

# 실시간 운영중인 네트워크 상에서 Self-Similarity 특성 및 Hurst 파라미터 측정

\*진성호, \*임재홍

\*한국해양대학교 전자통신공학과

## A Measurement of Self-Similarity Characteristic and Hurst Parameter on Real Time Operation Network

\*Sung-Ho Jin, \*Jae-Hong Yim

\*Dept. of Electronics & Communication Eng., Korea Maritime University

E-mail : deodato@thru.net.com

### 요 약

네트워크를 설계하고 서비스를 구현하는데 있어서 중요한 변수중의 하나는 트래픽의 특성을 파악하는 것이다. 기존의 트래픽 예측과 분석으로 Poisson 또는 Markovian을 기본으로 하는 모델을 사용했을 경우는 단기간의 의존성을 고려한 결과로써 실제 관측된 트래픽의 결과와는 상당히 다르다는 것이 밝혀졌다. 따라서 최근 실제 트래픽 모델과 유사한 모델로서 Self-Similarity 특성을 이용한 접근법이 대두되고 있다. 본 논문에서는 Self-Similarity의 장기간 의존성을 나타내기 위해서 실제 네트워크에서 측정된 데이터를 사용하여 Hurst 파라미터 H의 값을 추정하고 실시간 운영중인 네트워크 상에서 어느 정도의 Self-Similarity 특성을 가지고 있는지 분석한다.

### I. 서 론

1990년대 이후 네트워크 환경에서는 인터넷 사용자의 폭발적인 증가와 응용프로그램의 멀티미디어화로 트래픽의 양에 대해 급속적인 증가를 가져오고 있다. 그리고 사용자의 서비스 품질에 대한 요구가 점점 커지고 있기 때문에 지속적으로 네트워크 확충이 계속되고 있으나, 사용자들에게 충분한 만족을 주기에는 아직 미비한 실정이다. 이러한 서비스의 품질을 높이기 위해 다양한 기법이 개발되고 되고 있으며, 실제 제품을 생산하여 현장에서 운영을 하고 있다. 하지만 현장에 이러한 장비를 운영했을 시 그 장비의 기대치에 대한 효율을 가져오고 있지 못한 실정이다. 그러한 이유 중에 여러 가지가 있으나 중요한 요소중의 하나가 장비 설계 시 모델링을 하는 트래픽의 특성이 실제 운영중인 네트워크의 트래픽 특성과 많은 차이를 보인다는 것이다. 기존에는 포아송(Poisson)이나 마코비안(Markovian)을 기반으로 하는 큐잉모델을 사용하여 만든 스위치나 허브 등의 장비들은 실제 성능보다 작은 버퍼를 사용하게 한 것이다. 과거의 텍스트 기반의 네트워크 환경에서는 문제점이 드러나지 않았지만 현재의 트래픽 특성은 사용자의 증가와 특히 트래픽의 종류가 일반적인 텍스트가 아닌 멀티미디어화로 인해 그리고 대용량의 파일 전송이 많아지면

서 아주 버스트(burst)한 특성을 보이고 있기 때문에 기존의 모델을 사용한 경우 많은 문제점을 나타내고 있다. 최근에 이러한 문제점을 goruf하기 위해 여러 연구가 진행이 되고 있고 특히 Self-Similarity 특성에 관한 연구가 트래픽에 대한 모델로서 인증을 받고 있는 추세이다. 이에 본 논문에서는 Self-Similarity의 이론의 기본 개념에 대해 설명하고 실제 운영중인 네트워크에서 트래픽을 측정하여 어느 정도의 Self-Similarity 정도를 가지고 있는지를 파악한 후 이러한 특성을 나타내는 척도인 Hurst 파라미터 값을 추정해 낸다.

### II. Self-Similarity

Self-Similarity는 단지 최근에 들어서야 데이터 통신 트래픽 분석에 적용된 중요한 개념이다. 또한 몇몇 연구를 통해 Self-Similarity가 도처에 산재해 있다는 것이 밝혀졌다. Self-Similarity는 차원상의 서로 다른 확률대비율이나 서로 다른 스케일에서 보았을 때 동일하게 보이거나 동일하게 행동하는 것이 자기 유사한 현상을 말한다. 즉, 데이터 트래픽의 타임 시퀀스가 해상도의 정도에 상관없이 동일한 패턴으로 나타나는 것이다. 이것이 Self-Similarity의 핵심이다.

Self-Similar한 특징들은 실제의 현상에 대해서

무한하게 유지되는 것은 아니다. 그렇지만 아주 큰 범위의 스케일을 통해 대부분의 현상들은 Self-Similarity를 드러낼 것이다. 또한 네트워크 성능의 관점에서 가장 현저한 특징은 집단화의 지속성이며, 포아송 트래픽에서 집단화는 단기간 내에 발생하지만 장기간이 지나면 smooth out 되어 버린다.

### III. Self-Similar Data Traffic

주어진 패턴이 정확하게 서로 다른 스케일에서 반복되는 경우를 정확한 Self-Similarity라 한다. 이러한 Self-Similarity는 결정론적인 시계열에 대해 구성되어진다. 그러나 데이터 트래픽은 확률적인 과정으로 가장 잘 관찰되어지고, 통계적인 방법으로는 Self-Similarity를 언급할 수 있다<sup>[1]</sup>.

일반적으로 결정적이고 주기적인 신호는 시간 이동에 관하여 불변인 특성이 있다. 즉 그 신호는 시간상으로 여러 주기가 이동하더라도 동일한 신호이다. 이에 반해, 정상확률 과정에 대해서는 그 과정의 통계는 시간 이동에 불변이다. 또한 평균은 시간에 독립적이고 자기상관 함수는 단지 시간의 차이에만 의존한다<sup>[2]</sup>.

Self-Similar 확률과정은 기존의 논문에서 여러 가지 방법으로 정의되어져 왔다. 본 논문에서는 데이터 트래픽과 관련된 이산시간 확률과정에 대해서만 살펴본다.

#### 3.1 이산시간 정의

정상 시계열(stationary time series)  $x$ 에 대해,  $m$ -aggregated 시계열  $x^{(m)} = \{x_k^{(m)}, k=0,1,2,\dots\}$ 은 인접한  $m$ 크기의 블록을 겹침 없이(non-overlapping) 원래의 시계열을 합계함으로써 정의한다. 이것은 다음과 같이 표현된다<sup>[3]</sup>.

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

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

확률과정  $x$ 가 모든  $m=1,2,\dots$ 에 대해서 다음과 같다면 파라미터  $\beta$  ( $0 < \beta < 1$ )에 대하여 정확하게 Self-Similar하다고 한다.

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

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

Self-Similarity 특성을 나타내는 기준으로 Hurst 파라미터가 가장 많이 사용되고 있다. Hurst 파라미터  $H$ 는 통계적인 현상의 지속성에 대한 척도이고 확률과정의 장기간 종속에 대한 척도이다. 파라미터  $\beta$ 와 Hurst 파라미터  $H$ 는  $H=1-(\beta/2)$ 로 나타내어진다. 정상과정 및 에르고

딕 과정(ergodic process)에서는  $\beta=1$ 인 경우로써, 시간 평균의 분산은  $1/m$ 비율로 감소하게 되지만, Self-Similar 과정에서는 시간평균의 분산은 더욱 천천히( $1/m^\beta$ ) 감소하게된다. 그러므로  $H=0.5$ 의 값은 Self-Similarity의 부재를 나타내고,  $H$ 가 1에 가까울수록, 지속성의 정도 또는 장기간의 종속의 정도는 더욱 커진다.

확률과정  $x$ 가 충분히 큰 모든  $k$ 에 대해 다음과 같다면 근사적으로 Self-Similar하다고 한다.

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

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

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.

### IV. 분석환경

본 논문에서 실시간 운영중인 네트워크 트래픽 샘플을 측정하기 위해 사용한 네트워크 환경은 한국해양대학교 네트워크를 이용하였다. 샘플 측정은 C급 클래스인 203.230.252에서 측정을 하였고 샘플 측정 도구는 HP사의 Internet Advisor를 사용하였다. 전체 패킷 수는 신뢰성을 기하기 위하여 10만개를 1초 단위로 측정하였고 측정기간은 1999년 10월 5일부터 10월 6일까지이다. 샘플 측정의 신뢰성은 측정기간이 문제가 아니라 측정갯수에 의존하기 때문에 10만개 정도의 양이면 만족한 결과를 나타내어진다라고 예상된다. 측정 파라미터는 Ethernet 전체 패킷 수와 현재 일반적인 네트워크 상에서 주로 사용중인 프로토콜이 Internet을 이용하기 위한 IP와 학교나 연구실, 사무실에서 사용중인 PC의 주 운영체제가 MS사의 Windows를 기반으로 하여 파일 공유, 프린트 공유 등을 위한 NetBios가 주를 이루고 있다고 예상되어지기 때문에 IP와 NetBios 패킷만을 측정, 분석하였다.

### V. 분석 결과

측정한 샘플의 분석은 전체 Ethernet 패킷의 수와 같은 시간대에 IP 패킷 수 그리고 NetBios 패킷수로 분석하였다. Y축은 전체 Ethernet 패킷 수, IP 패킷 수, 그리고 NetBios 패킷 수를 각각 분리하여 나타내었다. 그리고 X축은 샘플 시간을 나타낸다. 그림 1은 패킷의 수로서 Ethernet 회선의 이용현황을 나타낸 것이다. 패킷의 수는 샘플 시간에 따라 많은 차이점을 보인다. 이것은 샘플 기간이 짧아 주, 야간이 1회 밖에 교차가 안되어

서 이러한 특징을 보이는 것이다. 그러나 이러한 특징도 샘플 기간을 장기간으로 했을 경우 본 논문에서 나타난 버스트함도 아주 작은 버스트한 성질로 나타내어 질 것이다. 그림 2와 3은 사용 프로토콜에 대해 패킷 수로 나타낸 것이다. Ethernet에서 주로 사용되는 프로토콜이 IP보다는 NetBios가 주로 사용된다는 것을 알 수 있다. 이러한 특징은 인터넷의 이용보다는 파일공유나 프린트 공유 등으로 인해 Ethernet 회선을 사용하는 것이 더 많다는 것을 보여주는 것이다.

전체적인 패킷 수와 사용 프로토콜에 대해 트래픽 특성을 살펴보면 버스트함이 잘 나타나고 있는 Self-Similarity 특성을 보여주고 있다.

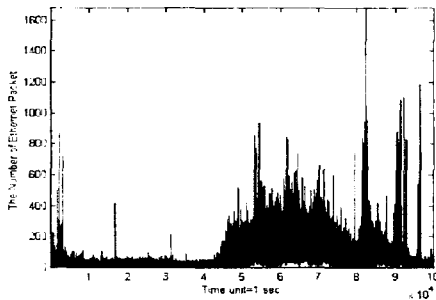


그림 1 전체 Ethernet 패킷

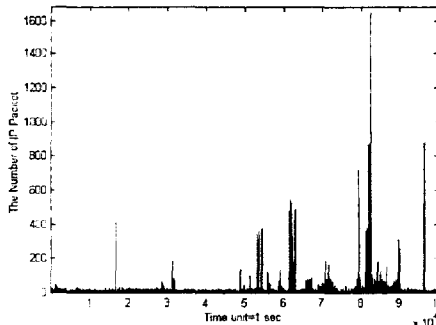


그림 2 IP 패킷

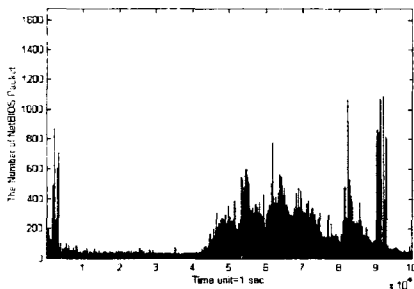


그림 3 NetBios 패킷

## VI. Hurst 파라미터 측정

앞에서 설명한 Self-Similarity를 나타내는 기준으로 많이 사용되는 Hurst 파라미터를 구하는 방법은 variance-time plot, R/S plot, Periodogram 방법 등 여러 가지가 있다. 이러한 방법으로 Hurst 파라미터를 추정했을 경우 약간의 오차는 있지만 무시할 정도라고 판단된다. 본 논문에서는 가장 간단하고 많이 사용되고 있는 variance-time plot를 사용했다.

### 6.1 Variance-time Plot

Self-Similarity 과정의  $m$ -aggregated 시계열  $X^{(m)}$ 에 대해 분산은 매우 큰  $m$ 에 대해 다음을 따른다.

$$Var(x^{(m)}) \approx \frac{Var(x)}{m^\beta} \quad (2)$$

여기에서, Self-Similarity 파라미터  $H=1-(\beta/2)$ 이다. 그리고,  $\log[Var(x)]$ 는  $m$ 과 무관한 상수이기 때문에 이 식은 다음과 같이 고쳐 쓸 수 있다.

$$\log[Var(x^{(m)})] \approx \log[Var(x)] - \beta \log(m) \quad (3)$$

이것은 log-log 그래프 상에  $m$ 에 대한  $Var(x^{(m)})$ 을 그리게 되며,  $-\beta$  값의 기울기를 가지는 직선이 나오게 된다. 즉 -1에서 0사이의 경사값은 Self-Similarity 특성을 암시하게 되는 것이다.

### 6.2 Hurst 파라미터 추정결과

그림 4, 5, 6은 본 논문에서 측정한 샘플을 사용해서 추정한 Hurst 파라미터 값이다. 그림 4, 5, 6에서 기울기가 0인 경우는 Hurst 파라미터가 1인 경우 즉 완벽한 Self-Similarity를 나타내는 경우이다. 중간의 점으로 된 그래프는 Hurst 파라미터 값을 나타내며, 하단의 참조선은 Hurst 파라미터가 0.5인 경우로서 Self-Similarity의 부재를 나타내고 있다.

Ethernet 패킷과 사용 IP, NetBios 프로토콜에 대해  $\beta$  값과 Hurst 파라미터를 구하면 표 1과 같다.

표 1 Ethernet, IP, NetBios 패킷에 대한  $\beta$  값과 Hurst 파라미터

	$\beta$	H
Ethernet 패킷	0.0891	0.9555
IP	0.1027	0.9486
NetBios	0.0968	0.9516

Hurst 파라미터 값을 살펴보면 대부분의 경우 Self-Similarity 특성이 잘 나타나고 있는 것을 알 수 있다. 분석 결과에서 IP 프로토콜이 NetBios보다 Hurst 파라미터 값이 약간 낮은 것은 전체 패킷 중에서 IP의 비율이 매우 낮기 때문이라고 판단할 수 있다. 즉 트래픽 양이 높아질수록 Self-Similarity 특성의 정도가 높아짐을 알 수 있다.

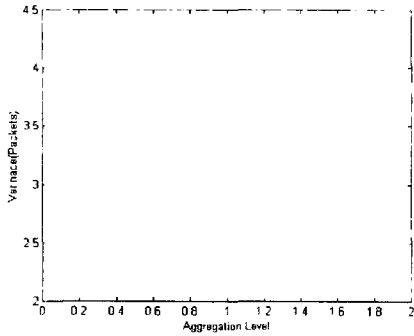


그림 4 전체 Ethernet 패킷에 대한 variance-time plot

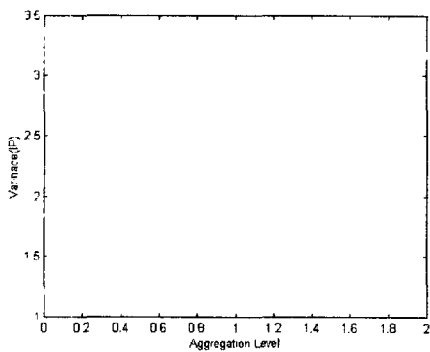


그림 5 IP 패킷에 대한 variance-time plot

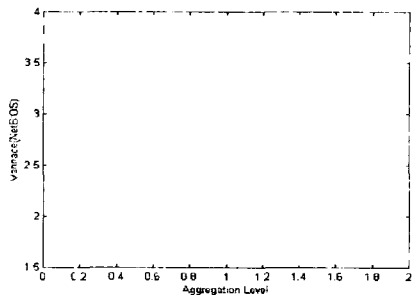


그림 6 NetBios 패킷에 대한 variance-time plot

## VII 결 론

본 논문에서는 실시간으로 운영중인 네트워크에서 패킷을 측정된 후 이에 따른 Self-Similarity 특성을 분석하고 Self-Similarity의 정도를 나타내는 Hurst 파라미터 값을 추정하였다. 본 논문에서는 한국해양대학교 네트워크를 대상으로 전체 패킷 수와 IP, NetBios 프로토콜의 패킷을 측정하였으며, 그 결과 주로 사용되는 프로토콜은 IP와 NetBios로 밝혀졌다. 이는 대규모의 백본 네트워크에서 사용되는 프로토콜과는 차이점이 있지만 소규모 네트워크 환경에 적용하기에는 충분하다고 판단된다. 그리고 측정 결과는 기존의 Poisson 또는 Markovian 모델에서 장기간에 걸쳐 모델링했을 때 트래픽의 버스트함이 없어지지만, 실제 운영중인 네트워크에서는 버스트함이 나타나는 Self-Similarity 특성을 보여주고 있는 것이 확인되었다. 또한 Hurst 파라미터를 추정하여 Self-Similarity 특성이 매우 높음을 파악하였다. 이러한 결과는 향후 네트워크 장비 설계 또는 네트워크 관리에 유용하게 활용될 수 있을 것으로 판단된다.

본 논문에서 분석한 네트워크는 소규모의 네트워크이기 때문에 대규모 백본망과는 프로토콜, 전체 패킷 수 등에 대해 많이 차이가 있을 것이다. 향후 과제는 대규모 백본망에서 트래픽을 측정해 보고 분석하여 기존의 모델과 어느 정도의 차이점이 있는가를 계속해서 파악하고 이를 사용해 Self-Similarity 모델의 적용성 및 타당성에 대해 검증해 보아야 할 것이다.

## 참 고 문 헌

- [1] William Stalling, High-Speed Network, Prentice Hall
- [2] Peyton Z. Peebles, JR. Probability, Random Variables, and Random Signal Principles, McGraw Hill
- [3] 김창호, 김동일 외, "트래픽에서의 장기간의 존성 및 Self-similar 특성", 한국통신학회 하계학술발표대회, p463-467, 1999
- [4] 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
- [5] M.S. Taqqu, V. Teverovsky, W. Willinger, "Estimators for long-range dependence: an empirical study", Fractal Vol.3.No.4 p785-788, 1995