

시뮬레이션에 의한 데이터 트래픽 발생모델에 관한 연구

구혜련, 임석구
천안대학교, 정보통신학부

A Study on the Data Traffic Generation Model by Simulation

Hye-Ryun Koo, Seog-Ku Lim
Div. of Information & Communication Engineering, Cheonan
University

요 약

자기 유사적인 트래픽 성질이 차세대 고속 통신망의 성능에 미치는 영향을 예측 분석하기 위해서는 자기유사 트래픽을 정확하고 효율적으로 모델링하기 위한 연구 및 분석 결과는 인터넷 망이나 새롭게 구현될 시스템에서 반드시 고려하여야 할 사항이라고 판단된다. 이러한 관점에서 본 논문에서는 인터넷 데이터 트래픽이 갖는 통계적 특성인 자기유사 트래픽을 발생할 수 있는 시뮬레이션 모델을 제시하고 각 방식의 성능을 분석한 후 이를 기초로 데이터 트래픽이 입력되는 모듈들을 실험할 수 있는 트래픽 발생 방식을 제안하고 마지막으로 결론을 맺는다.

1. 서론

차세대 고속 통신망을 구축하기 위한 시도는 두 가지 방향으로 진행되고 있다. 첫째는 고속의 패킷(Packet) 처리 능력을 가지는 기가비트 라우터(Gigabit Router)를 통하여 해결하려는 방향이며, 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. 지금까지 이루어진 음성 중심의 트래픽 모델링에서는 시간당 평균 호 발생률, 발생 간격의 분포, 호 유지시간(Holding Time), 그리고 쾨번시(Busy Hour)를 결정하는 것이 주요 과제였으며 이를 이용한 트래픽 엔지니어링은 음성호의 블러킹 확률과 지연시간을 최소화하기 위한 충분한 호 자원 확보에 중심을 두었던 것이 사실이다. 그러나 초고속 광대역 통신망의 등장과 함께 나타난 다양한 멀티미디어 서비스와 무선 이동통신에 대한 IS-95B 및 95C의 적용에 따른 데이터 통신 서비스가 활성화됨에 따라 데이터 트래픽을 고려한 트래픽 엔지니어링의 필요성이 대두되었

으며 이러한 서비스에서 발생하는 트래픽은 음성의 경우와 필요한 BHCA(Busy Hour Call Attempt)나 호 보유시간 뿐만 아니라 호의 통계적 특성을 알고 있어야만 QoS(Quality of Service)를 보장하는 효과적인 자원 확보를 위한 모델링이 가능하다.

현재 제공되는 많은 인터넷 서비스들의 동작 특성은 자기 유사성(Self-similar Property)이라는 기준에 고려되던 트래픽 특성과는 완전히 다른 성질들을 가진다는 사실이 밝혀지면서 이에 대한 연구가 최근 들어 활발히 진행 중이다. 1990년대 초반에 LAN(Local Area Network)에서의 트래픽이 장기간 의존성(LRD: Long-Range Dependency)을 가진다는 것이 발견된 이후 많은 MAN(Metropolitan Area Network) 트래픽, WAN(Wide Area Network) 트래픽, VBR(Variable Bit Rate) 비디오 트래픽, WWW(World Wide Web) 트래픽 및 SS No.7(Signaling System Number 7)에서의 트래픽 흐름들도 또한 장기간 의존성의 성질을 가진다는 것이 증명되었다[1],[2],[3].

자기 유사적인 트래픽 성질이 차세대 고속 통신망의 성능에 미치는 영향을 예측 분석하기 위해서는 자기유사 트래픽을 정확하고 효율적으로 모델링하기 위한 연구 및 분석 결과는 인터넷

망이나 새롭게 구현될 시스템에서 반드시 고려하여야 할 사항이라고 판단된다.

따라서 본 논문에서는 Self-Similar 트래픽을 발생시킬 수 있는 시뮬레이션 모델을 제시하고 각 방식의 성능을 분석한 후 이를 기초로 데이터 트래픽이 입력되는 모듈들을 실험할 수 있는 트래픽 발생 방식을 제안하고 마지막으로 결론을 맺는다.

2. 자기유사성을 갖는 호 발생 모델

2.1 트래픽 발생 방법

LRD(Long-range Dependence)는 자기유사 프로세스의 중요한 특성이며, 시뮬레이션에서 자기유사 프로세스는 종종 여러 도착 프로세스들을 합성함으로써 생성 가능하다. 많은 자기유사 트래픽 발생기가 제안되었는데, Fractional Stochastic Process, Fractal, Chaotic Map, Superposed On/Off Source, Wavelet Transform 등의 방법을 이용하였다. 본 논문에서는 참고문헌 [4], [5]에서 제안한 방법을 이용한다. 여기서 트래픽 발생기는 FGN 프로세스의 Power Spectrum을 평가하기 위하여 FFT(Fast Fourier Transform)를 이용하였다.

자기유사 트래픽을 발생하는 과정은 다음과 같다.

- ① `fft_gen`을 이용하여 단위 슬롯 당 호 도착개수를 발생시킨다. 여기서 슬롯의 길이는 1초임을 가정한다. 따라서 `fft_gen`은 4096개의 단위 슬롯 당 호 도착개수를 발생시킨다. `fft_gen`에서 첫 번째로 발생하는 엔트리는 $t=[0, 1]$ 사이에 도착하는 호의 개수를 의미하며, 두 번째 엔트리는 $t=[1, 2]$ 사이에 도착하는 호의 개수를 의미하고, 4096번째 발생하는 엔트리는 $t=[4095, 4096]$ 사이에 도착하는 호의 개수를 의미한다.
- ② 슬롯 당 호의 도착 개수를 각 슬롯 지속시간동안에서의 도착 시간으로 변환한다. 이 과정에서 여러 가지 방법이 가능하다. 예를 들어 호의 도착시간간격은 균일한 경우, 슬롯 당 평균 호의 도착수가 200(호/초)이면 호의 도착시간간격은 $0.005(=1/200)$ 이다.
- ③ 시뮬레이션 모델에 적용하기 위하여 각 호의 도착을 Scheduling한다. Scheduling은 event 노드를 이용하였다.
- ④ 시뮬레이션 모델에서는 이를 처리한다. 이를 시뮬레이션 전용언어인 SLAM II 네트워크 모델로 적용하였을 경우를 그림 1에 나타내었다.

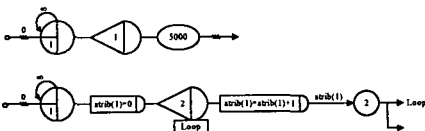


그림 1. 자기유사 트래픽을 발생시키는 모델

2.2 도착 수를 도착 시간으로의 변환과정

`fft_gen` 발생기를 이용하여 자기 유사적인 트래픽을 발생시키기 위해서는 총 12가지의 파라미터를 입력하여야 하는데, 이 중에서 가장 중요한 5가지는 Hurst 파라미터, 평균, 분산, 샘플 수, 그리고 출력 파일명이다. 예를 들어 `fft_gen`이 211.053, 81.867, 4.211, -38.210의 도착 수를 발생시킨 경우 bin 0에는 211이 할당되고, bin 1에는 82, bin 3에는 4, bin 4에는 0이 각각 할당된다. Bin 폭이 0.1(초)인 경우 Bin 0에 할당된 211개의 도착 수는 균일한 분포인 경우 첫 번째 도착시간은 $t=0.0004739(=0.1/211)$, 두 번째 도착시간은 $t=0.0009479$ 등이다. 그러나 실제 데이터의 도착시간 간격은 균일하지 않으므로 본 논문에서는 4가지 방법을 비교 분석하고자 한다.

2.1.1 버스트성이 가장 강한 경우(방법1)

이 경우는 각 bin의 시작시점에 모든 데이터가 도착하는 경우이다.

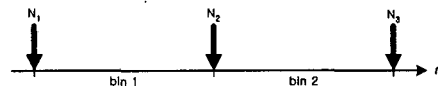


그림 2. 버스트성이 가장 강한 경우

2.1.2 어느 정도의 버스트성이 있는 경우(방법2)

이 경우는 각 bin의 일정시점마다 데이터가 버스트하게 도착하는 경우로서 비디오 트래픽을 모델링하는데 많이 사용되는 방법이다.

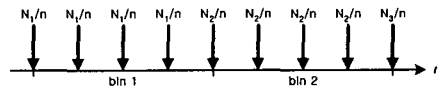


그림 3. 어느 정도의 버스트성이 있는 경우

2.1.3 도착시간간격이 균일한 경우(방법3)

이 경우는 각 bin에 할당된 데이터가 bin에 균일하게 분포하여 도착하는 경우로서 실제의 데이터 도착과는 거리가 있는 방법이다.

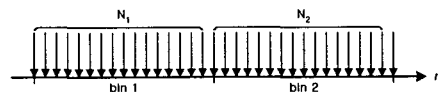


그림 4. 도착시간 간격이 균일한 경우

2.1.4 지수분포를 따르는 경우(방법4)

이 경우는 각 bin에 할당된 데이터가 아래와 같은 밀도함수와 분포함수에 따라 도착하는 경우로서 실제의 큐잉 동작을 묘사하는데 매우 유용하다[6].

$$f(x) = \frac{e^{-x}}{1 - e^{-1}} \quad (1)$$

$$F(x) = \frac{1 - e^{-x}}{1 - e^{-1}} \quad (2)$$

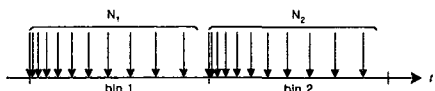


그림 5. 도착시간 간격이 지수분포를 따르는 경우

본 논문에서는 식 (1)의 분포를 갖는 랜덤변수를 발생시키기 위해 역변환(Inverse Transform)법을 사용하였다. 우선 역변환법에 대해 살펴보면 다음과 같다.

발생시키려는 연속형 확률변수를 X, 이것의 분포함수를 F라 하자. 여기에서 F는 연속형이며 증가함수 ($0 < F(x_1) \leq F(x_2) < 1$, $x_1 < x_2$)가 된다. 그러면 다음의 알고리즘은 분포함수 F를 가지는 확률변수 X를 발생하게 된다.

- 균일분포 U(0, 1)의 난수를 발생한다.
- $X = F^{-1}(U)$ 로 놓고 그 값을 되돌려 준다.

여기서 F^{-1} 은 F의 역함수를 나타내며 $0 \leq U < 1$ 이며 F의 범위가 [0, 1]이기 때문에 $F^{-1}(U)$ 은 항상 존재한다. 따라서 식 (2)의 분포를 가지는 확률변수를 발생시키기 위해 식 (2)의 역함수를 구하면 다음과 같다.

$$X = F^{-1}(U) = -\ln[1 - (1 - e^{-1})U] \quad (3)$$

2.3. 시뮬레이션 모델

그림 6에는 시뮬레이션 모델을 나타내었는데, 하나의 트래픽 소스는 하나의 출력링크에 의해 서비스되는 공통 버퍼에 입력된다. 시뮬레이션에서 설정되어야 할 파라메타로는 네트워크 구성 측면에서는 버퍼 크기와 출력링크 용량이고 트래픽 소스의 측면에서는 평균, 분산계수, Hurst 파라메타이다. 버퍼의 크기와 출력링크의 용량은 명확하게 정해지며, 시뮬레이션 모델에 사용된 입력 트래픽 소스의 평균은 $m=525(\text{cells/sec})$, 분산계수는 $\sigma=23.9(\text{cells-sec})$, 출력링크의 용량 $C=600(\text{cells/sec})$, Hurst Parameter는 $H=0.83$ 으로 설정하였다. Hurst 파라메타는 원하는 Hurst 파라메타를 갖는 트래픽을 발생하여 이를 입력시키도록 구성한다. 셀 손실율(CLR: Cell Loss Rate)은 시뮬레이션에서

입력 파라메타로 사용하지 않으며, 이는 시뮬레이션 출력으로 얻을 수 있는 결과이다.

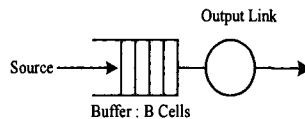


그림 6. 시뮬레이션 모델

3. 시뮬레이션 결과

앞에서 설명한 4가지 방법의 성능을 그림 7~그림 9에 나타내었다. 그림 7에는 버퍼 크기의 변화에 따른 셀 손실률을 나타내었는데, 예측한 바와 같이 가장 버스트한 트래픽을 발생시키는 방법1의 셀 손실률이 가장 높으며, 그 다음은 방법2 순이며, 방법3과 방법4는 거의 비슷한 셀 손실률을 보여주고 있다. 그림 8에는 버퍼 크기의 변화에 따른 큐 길이의 변화를 나타내었는데 모든 구간에 균일하게 트래픽이 입력되는 방법3이 큐 길이가 가장 작으며, 방법1에서의 큐 길이가 가장 크다. 그림 9에는 버퍼 크기의 변화에 따른 출력링크의 점유율의 변화를 나타내었는데, 예측한 바와 같이 방법3이 가장 효율적으로 출력링크를 사용한 반면, 버스트성이 가장 강한 방법1은 매우 비효율적으로 출력링크를 사용함을 알 수 있다.

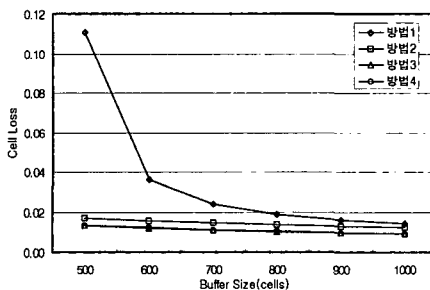


그림 7. 셀 손실률의 변화

실제 특정 서비스시스템의 성능을 확인하기 위한 실험에서 입력 트래픽 발생 형태를 선택하는 것은 매우 중요하다. 분석한 바와 같이 방법1은 실제 트래픽의 상황과는 거리가 있어서 이를 기준으로 시스템을 설계하는 경우 Over Engineering하는 결과를 초래할 것이며, 일정한 트래픽이 인가되는 방법3을 이용하여 설계하는 경우에는 오히려 Under Engineering하는 결과를 낳을 것이다.

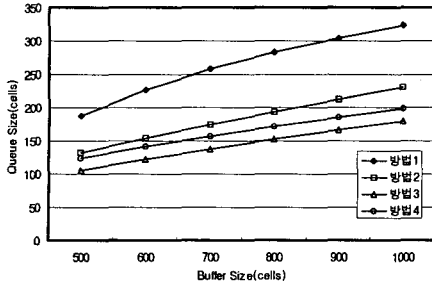


그림 8. 큐 길이의 변화

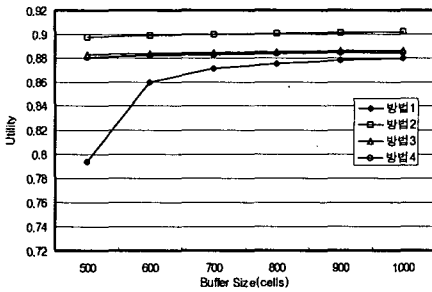


그림 9. 출력링크 점유율의 변화

비디오 트래픽을 모델링하는데 사용되는 방법2는 특정 서비스 시스템이 비디오 트래픽만을 처리하는 경우에는 좋은 방법이 된다. 그러나 실제 트래픽은 음성, 데이터, 화상 등과 같이 멀티미디어 환경에서 동작할 것이므로 이러한 관점에서 방법4가 가장 적합한 트래픽 발생방법이 될 것이다.

4. 결론

현재까지 이루어진 음성 중심의 트래픽 모델링에서는 시간당 평균 호 발생률, 발생 간격의 분포, 호 유지시간(Holding Time), 그리고 최번시(BHCA: Busy Hour Call Attempt)를 결정하는 것이 주요 과제였으며 이를 이용한 트래픽 엔지니어링은 음성호의 Blocking 확률과 지연시간을 최소화하기 위한 충분한 호 자원 확보에 중점을 두었다. 그러나 초고속 광대역 통신망의 등장과 함께 나타난 다양한 멀티미디어 서비스와 무선 이동통신에 대한 IS-95B 및 95C의 적용에 따른 데이터 통신 서비스가 활성화됨에 따라 데이터 트래픽을 고려한 트래픽 엔지니어링의 필요성이 대두되었으며 이러한 멀티미디어 서비스에서 발생하는 트래픽은 음성의 경우에 필요한 BHCA나 호 보유시간 뿐만 아니라 호의 통계적 특성을 알고 있어야만 QoS(Quality of Service)를 보장하는 효과적인 자원 확보를 위한 모델링이 가능하다.

IS-95B 및 95C 시스템에서 지원하고자 하는 서비스가 음성뿐만이 아닌 고속 인터넷 서비스와 동영상 서비스와 같이 광대역

을 필요로 하는 서비스이기 때문에 기존의 Markovian 모델에 기초한 음성 위주의 트래픽 모델링으로는 실제 서비스 트래픽을 묘사할 수 없으며, 이러한 모델링을 통한 시스템 설계와 성능 예측은 부정확한 시스템 파라미터의 설정과 잘못된 성능 분석 결과를 초래할 수 있다. 따라서 시스템 설계 및 성능 분석의 수행을 위해서 광대역 서비스 트래픽의 특성을 보다 정확하게 모델링할 수 있어야 한다.

1990년대 초반에 이더넷 트래픽이 장기간 의존성(LRD: Long-Range Dependency)을 가진다는 것이 발견된 이후 많은 MAN 트래픽, WAN 트래픽, VBR 비디오 트래픽, WWW 트래픽 및 SS No.7에서의 트래픽 흐름들도 또한 장기간 의존성의 성질을 가진다는 것이 증명되었다. 이러한 장기간 의존성 성질을 표현하기 위한 모델로는 자기유사(Self-similar) 모델이 있는데, 자기 유사적인 트래픽 성질이 차세대 고속 통신망의 성능에 미치는 영향을 예측 분석하기 위해서는 자기유사 트래픽을 정확하고 효율적으로 모델링하기 위한 방법에 대한 연구 및 분석 결과는 인터넷 망이나 새롭게 구현되어지는 시스템에서 반드시 고려하여야 할 사항이라고 판단된다.

이러한 관점에서 본 논문에서는 Self-Similarity가 시스템 성능에 미치는 영향을 분석하기 위하여 실질적으로 데이터 트래픽이 입력되는 시스템 모듈들을 실험할 수 있는 트래픽을 발생시키기 위한 시뮬레이션 모델에 대해 연구하였다. Self-Similar 트래픽을 발생시킬 수 있는 방법으로서 4가지 방식을 제안하였는데, 이들은 각각 트래픽의 버스트성을 고려하였다. 각 방식의 성능을 비교 분석한 결과, 패킷 데이터의 도착시간간격이 지수분포를 따르는 경우인 방식4가 가장 적합한 방식임을 제시하였다.

추후에는 실질적인 데이터 트래픽의 특성 파라미터를 측정할 후, 제안한 방식으로 시뮬레이션을 수행할 필요가 있다.

참고문헌

- [1] 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 (March, 1994): 544-551.
- [2] M. Crovella and A. Bestavros, "Self-similarity in World-wide Web Traffic: Evidence and Possible Causes", *Proc. of ACM Sigmetrics Conference on Measurements and Modeling of Computer Systems* (May 1996): 160-169.

- [3] Robert Geist and James Westall, "Simulation Modeling Self-Similarity in Network Traffic", Proc. of ICC'98, (August 1998): 249-254.
- [4] M. Garrett, W. Willinger, "Analysis Modeling and Generation of Self-similar VBR Traffic", Proc. of ACM SIGCOMM 94 (August 1994): 269-280.
- [5] Schuler, C. [†]ff_t_gn . Research Institute for Open Communication Systems, GMD FOKUS, Hardenbergplatz 2, D-10623 Berlin, Germany.
- [6] V. Paxson. [†]Fast Approximation of Self-Similar Traffic, Technical Report LBL-36750, Lawrence Berkeley Laboratory and EECS Division, University of California, Berkeley, April 1995.