

인터넷에서의 데이터 트래픽 특성분석

임석구*, 이종규**

*천안대학교 정보통신학부

**LG 전자 CDMA 시스템연구소

e-mail : sklim@infocom.cheonan.ac.kr, havdream@lge.com

Analysis for Data Traffic Characteristics in Internet

Seog-Ku Lim*, Jong-Kyu Lee**

*Div. of Information & Communication Engineering, Cheon-An University

** LG Electronics Inc., CDMA System Research Lab

요 약

현재 제공되는 인터넷 서비스들의 동작 특성은 기존에 고려되던 트래픽 특성과는 완전히 다른 자기 유사성(Self-similarity)이라는 성질을 가진다는 것이 증명되었다. 자기 유사성은 장기간 의존성으로 표현되는데, 이것은 단기간 의존성 성질을 갖는 기존의 모델인 포아송(Poisson) 모델과는 상반되는 개념이다. 따라서 차세대 통신망의 설계 및 디멘셔닝을 위해서는 무엇보다도 데이터 트래픽의 주요 특성인 버스트성(Burstiness)과 자기유사성이 반영된 트래픽 모델이 요구된다. 여기서 자기유사성은 허스트 파라미터(Hurst Parameter)로 특성화 될 수 있다. 이러한 관점에서 본 논문에서는 데이터 트래픽 특성이 서로 다른 다수의 데이터 트래픽의 통합되어 통신망에 입력되는 경우 주요 파라미터인 Hurst Parameter의 변화를 다양한 환경하에서 분석하였고 이를 시뮬레이션 결과와도 비교하였다.

1. 서 론

세계 각국에서는 21 세기 지식 정보사회의 도래를 대비하여 국가 정보 하부구조를 구축하는데 인터넷을 근간으로 하고자 하는 노력이 활발히 추진되고 있다. 그러나 인터넷 사용자가 급속하게 증가하면서 인터넷은 느린 속도와 제한적인 멀티미디어 서비스, 보안 문제와 같은 사용자의 불만에 직면하게 되었고, 인터넷 주소의 고갈로 인해 보다 진일보한 새로운 형태의 서비스가 필요하게 되었다. 차세대 고속 통신망을 구축하기 위한 시도는 두 가지 방향으로 진행되고 있다고 분석된다. 첫째는 고속의 패킷 처리 능력을 가지는 기가비트 라우터(Gigabit Router)를 통하여 해결하자는 방향이다. 다른 방법은 데이터링크 계층의 스위칭 기술을 이용하여 패킷을 고속으로 전달하고자 하는 것이다.

그러나 이러한 문제 해결에 앞서 기본적으로 고려해야 할 사항은 바로 인터넷 망에 흐르는 트래픽의 특성 분석이다. 인터넷에서 지원하고자 하는 서비스가 음성 뿐만이 아닌 고속 인터넷 서비스와 동영상 서비스와 같이 광대역을 필요로 하는 서비스이기 때문에 기존의 Markovian 모델에 기초한 음성

위주의 트래픽 모델링으로는 실제 서비스 트래픽을 모사할 수 없으며, 이러한 모델링을 통한 시스템 설계와 성능 예측은 부정확한 시스템 파라미터의 설정과 잘못된 성능 분석 결과를 초래할 수 있다.

현재 제공되는 많은 인터넷 서비스들의 동작 특성은 자기 유사성(Self-similarity)이라는 기존에 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], [2], [3], [4], [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 고려하여야 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 트래픽의 통합되어 통신망에 입력되는 경우 주요 파라미터인 허스트 파라미터의 변화를 다양한 환경하에서 분석하였고 이를 시뮬레이션 결과와도 비교하였다. 또한 차세대 통신망 설계와 디멘저닝 단계에서 요구되는 채널 및 중계선과 같은 자원들의 효율적인 물량산출 및 설계를 위해 필요한 유효 대역폭(Effective Bandwidth)관점에서도 분석하였다.

2. Self-Similar 정의 및 유효 대역폭

Self-similarity 는 기존의 Fractal 과 Chaos 이론에 기초한 개념이며, 차원(Dimension)상의 서로 다른 확대비율이나 서로 다른 스케일에서 보았을 때, 동일하게 보이거나 동일하게 행동하는 자기 유사한 현상이다. 참고문헌 [6]에서 설명한 인위적인 고찰을 토대로 Self-similar 트래픽에 대한 어떠한 통찰력을 얻을 수 있다. 망 성능 관점에서 Self-Similarity 가 갖는 가장 큰 특징은 “집단화의 지속성(Persistence Of Clustering)”이다. 포아송 트래픽의 경우 집단화는 짧은 기간동안 나타나고 시간이 지나면 무디어진다. 이러한 장기간에 걸친 집단화의 완화를 예측하여 버퍼를 갖는 시스템 설계에 적용할 때, 적당한 크기의 버퍼로도 충분할 것으로 사료되지만 버스트한 특성 자체 때문에 실제 큐 길이는 포아송 트래픽의 경우로 예상된 길이보다 증가할 수 있다.

2.1 Self-Similar 정의

Self-similar 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 $H(0.5 \leq H \leq 1)$ 을 갖는 통계적 Self-Similar 프로세스라 한다. 이 관계는 다음과 같은 세가지 조건에 의해서 표현될 수 있다.

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

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

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

여기서 파라미터 H 는 Hurst 파라미터 또는 Self-Similarity 파라미터로서 알려져 있는데, 이것은 self-similarity 의 주요 척도이다. 보다 정확히 말하면, H 는 통계적 현상의 지속성에 대한 척도이고 Stochastic Process 의 Long-Range 의존성의 길이에 대한 척도이다. H 값이 0.5 이면 Self-Similarity 가 없음을 의미하고 1 에 가까우면 지속의 정도 또는 Long-Range 의존성의 정도가 크다는 것을 의미한다.

2.2 유효 대역폭(Effective Bandwidth)

기존의 음성 트래픽과는 달리 자기유사성을 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 자기유사성이 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 디멘저닝 시에 유용하게 사용될 수 있다.

본 논문에서는 유효 대역폭을 산출하기 위한 해석적 모델로써 Norros 가 제안한 자기유사 트래픽 소스에 대한 유효 대역폭(C)는 식(4)와 같이 정의된다[6].

$$C = m + \left[\kappa(H) \sqrt{-2 \ln \epsilon} \right]^{1/H} (ma)^{1/(2H)} B^{-(1-H)/H} \tag{4}$$

- m : 트래픽 스트림의 평균 비트율(bits/sec)
- a : 트래픽 스트림의 분산계수(Variance Coefficient)(bit-sec)
- H : Hurst 파라미터($0.5 \leq H \leq 1$)
- B : 버퍼 크기(bits)
- ϵ : Target Cell Loss Rate
- $\kappa(H) = H^H (1-H)^{1-H}$

식(4)를 이론적으로 유도하는데 있어서 기반이 되는 두 가지 가정사항은 매우 많은 iid.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.

3. Self-Similar 트래픽 파라미터 및 통합 트래픽 분석

Norros 가 제시한 유효 대역폭 공식은 자기 유사 트래픽의 통계적인 다중화에 관련하여 이론적인 기초를 제공한다. 여기서 중요한 파라미터는 분산계수 a 와 Hurst 파라미터 H 인데 평균 m 과 함께 트래픽 소스의 통계적인 특성을 나타내는 파라미터이다.

3.1 Homogeneous Traffic

평균이 m 이고 분산계수가 a 이며, Hurst 파라미터가 H 인 두 개의 동일한 트래픽 스트림이 다중화되는 경우 통합 트래픽 스트림은 평균은 $2m$ 이고 분산계수와 Hurst 파라미터는 각각 a 와 H 이다. 이러한 결과는 트래픽 소스가 서로 독립적이라는 사실로부터 유도 가능한데, 평균은 더하면 되고, 분산도 더하면 되므로 분산계수는 a 가 된다.

3.2 Heterogeneous Traffic

평균이 m_1 이고 분산계수가 a_1 이며, Hurst 파라미터가 H 인 트래픽 스트림과 평균이 m_2 이고 분산계수가 a_2 이며, Hurst 파라미터가 H 인 트래픽 스트림이 다중화되는 경우 통합 트래픽 스트림은 평균은 $m=m_1+m_2$ 이고 분산계수 a 는 $a = \frac{m_1 a_1 + m_2 a_2}{m_1 + m_2}$ 이다. 이러한 결과는 트래픽 소스가 서로

독립적이라는 사실로부터 유도 가능한데, 평균은 더하면 되고, 분산도 더하면 되므로 분산계수는 가중평균의 형태가 된다.

3.3 Multiple Source

평균이 m 인, 분산계수 a , Hurst 파라미터 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(단일) 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 1.에는 자기 유사 트래픽 소스의 통합에 관련된 이론적인 결과를 요약하였다.

표 1. 통합된 자기유사 트래픽의 통계적인 특성

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

3.4 허스트 상수(Hurst Parameter)

Hurst 파라미터가 H_1 인 트래픽 스트림과 H_2 인 트래픽 스트림이 다중화되는 경우 통합 트래픽 스트림의 Hurst 파라미터 H 를 결정하는 것은 매우 중요하다. 참고문헌 [8]에서 이와 같은 경우의 통합 트래픽의 허스트 상수는 $H = \max(H_1, H_2)$ 임을 제시하였다. 이는 직관적으로 장기간에 걸쳐 통합 트래픽 스트림의 분산의 감쇄는 큰 값의 H 값에 지배되는 것처럼 보이나 이와 같이 통계적 특성이 서로 다른 트래픽이 통합되는 경우 평균이나 분산값의 가중평균의 형태로 표현되는 것이 실제의 트래픽 특성을 매우 잘 반영할 수도 있다. 특히, 입력되는 트래픽의 평균이 매우 상이한 경우에는 평균값이 큰 트래픽의 허스트 상수값에 의해 통합 트래픽의 허스트 상수값은 좌우될 것으로 예측할 수 있다.

이러한 관점에서 본 논문에서는 표 2 와 같이 2 가지 근사식을 적용한 경우를 비교하기 위하여 참고문헌 [9]에서 제안한 방법을 이용하여 Self-Similar 트래픽을 발생시킨다.

트래픽 발생기는 FGN(Fractional Gaussian Noise) 프로세스의 Power Spectrum 을 평가하기 위하여 FFT(Fast Fourier Transform)를 이용하였다.

표 2. 통합 트래픽의 허스트 파라미터 근사식

Case A	Case B
$H_g = \max(H_1, H_2)$	$H_g = \frac{\sum_i v_i H_i}{\sum_i v_i}$

4. 시뮬레이션 및 Numerical Example

그림 1 에는 2 개의 트래픽이 통합되는 경우를 나타내었는데, 입력되는 트래픽의 평균이 각각 $m=75.5\text{Mbps}$, 분산계수 $a=1.7 \times 10^7$ 이고 허스트 상수값은 하나는 $H=0.8$ 로 고정시키고 다른 트래픽의 허스트 상수값은 0.7 에서 0.9 까지 변화시킨 경우를 시뮬레이션 결과와 함께 나타내었다. Case B 가 시뮬레이션 결과와 일치함을 알 수 있다.

그림 2 에는 입력되는 총 입력 트래픽이 $m=155\text{Mbps}$, 분산계수 $a=1.7 \times 10^7$ 이고 $H=0.8, 0.9$ 인 경우 트래픽의 상대적인 비율을 변화시킨 경우를 시뮬레이션 결과와 함께 나타내었다. Case B 가 시뮬레이션 결과와 일치함을 알 수 있다.

그림 1 과 그림 2 의 결과로부터 통합 트래픽의 H 는 입력 트래픽 중에서 가장 큰 H 에 의해 지배되는 Case A 보다는 분산값의 가중평균의 형태로 표현되는 Case B 와 매우 잘 일치함을 알 수 있었다.

그림 3 에는 표 2 에 제시한 근사식을 적용한 경우, 버퍼크기(B)의 변화에 따른 유효 대역폭의 변화를 나타내었다. 여기서 통합되는 각각의 트래픽은 $m=75.5\text{Mbps}$, $a=1.7 \times 10^7$, $H=0.8$ 과 0.9 이다. 또한 $\epsilon=10^{-6}$ 이며, 버퍼 크기는 $B=10^8 \sim 10^9$ 로 변화하였다. 버퍼크기가 증가함에 따라 유효 대역폭은 총 입력 트래픽인 155Mbps 로 접근함을 알 수 있다. 표 2 의 근사식 중에서 Case B 가 시뮬레이션 결과 매우 근접함을 알 수 있었다. 이러한 결과는 Case B 의 허스트 파라미터가 실제의 값과 근접하고 있기 때문에 나타난 결과로 분석할 수 있다.

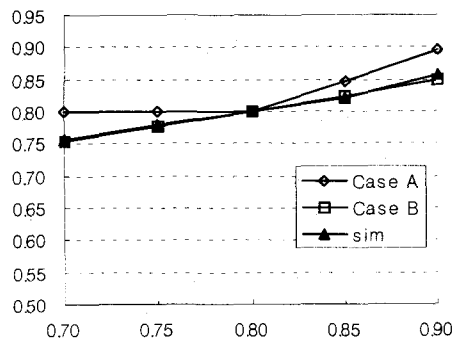


그림 1. 트래픽이 통합되는 경우의 H 의 변화

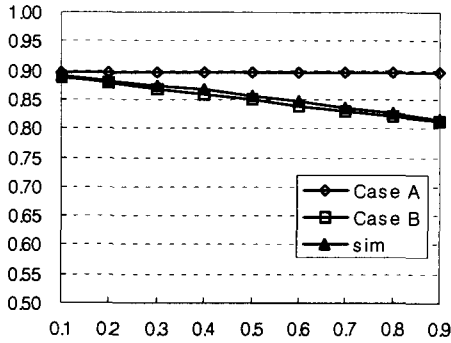


그림 2. 트래픽의 비율에 따른 통합 트래픽의 H의 변화

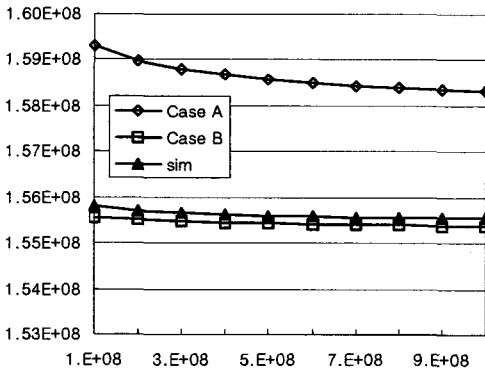


그림 3. 버퍼크기에 따른 유효 대역폭의 변화

5. 결론

현재 제공되는 인터넷 트래픽은 자기 유사성(Self-similar Property)이라는 기존에 고려되던 트래픽 특성과는 완전히 다른 장기간 의존성의 성질들을 가진다는 것이 증명되었다 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 고려하여야 할 사항이라고 판단된다.

자기유사성은 Hurst 파라미터로 특성화 될 수 있다. 이러한 관점에서 본 논문에서는 데이터 트래픽 특성이 서로 다른 다수의 데이터 트래픽의 통합되어 통신망에 입력되는 경우 주요 파라미터인 Hurst 파라미터의 변화를 다양한 환경하에서 분석하였고 이를 시뮬레이션 결과와도 비교하였다. 분석결과 통계적인 특성이 서로 다른 트래픽이 통합되는 경우 입력 트래픽 중에서 가장 큰 Hurst 파라미터에 의해 지배되는 Case A 보다는 분산값의 가중평균의 형태로 표현되는 Case B 가 시뮬레이션 결과와 매우 잘 일치함을 알 수 있었다. 이러한

결과는 버퍼크기를 변화시키면서 유효 대역폭의 변화를 제시한 결과에서도 동일한 결과를 얻을 수 있었다.

앞으로는 실제 인터넷 상에서 특정한 트래픽의 Hurst 파라미터를 근간으로 보다 다양한 환경에서의 시뮬레이션이 필요할 것으로 사료된다.

참고문헌

- [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] William Stallings, "High-Speed Networks and Internets : Performance and Quality of Service" Second Edition, Prentice Hall, pp. 220-222, 2002
- [7] 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.
- [8] Z. Deng, "Modeling and Analysis of Self-Similar Video Traffic," M. Sc. Thesis, Department of Computer Science, University of Saskatchewan, September 1996
- [9] Schuler, C. "fit_fgm". Research Institute for Open Communication Systems, GMD FOKUS, Hardenbergplatz 2, D-10623 Berlin, Germany.