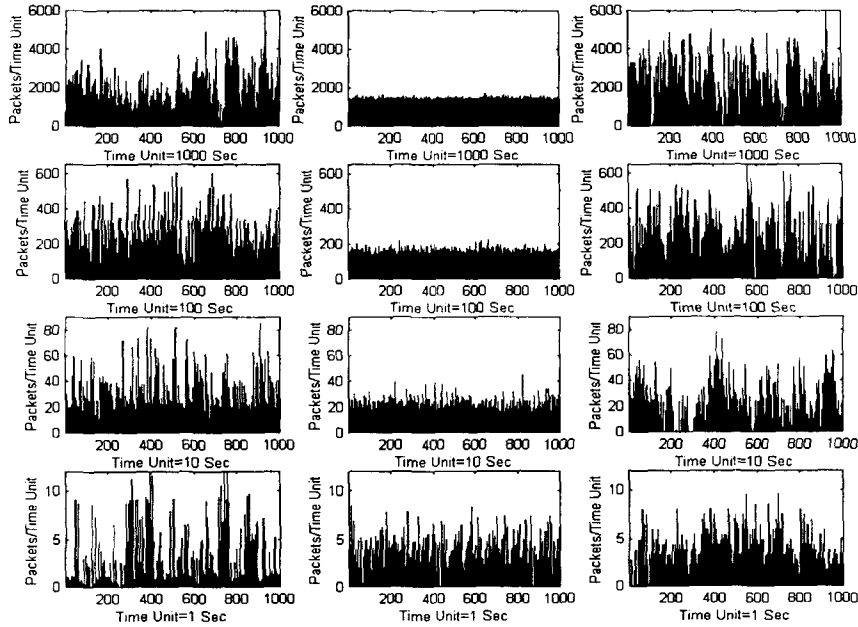


그림 4 aggregation레벨에 따른 실제 트래픽과 합성된 트래픽의 비교

는 것을 볼 수 있다. 즉, 분산이 aggregation의 레벨이 높아짐에 따라 급격하게 zero로 수렴하게 된다. 여기에서 알 수 있듯이 기존의 Poisson 가정을 사용하는 큐잉분석이 네트워크 트래픽을 모델링 하는데 적합하지 않음을 단적으로 알 수 있다.

의 Poisson 가정으로 생성한 트래픽의 경우 0.5에 가까운 H값을 가짐을 알 수 있다. 즉, 거의 백색 잡음에 가까운 특성을 나타낸다. 여기에서도 알 수 있듯이, self-similar 모델링에 의해 발생된 트래픽의 실제 트래픽 특성을 잘 나타내는 반면, Poisson가정의 경우 실제 트래픽의 버스트한 성질을 제대로 나타내지 못함을 알 수 있다.

4.2 Self-similarity의 강도 측정

Hurst 파라미터 H 가 self-similarity의 강도(intensity)를 나타내는 척도이다. 이러한 Hurst 파라미터를 추정하는 방법에는 여러 가지가 있지만, 본 논문에서는 가장 많이 사용하고 간단한 Variance-time plot법을 사용해서 Self-similarity의 강도를 측정하였다^[2]. 그림 5는 그림 4의 3가지 트래픽에 대한 log-log그래프 상의 V-T plot을 나타낸다. 비교를 위해 각 시계열의 분산을 1로 정규화 시켰다.

각 직선의 기울기는 $-\beta$ 을 나타낸다. 기울기를 추정하기 위해 최소자승 직선 근사법(least square line fitting)을 사용하여 $\beta(H=1-\beta/2)$ 를 구한 뒤 Hurst 파라미터 H 값을 구해보면 표 1과 같다. 그림 5와 표 1에서 알 수 있듯이 실제의 트래픽과 self-similar 모델링에 의해 인위적으로 생성된 트래픽의 경우 1에 가까운 아주 높은 Hurst 파라미터 값을 가짐을 알 수 있다. 이는 self-similarity의 강도가 아주 강함을 나타낸다. 이에 반해서 기존

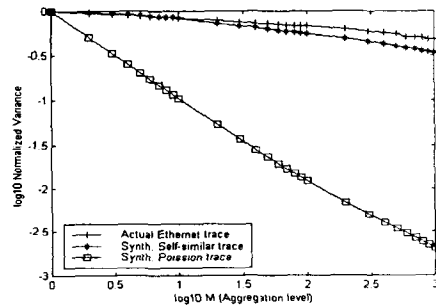


그림 5 세 가지 트래픽의 Variance-Time Plot

표 1 trace별 Hurst 파라미터 추정값

트래픽 trace	β	H
Actual Ethernet trace	0.1054	0.9473
Synth. Self-similar trace	0.1688	0.9156
Synth. Poisson trace	0.9478	0.5261

4.3 장기간 의존성(LRD)

LRD 특성을 고찰하기 위해 각 trace의 자기상관함수를 구했다. 그림 6은 시차(lag)에 따른 자기상관함수를 나타낸다.

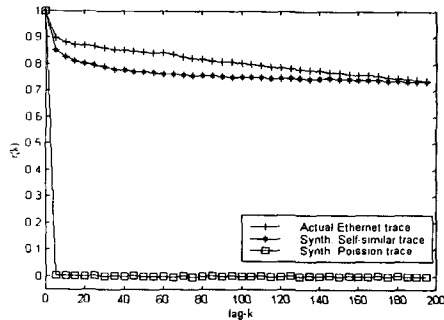


그림 6 시차(lag-k)에 따른 자기상관함수

여기에서도 알 수 있듯이 실제의 트래픽 데이터와 "Pareto100" 데이터 둘 다 시차(lag-k)에 따라서 아주 높은 상관성을 지니는 것을 알 수 있다. 즉, 장기간 의존성(LRD)을 가진다. 이에 반해, Poisson 가정에 의한 트래픽의 경우 시차 k가 0의 경우에는 1의 상관을 가지는 반면에 시차 k가 커짐에 따라 급격하게 0으로 수렴함을 알 수 있다. 이는 백색 잡음과 같은 특성이며, 단기간 의존성(SRD)을 가진다.

V. 결론

네트워크 트래픽의 특성을 이해하는 것은 네트워크 설계 및 성능분석의 핵심이다. 그러나 기존의 Poisson 가정의 경우 단기간의존성만을 고려한 경우로써 실제의 트래픽 특성이 제대로 반영할 수 없었다.

본 논문에서는 Ethernet LAN상의 백만 개의 패킷 트래픽을 측정하였고, 이것을 self-similar 모델링에 의한 트래픽 및 기존의 Poisson 가정에 의한 트래픽과 비교 분석하였다. self-similarity의 강도를 나타내는 Hurst 파라미터의 경우 실제의 데이터는 1에 가까운 값(0.9473)을 가짐을 알 수 있었고, Poisson 트래픽의 경우 0.5에 가까운 값(0.53)을 가짐을 알 수 있었다. 즉, 이는 기존의 Poisson 가정의 경우 실제 네트워크 트래픽 특성(self-similarity, LRD)을 제대로 반영하지 못함을 의미한다. 그리고 장기간 의존성 테스트의 경우 실제 트래픽과 "Pareto100" 트래픽은 시차 k에 따라 자기상관함수가 매우 천천히 감소함을 알 수 있었다. 즉, 장기간 의존성(LRD)을 가진다는 것을 알 수 있다. 이에 반해 Poisson 트래픽의 경우 시

차 k에 따라 자기상관함수가 매우 급격히 줄어들며(단기간 의존성: SRD)을 알 수 있었다.

추후 연구과제로서는 이러한 self-similar 특성이 네트워크의 성능(지연 및 손실특성)에 미치는 영향과 실제의 트래픽 특성을 반영한 큐잉모델에 대한 연구가 요구된다.

참고문헌

- [1] Schroeder, M., Fractals, Chaos, Power Laws :Minutes from an Infinite Paradise. Freeman, 1991.
- [2] M.S.Taqqu, V.Teverovsky, W.Willinger, Estimators for long-range dependence: an empirical study. Fractal Vol.3.No.4 p785-788, 1995
- [3] W.Willinger, M.S.Taqqu, R. Sherman, Self-Similarity Through High-Variability, ACM/Sigcomm, 1995
- [4] Peyton Z. Peebles, JR. Probability, Random Variables, and Random Signal Principles, McGraw Hill, p134-198, 1993
- [5] Leland, W., Taqqu, M., Willinger, W., Wilson, D. On the Self-similar Nature of Ethernet Traffic(Extended Version), IEEE/ACM Transaction on Networking, Feb, 1994
- [6] Willinger, W., Wilson, D., Taqqu, M. Self-similar Traffic Modeling for High-speed Networks, ConneXions, Nov, 1994
- [7] Grossulauser, M. Bolot, J. On the Relevance of Long-range Dependence in Network Traffic, Proceedings, SIGCOMM'96, Aug, 1996
- [8] T., Taralp, M., Devetsikiotis, I., Lambadaris, Efficient Fractional Gaussian Noise Generation Using the Spatial Renewal Process, IEEE, 1998
- [9] P., Pruthi, Heavy-Tail ON/OFF Source Behavior and Self-Similar Traffic, ICC'95, 1995
- [10] H., P., Schwefel, Modelking of Packet Arrivals Using Markov Modulated Poission Process with Power-Tail Bursts, Connecticut Univ., 1997
- [11] 김창호, 김동일 외, 데이터 트래픽에서의 Self-similar 특성, 춘계해양정보통신학회, p146-151, 1999
- [12] 김창호, 김동일 외, 트래픽에서의 장기간 의존성 및 Self-similar 특성, 하계통신학회, p463-467, 1999