

자기유사성을 갖는 데이터 트래픽의 통계적인 특성

Statistical Characteristics of Self-similar Data Traffic

구혜련, 홍경호, 임석구
천안대학교

Koo Hye-Ryun, Hong Keong-Ho, Lim Seog-Ku
Cheonan University

Abstract

Recent measurements of local-area and wide-area traffic have shown that network traffic exhibits at a wide range of scales - Self-similarity. Self-similarity is expressed by long term dependency, this is contradictory concept with Poisson model that have relativity short term dependency. Therefore, first of all for design and dimensioning of next generation communication network, traffic model that are reflected burstness and self-similarity is required. Here self-similarity can be characterized by Hurst parameter. In this paper, when different many data traffic being integrated under various environments is arrived to communication network, Hurst Parameter's change is analyzed and compared with simulation results.

I. 서론

현재 제공되는 많은 인터넷 서비스들의 동작 특성은 자기 유사성(self-similarity)이라는 기준에 고려되던 트래픽 특성과는 완전히 다른 장기간 의존성(Long Range Dependency)의 성질들을 가진다는 것이 증명되었다. 이것은 단기간 의존성(Short Range Dependency) 성질을 표현하는 기존의 모델인 포아송(Poisson) 모델과는 상반되는 개념으로서 효율적인 네트워크 디멘셔닝(dimensioning) 절차와 트래픽 관리 기능 등에 상당히 중요한 영향을 미칠 것이며, 현재 전 세계적으로 이러한 트래픽 모델에 관한 연구가 활발히 진행 중이다[1~5].

자기 유사적인 트래픽 성질이 차세대 고속 통신망의 성능에 미치는 영향을 예측 분석하기 위해서는 자기유사 트래픽을 정확하고 효율적으로 모델링하기 위한 방안에 대한 연구 및 분석 결과는 인터넷 망이나 새롭게 구현되어지는 시스템에서 반드시 고려하

여야 할 사항이라고 판단된다. 여기서 자기유사성은 허스트 파라미터(Hurst Parameter)로 특성화 될 수 있다. 이러한 관점에서 본 논문에서는 데이터 트래픽 특성이 서로 다른 다수의 데이터 다수의 데이터 트래픽이 통합되어 통신망에 입력되는 경우 주요 파라미터인 허스트 파라미터의 변화를 다양한 환경에서 분석하였고 이를 시뮬레이션 결과와도 비교하였다. 또한 차세대 통신망 설계와 디멘셔닝 단계에서 요구되는 채널 및 중계선과 같은 자원들의 효율적인 물량산출 및 설계를 위해 필요한 유효 대역폭(Effective Bandwidth)관점에서도 분석하였다.

본 논문의 구성은 다음과 같다. 서론에 이어 II장에서는 자기유사성의 특성과 유효대역폭을 설명하고, III장에서는 데이터 통합에 따른 허스트 파라미터의 변화를 분석하고 IV장에서는 시뮬레이션 수행 및 그 결과를 분석하며 마지막으로 V장에서는 결론을 맺는다.

II. 자기 유사성 및 유효 대역폭

2.1 자기 유사성의 정의

자기 유사성은 기존의 Fractal과 Chaos 이론에 기초한 개념이며, 차원(Dimension)상의 서로 다른 확률 대비율이나 서로 다른 스케일에서 보았을 때, 동일하게 보이거나 동일하게 행동하는 자기 유사한 현상이다.

자기 유사성을 갖는 확률 과정의 일반적 정의는 연속 시간 변수의 직접적인 스케일링에 근간을 두고 있다. 확률 과정 $x(t)$ 가 임의의 실수 $a (>0)$ 에 대해서 $a^{-H}x(at)$ 와 같은 통계적 특징을 지니면, $x(t)$ 는 파라미터 $H(0.5 < H < 1)$ 을 갖는 통계적 자기 유사성을 갖는 프로세스라 한다. 이 관계는 다음과 같은 세 가지 조건에 의해서 표현될 수 있다.

$$\text{평균 : } E[x(t)] = \frac{E[x(at)]}{a^H} \quad (1)$$

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

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

여기서 파라미터 H 는 허스트 파라미터 또는 자기 유사성 파라미터로서 알려져 있는데, 이것은 자기 유사성의 주요 척도이다. 보다 정확히 말하면 H 는 통계적 현상의 지속성에 대한 척도이고 Stochastic Process의 장기간 의존성(Long-Range Dependence)의 길이에 대한 척도이다. H 값이 0.5이면 자기 유사성이 없음을 의미하고 1에 가까우면 지속의 정도 또는 장기간 의존성의 정도가 크다는 것을 의미한다.

2.2 유효 대역폭

기존의 음성 트래픽과는 달리 자기 유사성을 가지는 데이터 트래픽의 경우 시스템이나 네트워크 내에서 수많은 통합(Aggregation) 과정을 거치더라도 분산이 거의 줄어들지 않기 때문에 단순히 평균 트래픽의

양으로 시스템을 설계하는 데는 무리가 따른다. 따라서 데이터 트래픽의 특징인 버스트성과 자기 유사성이 반영된 유효 대역폭으로 네트워크 트래픽을 수용하는데 필요한 전송용량 및 큐 모델의 서버용량 등을 평가할 필요가 있다. 이러한 유효 대역폭은 호 수락 제어(CAC : Call Admission Control)나 네트워크 디맨저닝 시에 유용하게 사용될 수 있다.

본 논문에서는 유효 대역폭을 산출하기 위한 해석적 모델로써 Norros가 제안한 자기 유사 트래픽 소스에 대한 유효 대역폭(C)은 식(4)와 같이 정의된다[6].

$$C = m + [x(H)\sqrt{-2 \ln \varepsilon}]^{1/H} (ma)^{1/(2H)} B^{-(1-H)/H} \quad (4)$$

여기서,

m : 트래픽 스트림의 평균 비트율(bits/sec)

a : 트래픽 스트림의 분산계수(bit-sec)

H : 허스트 파라미터($0.5 < H < 1$)

B : 버퍼 크기(bits)

ε : 목표로 하는 셀 손실률(Target Cell Loss Rate)

$$x(H) = H^H(1-H)^{1-H}$$

식(4)를 이론적으로 유도하는데 있어서 기반이 되는 두 가지 가정사항은 매우 많은 i.i.d. (independent, identically distributed)한 트래픽이 통합되어졌을 때 Gaussian 특성을 따른다는 것과, 큐 길이의 분포의 꼬리부분이 Weibull 분포를 따른다는 가정이다. 따라서 이 식은 적은 버퍼 크기에 대해서 오차가 많은 단점이 있다.

III. 트래픽 파라미터 및 통합 트래픽 분석

3.1 트래픽 분석

평균이 m 인, 분산계수 a , 허스트 파라미터가 H 인 N 개의 트래픽 스트림이 통합되는 경우를 고려한다.

여기서 통합 트래픽 스트림의 부하는 M 으로 일정하다고 가정한다. 단일 소스인 $N=1$ 인 경우 소스의 특성은 m, a, H 이다. $N=2$ 인 경우 두 개의 소스는 각각 $m/2, a/2, H$ 이다. 앞에서의 결과로부터 $N=1$ 인 경우와 $N=2$ 인 경우의 유효 대역폭은 같지 않다. 단일 소스인 $N=1$ 인 경우 통합 트래픽 스트림은 M, a, H 의 특성을 가지며, 소스가 두 개인 $N=2$ 인 경우는 $M, a/2, H$ 의 특성을 가진다. 따라서 $N=2$ 인 경우의 유효 대역폭은 $N=1$ 인 경우의 유효 대역폭보다 작음을 알 수 있다. 다른 말로 표현하면 통계적인 다중화 이득이 있음을 의미한다. 즉, 통합 스트림의 유효 대역폭은 각 트래픽 스트림의 유효 대역폭의 합보다 작다. 표 1에는 자기 유사 트래픽 소스의 통합에 관련된 이론적인 결과를 요약하였다.

[표 1] 통합된 자기유사 트래픽의 통계적인 특성

| 평균 | 분산 | 분산계수 |
|--------------------|--------------------|---|
| $m_g = \sum_i m_i$ | $v_g = \sum_i v_i$ | $a_g = \frac{\sum_i m_i a_i}{\sum_i m_i}$ |

3.2 허스트 파라미터(Hurst Parameter)

허스트 파라미터가 H_1 인 트래픽 스트림과 H_2 인 트래픽 스트림이 다중화되는 경우 통합 트래픽 스트림의 허스트 파라미터 H 를 결정하는 것은 매우 중요하다. 통합 트래픽의 H 를 결정하기 위한 관련 연구결과 중에서 참고문헌 [7]에서는 다수의 트래픽이 다중화되는 경우 통합 트래픽의 허스트 파라미터는 $H = \max(H_1, H_2)$ 임을 제시하였다. 그러나 이 결과는 직관적으로 장기간에 걸쳐 통합 트래픽 스트림의 분산의 감쇄는 큰 값의 H 값에 지배되지 않을 것이며, 통계적 특성이 서로 다른 트래픽이 통합되는 경우에서 통합 트래픽의 H 는 평균이나 분산값의 가중평균의 형태로 표현되는 것이 실제의 트래픽 특성을 매우 잘 반영할 것으로 예측할 수 있다. 특히, 입력되는 트래픽의 평균이 매우 상이한 경우에는 평균이

나 분산이 큰 트래픽의 허스트 파라미터 값에 의해 통합 트래픽의 허스트 파라미터 값은 좌우될 것으로 예측할 수 있다.

이러한 관점에서 본 논문에서는 다양한 자기 유사 트래픽 특성을 갖는 다수의 트래픽이 통합되는 경우 통합 트래픽의 허스트 파라미터를 효율적으로 예측할 수 있는 표 2의 방식 B를 제안하고, 제안한 근사식의 효율성을 입증하기 위하여 시뮬레이션을 수행하였으며, 참고문헌 [7]에서 제안한 방식인 표 2의 방식 A와 비교 분석하였다. 시뮬레이션 모델에서 입력 트래픽은 자기 유사성을 갖도록 트래픽을 발생시켰다. 트래픽 발생기는 FGN(Fractional Gaussian Noise) 프로세스의 Power Spectrum을 평가하기 위하여 FFT(Fast Fourier Transform)를 이용하였다 [8].

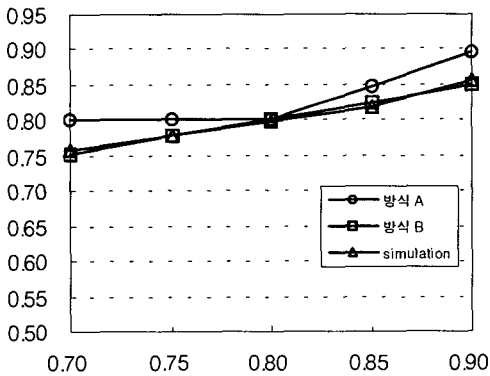
[표 2] 통합 트래픽의 허스트 파라미터 근사식

| 방식 A | 방식 B |
|-------------------------------|---|
| $H_g = \max(H_1, H_2, \dots)$ | $H_g = \frac{\sum_i v_i H_i}{\sum_i v_i}$ |

IV. 시뮬레이션

본 논문에서 제안한 다수의 트래픽이 통합되는 경우 통합 트래픽의 허스트 파라미터를 효율적으로 예측할 수 있는 방식 B의 효율성을 입증하기 위하여 시뮬레이션을 수행하였다. 시뮬레이션 전용 언어인 SLAM II를 사용하였으며, 각 데이터 점은 10만 이상의 데이터를 평균하여 구하였다.

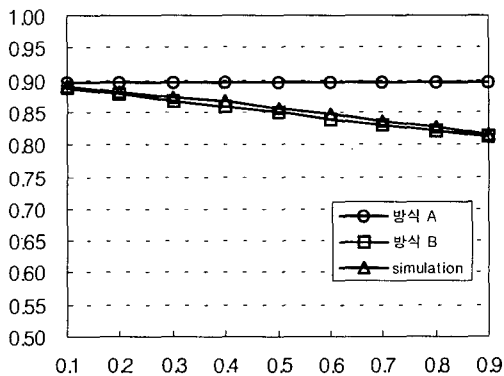
그림 1에는 2개의 트래픽이 통합되는 경우를 나타내었는데, 입력되는 트래픽의 평균이 각각 $m=75.5\text{Mbps}$, 분산계수 $a=1.7107$ 이고 허스트 파라미터 값은 하나는 $H=0.8$ 로 고정시키고 다른 트래픽의 허스트 파라미터 값은 0.7에서 0.9까지 변화시킨



▶▶ 그림 1. 트래픽이 통합되는 경우의 H의 변화

경우를 시뮬레이션 결과와 함께 나타내었다. 방식 B가 시뮬레이션 결과와 일치함을 알 수 있다.

그림 2에는 입력되는 총 입력 트래픽이 $m=155\text{Mbps}$, 분산계수 $a=1.7107$ 이고 $H=0.8, 0.9$ 인 경우 트래픽의 상대적인 비율을 변화시킨 경우를 시뮬레이션 결과와 함께 나타내었다. 방식 B가 시뮬레이션 결과와 일치함을 알 수 있다.



▶▶ 그림 2. 트래픽의 비율에 따른 통합 트래픽의 H의 변화

본 논문에서는 방식 B를 적용하여 인터넷 서비스 비율에 따른 유효 대역폭을 산출하기 위한 예로써 표 3과 같이 6가지 주요 인터넷 서비스를 고려한다. 또한 표 3은 차세대 이동 통신망에서의 자기 유사성에 관한 기존의 논문들의 연구결과에서 측정된 서비스

별 허스트 파라미터 H 값들을 정리해서 나타낸 것이다. 여기에 분석한 허스트 파라미터 값들은 기존의 유선망과 이동망에서의 데이터 트래픽의 특성이 크게 다르지 않을 것이라는 가정 하에서, 유선망에서 사용되는 트래픽을 바탕으로 분석된 값들이다.

[표 3] 주요 인터넷 서비스별 허스트 파라미터

| 서비스 유형 | Hurst Parameter 값(H) |
|--------|----------------------|
| VoIP | 0.8 |
| ftp | 0.9~0.96 |
| http | 0.74~0.76 |
| pop3 | 0.75 |
| video | 0.8~0.9 |
| telnet | 0.6 |

네트워크에서 발생하는 트래픽은 주로 다음과 같은 6가지의 프로토콜들로 구성된다. 인터넷을 이용한 음성통신인 VoIP, 파일전송을 위한 ftp, 웹상에서의 네비게이션을 위한 http, 이-메일을 위한 pop3, 원격접속을 위한 telnet 등이 있다. 인터넷 트래픽 중에서 telnet의 비율은 매우 미약하므로 네트워크 성능에 미치는 영향은 크지 않을 것으로 보인다.

실제적으로 인터넷에서 트래픽의 비율은 음성과 비디오는 10%, 데이터는 90% 정도이다. 가까운 미래에 음성과 비디오의 비율은 증가될 것으로 예측된다. 이러한 이유 때문에 본 논문에서는 음성과 비디오 트래픽 비율과 데이터 트래픽의 비율을 변화하면서 시뮬레이션을 수행하였다. 인터넷 데이터 트래픽은 http가 42%, ftp가 20%, pop3가 34%, 비디오가 3%, telnet이 1%로 구성된다. 표 4는 서로 다른 트래픽 비율을 요약하였으며 앞으로 예측되는 트래픽의 비율을 또한 정리하여 나타내었다[9].

[표 4] 서비스 비율에 대한 유효 대역폭

| 서비스 유형 | CASE I | CASE II | CASE III |
|----------|------------|------------|------------|
| VoIP | 10% | 27% | 45% |
| ftp | 16% | 13% | 10% |
| http | 40% | 30% | 20% |
| pop3 | 30% | 24% | 17% |
| video | 3% | 5% | 7% |
| telnet | 1% | 1% | 1% |
| 허스트 파라미터 | 0.765 | 0.776 | 0.792 |
| 유효 대역폭 | 45.191Kbps | 45.219Kbps | 47.352Kbps |

인터넷 서비스별 유효 대역폭을 산출하기 위한 예로써 표 4와 같이 3가지 경우의 서비스 사용비율에 따른 유효 대역폭을 산출하였다. 여기서 버퍼 크기 B 는 16K cell, 목표로 하는 셀 손실률 ϵ 은 잘 알려진 값인 10^{-6} 을 사용하였다. 여기서 각 파라미터 값들 중 분산계수(Variance Coefficient) 값은 편의상 평균 데이터 속도의 30%로 설정하였다. 추후 정확한 시스템 디멘저닝의 적용을 위해, 분산계수 및 허스트 파라미터 값에 대해서는 서비스 형태별로 실측할 필요가 있다.

표 4의 결과에 의하면 음성과 비디오 트래픽의 비율이 증가하고 다른 트래픽의 비율이 감소함에 따라 통합 트래픽의 허스트 파라미터 값은 증가하므로 유효 대역폭은 증가함을 알 수 있었다. 이러한 결과는 시뮬레이션 결과에 의해서도 확인할 수 있었다.

V. 결론

현재 제공되는 인터넷 트래픽은 자기 유사성이라는 기준에 고려되던 트래픽 특성과는 완전히 다른 장기간 의존성의 성질들을 가진다는 것이 증명되었다. 자기유사성은 허스트 파라미터로 특성화 될 수 있다. 이러한 관점에서 본 논문에서는 데이터 트래픽 특성이 서로 다른 다수의 데이터 트래픽의 통합되어 통신망에 입력되는 경우 주요 파라미터인 허스트 파라미터의 변화를 다양한 환경 하에서 분석하였고 이를 시뮬레이션 결과와도 비교하였다. 분석결과 통계적인 특성이 서로 다른 트래픽이 통합되는 경우 입력 트래픽 중에서 가장 큰 허스트 파라미터에 의해 지배되는 방식 A 보다는 분산값의 가중평균의 형태로 표현되는 방식 B가 시뮬레이션 결과와 매우 잘 일치함을 알 수 있었다. 이러한 결과는 버퍼크기를 변화시키면서 유효 대역폭의 변화를 제시한 결과에서도 동일한 결과를 얻을 수 있었다. 마지막으로 인터넷에서 서비스 중에서 대표적인 서비스의 허스트 파라미터를 고려하여, 각 서비스 비율에 따른 유효 대역폭을 산출하

였다.

앞으로는 분산계수 및 허스트 파라미터 값은 실측을 통해 서비스 형태별로 실측할 필요가 있으며, 이러한 값들을 근간으로 보다 다양한 환경에서의 시뮬레이션이 필요할 것으로 사료된다. 또한 추후로는 3GPP2의 [10]에서 모델링한 가입자 트래픽 모델을 근간으로 하여 다양한 환경에서 가입자당 유효대역폭을 분석 및 시뮬레이션 해야 할 것이며, 이를 바탕으로 고속 데이터 시스템인 1xEV-DO/DV의 망 설계 및 디멘저닝을 수행해야 한다.

■ 참고 문헌 ■

- [1] M. E. Crovella, A. Bestavros, "Self-similarity in World Wide Web traffic: evidence and possible causes," *IEEE/ACM Transactions on Networking*, Vol.5, No.6, pp.835-846, 1997.
- [2] D. E. Duffy, A. A. McIntosh, M. Rosenstein, W. Willinger, "Statistical Analysis of CCSN/SS7 Traffic Data from Working CCS Subnetworks," *IEEE Journal on Selected Areas in Communications*, Vol.12, No.3, pp.544-551, 1994.
- [3] M. Garrett, W. Willinger, "Analysis Modeling and Generation of Self-similar VBR Traffic", *Proc. of SIGCOMM '94*, pp.269-280, August 1994.
- [4] W. Leland, M. Taqqu, W. Willinger, and D. Wilson, "On the Self-similar Nature of Ethernet Traffic", *Proc. of SIGCOMM '93*, pp.183-193, Sept. 1993.
- [5] V. Paxson, S. Floyd, "Wide area traffic: the failure of Poisson modeling," *IEEE/ACM Transactions on Networking*, Vol.3, No.3, pp.226-244, 1995.
- [6] I. Norros, "On the use of fractional Brownian motion in the theory of connectionless networks," *IEEE Journal on Selected Areas in Communications*, Vol.13, No.6, pp.953-962, 1995.
- [7] A. Patel and C. Williamson, "Effective Bandwidth of Self-Similar Traffic Source: Theoretical and Simulation Results," *Proceeding of the IASTED Conference on Applied Modeling and Simulation*, Banff, AB, July 1997.
- [8] Schuler, C. "fft_fgn". Research Institute for Open Communication Systems, GMD FOKUS,

Hardenbergplatz 2, D-10623 Berlin, German

- [9] Alejandro Quintero , Yacine Elalamy and Samuel Pierre, "Performance evaluation of a broadband wireless access system subjected to heavy load", Vol.27, Issue 9, pp.781-791, June 2004.
- [10] "1xEV-DV Evaluation Methodology-Addendum [V5]," 3GPP2 WG5 Evaluation AHG, July 16, 2001.