

자기유사성을 고려한 VoIP 트래픽 생성 시뮬레이션 방법의 연구

Study of the Simulation of VoIP Traffic Generation with Considering Self-Similarity

이계신, 김재범, 이창현, 김윤배
성균관대학교 시스템 경영공학부

Abstract

VoIP는 인터넷 프로토콜(IP)를 이용하여 음성을 데이터 packet처럼 전송하는 것을 의미한다. 최근 VoIP 기술의 도입으로 기존 망 성능 관리에 대한 관심이 높아지고 있다. 보다 원활한 기술 구현을 위해서는 VoIP 트래픽에 대한 체계적인 분석과 위험성 검증할 수 있는 도구가 필요하다. 또한 기존의 트래픽 시뮬레이션 기법에서 실제 망에서의 자기유사성을 적용한 사례가 적다는 것 또한 본 연구가 행하여진 동기이다. 본 연구에서는 자기유사성을 반영하여 소량의 샘플을 갖고 전체 VoIP 망 트래픽을 생성할 수 있는 방법론을 개발하고자 시도하였다.

1. Introduction

VoIP는 데이터를 전송하기 위해 개발한 인터넷 프로토콜(IP)에 음성을 동시에 실어 나르는 기술로서 인터넷을 통한 전화서비스를 비롯해 팩스, 영상회의, 멀티미디어 메시징 등의 부가서비스를 지원할 수 있다. 근래에는 컴퓨터전화통합(CTI)분야나 통합 메시징 시스템(UMS) 등이 VoIP 기술과 결합하여 인터넷 통합솔루션을 지원하고 있다. VoIP 이용자는 시내전화회선을 통해 바로 인터넷에 접속할 수 있고 원하는 지역의 이용자와 통화할 수 있다.[3]

이러한 VoIP 기술을 구현하는 과정에서 기술 특성상 망 성능관리가 중요한 문제로 부각되고 있다. 인터넷 트래픽에 추가적인 음성 데이터 트래픽이 추가됨으로 해서 생기는 대역폭의 확대는 필연적이다. 패킷의 크기가 커질 경우 패킷 조립에 의하여 대기 시간과 처리시간의 증가하는 것은 당연하기 때문이다. 이는 기본적으로 VoIP Packet Format 자체의 용량이 크다는데 기인한다. 따라서, 임의의 기업에서 기존 인터넷망에 새로운 VoIP 시스템을 도입하여 이를 인터넷 통합 솔루션으로서 이용하려 할 때에 기

존 망에 용량 문제가 발생하게 되는 것이다.[9]

다양한 VoIP 프로토콜의 방식 중에서 기존에 가장 많이 쓰여지고 있는 H.323 프로토콜이다. 현재 H.323 protocol을 기반으로 서비스 중인 대부분의 기업들은 기간 통신 사업자의 PSTN 라인과 IP망을 대역하여 이를 기반으로 사업을 하고 있다. 대부분의 기업들의 투자비는 최초 장비 구매 후 망을 구성하기 위한 IP 망 및 PSTN 라인 대역 비용에 상당 부분이 투입되고 있다. 그러므로, H.323 protocol하에서 발생하는 트래픽을 분석하고 이를 재생성할 수 있다면 이러한 오류 및 실패에 따른 비용 및 노력을 줄일 수 있을 것이다.[3][9]

본 연구에서는 단일 사용자가 만들어내는 패킷을 대상으로 이들의 통계적인 특성을 파악하여 전체 사용자가 만들어내는 트래픽을 재구현할 수 있는 시뮬레이션 기법을 개발하고자 한다. 시뮬레이션 기법의 정확성을 위해 전체 망에 접속하고자 하는 다중의 사용자가 만들어내는 실제 트래픽과 시뮬레이션 결과 값이 비교 검증될 것이다.

2. 인터넷 트래픽의 자기유사성

최근 인터넷 트래픽의 특성은 기존 많은 연구 ([1][2][4][8])에서 제시되고 있듯이 Self-Similarity (자기 유사성)로 대변할 수 있다. Self-Similar 현상은 시간 스케일(milliseconds, seconds, minutes 등)을 달리 하더라도 그 구조적 유사성이 유지된다는 것을 의미한다. 이것은 Poisson Process나 Markovian Arrival Process와 비교하면 명확하게 설명될 수 있다.

만약, 인터넷 트래픽이 Poisson Process를 따른다고 한다면 Short-Range Dependence(SRD)의 성격에 의해서 시간 스케일을 충분히 넓힐 경우 평균화에 의해 매끄러운 형태를 가지게 될 것이다. 하지만 실제 인터넷 트래픽에서는 광범위한 시간 스케일에서도

버스티(bursty)한 현상을 발견할 수 있다. 이로 인해 인터넷 트래픽은 SRD가 아닌 Long-Range Dependence(LRD)를 가지게 된다.([1][2])

다음은 트래픽이 자기유사성을 갖는지 검증하기 위해 자주 사용되는 방법이다.

Hurst 모수

어떤 확률과정(stochastic process) $X(t)$ 가 있을 때, 임의의 실수 $a > 0$ 에 대하여 과정 $a^{-H}X(at)$ 가 $X(t)$ 와 같은 통계적 특성을 가진다면 확률과정 $X(t)$ 는 파라미터 $H(0.5 \leq H \leq 1)$ 를 갖는 통계적 자기유사 확률과정(statistical self-similar stochastic process)이라고 한다. 이러한 관계는 다음의 세 가지 조건에 의해 표현된다.

$$E[X(t)] = \frac{E[X(at)]}{a^H}$$

$$Var[X(t)] = \frac{Var[X(at)]}{a^{2H}}$$

$$R[X(t), X(s)] = \frac{R[X(at), X(as)]}{a^{2H}}$$

$$= \frac{E[(X(t) - \mu_t)(X(s) - \mu_s)]}{\sqrt{Var[X(t)] \cdot Var[X(s)]}}$$

$$(E[X(t)] = \mu_t, E[X(s)] = \mu_s)$$

Hurst 파라미터 또는 Self-Similarity 파라미터라고 하는 파라미터 H 는 Self-Similarity의 주요 척도가 된다. 보다 정확하게 말하면, 통계적 현상의 지속성과 확률과정의 장기간 의존성(LRD)에 대한 척도가 되는 것이다. 만약, $H=0.5$ 이면 Self-Similarity가 없음을 의미하며 H 가 1에 가까워짐에 따라 지속성(persistence)과 LRD의 정도가 강해짐을 의미한다.

Variance-Time Plot

어떤 Self-Similar process의 집단화된 시계열 $X^{(m)}$ 에 대하여 분산은 m 이 충분히 클 때 아래의 식을 따른다.

$$Var[X^{(m)}] \sim \frac{Var[X]}{m^\beta}$$

위 식의 양변에 상용로그를 취하면 아래와 같이 나타낼 수 있다.

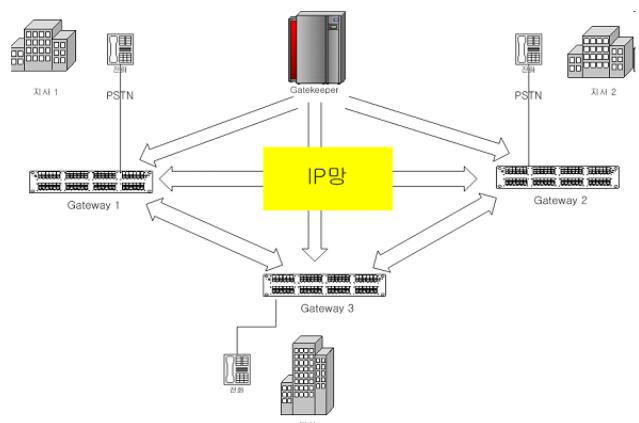
$$\log_{10}[Var(X^{(m)})] \sim \log_{10}[Var(X)] - \beta \log_{10}(m)$$

위 식에서 $\log[Var(X)]$ 은 m 과 무관한 상수이므로 $\log\text{-}\log$ 축 상에서 m 에 대하여

$\log[Var(X^{(m)})]$ 을 그린다면 기울기가 $-\beta$ 인 직선이 나타날 것이다. 기울기가 -1과 0사이의 값을 가지면 Self-Similarity를 의미한다.

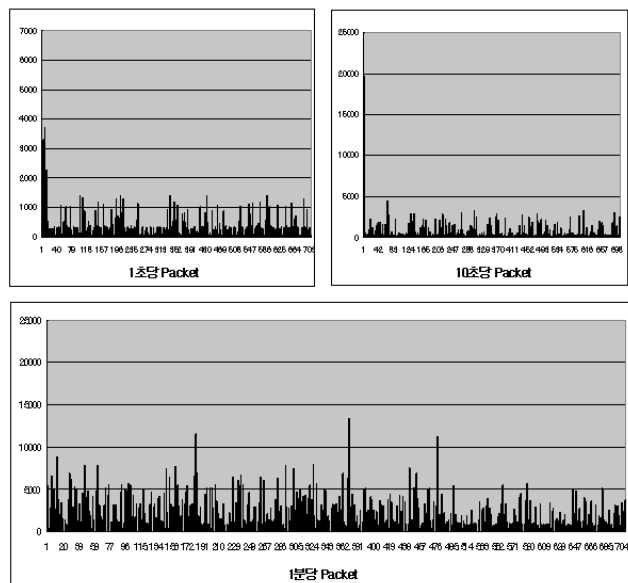
3. Voip 트래픽 분석

본 연구의 VoIP traffic 분석을 위해 I사의 VoIP 시스템이 있는 H사 IDC에서 DNA323과 agilent advisor 라는 프로그램을 이용하여 2003년 8월 1일부터 동년 동월 5일까지의 packet을 수집하였다. I사 VoIP시스템의 Hub에는 1대의 게이트키퍼와 다수의 게이트웨이 그리고 H.323 Terminal 만이 존재한다.



<그림 3.1> I사 Voip 시스템

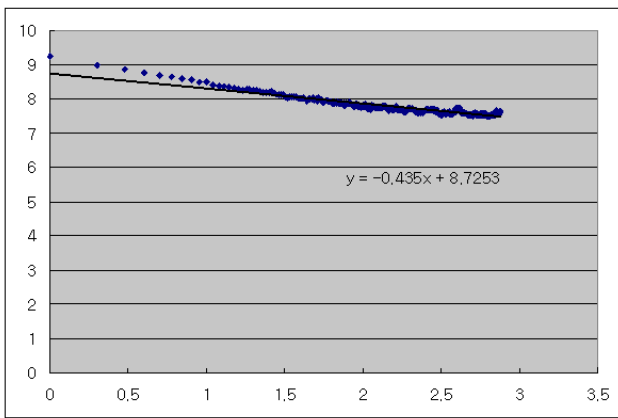
네트워크 상에서의 Packet이 Self-Similar 성질을 가지고 있다고 한다면, 실제 게이트키퍼와 게이트웨이가 있는 IP 네트워크에서 H.323 기반의 VoIP 패킷의 데이터 역시 같은 성질을 가지고 있어야 한다. 이를 확인하기 위하여 앞서 언급한 Self-Similar 특성 검증 기법을 사용하였다.



<그림 3-2> 시간 스케일에 따른 packet 양

<그림 3-2>는 실제 게이트키퍼와 게이트웨이 가 있는 IP 네트워크에서 H.323 기반의 VoIP 패킷을 보여주고 있다. 각각 시간 스케일 1초 10초 1분으로 달리 하여 표현하였다. 스케일이 다른 각각의 그래프가 모두 버스티한 경향을 나타내고 있다.

집단화 크기 m 과 $\log_{10}[Var(X^{(m)})]$ 을 log-log 축 상에 나타내었을 때, 기울기 $-\beta$ 가 -1과 0 사이의 값을 갖는다면, 그 데이터는 Hurst 파라미터 $H = 1 - (\beta/2)$ 인 자기 유사성이 있는 데이터라고 추정할 수 있다.



<그림 3-3> V-T Plot

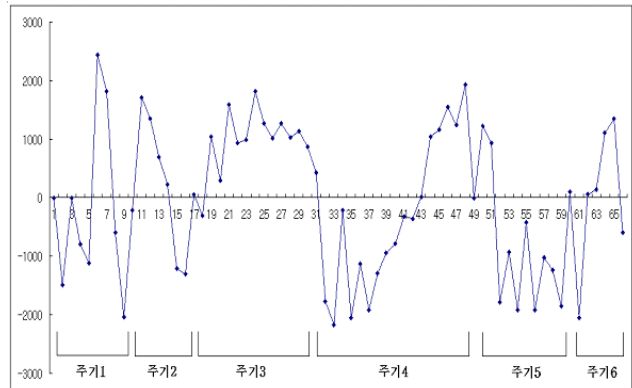
<그림 3-3>은 실제 데이터의 $\log_{10} Var[X^{(m)}]$ 과 기울기 -1을 log-log 축 상에 표시한 것이다. 위의 그래프를 보았을 때, $\log_{10} Var[X^{(m)}]$ 의 데이터의 패턴은 대략적으로 기울기가 -1보다는 크고 0보다는 작다는 것을 알 수 있다. 또한 회귀분석을 한 결과 기울기 $-\beta$ 는 -0.435로 구하였다. V/T Plot을 이용한 Hurst 파라미터를 $H_{V/T}$ 라고 하였을 때, 다음과 같이 나타낼 수 있다.

$$H_{V/T} = 1 - (\beta/2) = 1 - \frac{0.435}{2} = 0.7825$$

4. VoIP 트래픽 재생성 시뮬레이션 분석

H.323 시스템에 접속하는 사용자의 도착 간격이 Poisson Process를 따르는 지수분포를 입력 분포로 갖는다고 가정하였다. 시간대는 호의 양이 많은 오전 8시부터 저녁 23시까지로 한정하였다. 이는 실제 호의 양이 미미한 새벽 시간대의 분석이 전체 시스템에 미치는 영향은 거의 없기 때문이다. 해당 시간대

의 전화를 사용한 사용자의 전화 사용시간은 상호간에 약간씩의 차이가 있기는 하지만 대부분 비슷하며 평균 5분 정도의 시간을 갖는 지수 분포를 갖는다. 단일 사용자가 생성해내는 Packet은 대부분의 경우 처음의 Packet의 byte가 다음 Packet의 byte에 상관관계가 있는 자기 상관관계가 강하므로 이러한 자기 상관 관계를 다른 사용자에게도 적용하기 위하여 임계값 부트스트랩(Threshold Bootstrap)([5][6][7])을 사용하여 다른 사용자들의 Packet을 생성한다. 표본 평균은 초당 2175 byte의 패킷을 생성해내므로 chunk는 6이며 각각의 주기가 재추출 단위가 된다.



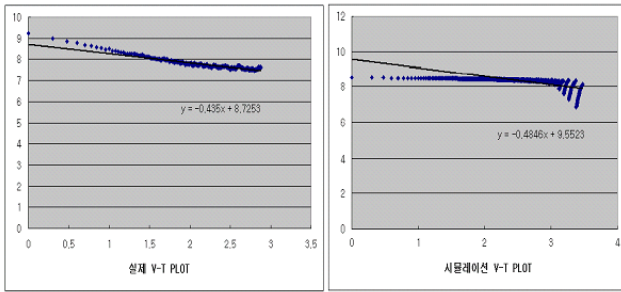
<그림 4-1> 사용자 Packet 양에 대한 부트스트랩

실제 트래픽을 본 연구의 시뮬레이션 방법으로 유사하게 재생성 해냈는지를 확인하기 위하여, 시뮬레이션 결과 값이 실제 데이터와 같은 Self-Similar 성질을 가지고 있으며, 실제 데이터의 평균값과 유사한 평균값을 가지고 있고 그 Hurst 모수 또한 유사한 값을 가지고 있는지를 검증한다.

Self-Similarity의 정도를 나타내는 Hurst 모수를 체크해 보면 더욱 상세한 정보를 알 수 있다. <표 4-1>은 실제 트래픽 데이터의 평균 Packet 과 그 Hurst 모수를 시뮬레이션 결과 평균 Packet과 Hurst 모수를 비교한 것이다.

	실제값	결과값	오차율
평균	31752.11	31458.33	0.9%
Hurst 모수	0.7825	0.7577	3.1%

<표 4-1> 실제값과 시뮬레이션 결과값 비교



<그림 4-2> 실제 데이터와 결과 값의 V-T Plot

<표 4-1>과 같이 실제값과 시뮬레이션 결과값의 평균값과 Hurst 모수의 값이 유사하다는 것을 알 수 있다.

5. Results

본 연구에서 제시한 새로운 트래픽 생성 방법은 VoIP 망에 접속하는 개인이 만들어내는 Packet의 특성을 파악하고 이 작은 양의 Sample을 사용하여 전체 트래픽을 예측하고 재생성해 내는 것이다.

본 연구에서 사용된 실측 데이터는 별정 통신의 한계 상 상당히 제한된 sample과 트래픽 양만을 사용하였고 부분적인 통계적인 검증을 하였지만 전체적인 검증을 했다고 단언할 수 없기 때문에 이러한 방법이 얼마만큼 효율적이고 유의미하다고 말할 수는 없다. 좀더 많은 수의 트래픽을 다루어 보고 또 인터넷망의 트래픽에 대한 각종 성질이 좀더 밝혀진다면 추후에 더욱 효과적인 방법을 개발할 수 있을 것이다. 본 연구는 PSTN에서 VoIP로 나아가는 이때에 적정 용량의 산출과 VoIP망 배치에 좀 더 과학적인 방법론의 하나로서 가치가 있다고 생각한다.

Reference

- [1] Crovella, M. E., Bestavros, A.(1997), Self-Similarity in World Wide Web Traffic : Evidence and Possible Cause, IEEE/ACM Transaction on Networking, 5(6)
- [2] Crovella, M. E., Lipsky, L.(1999), Simulation with Heavy-Tailed Workloads", Self-Similar Network Traffic and Performance Evaluation, pp. 89-97
- [3]James Toga, Jorg Ott,(1999), ITU-T standardization activities for interactive multimedia communication on packet-based networks : H.323 and related recommendations, computer networks 31

[4]Jim W. Roberts(2001), Traffic Theory and the Internet, IEEE Communication Magazine

[5] KIM, Y., HADDOCK, J., and WILLEMAIN, T. R. (1993a), The Binary Bootstrap: Inference with Autocorrelated Binary Data, Communications in Statistics: Simulation and Computation. 22, 205-216.

[6] KIM, Y., WILLEMAIN, T. R., HADDOCK, J. and RUNGER, G. (1993b) ,Simulation Output Analysis Using the Threshold Bootstrap, Technical Report no. 37-93-378, Department of Decision Sciences and Engineering Systems, Rensselaer Polytechnic Institute, Troy, New York.

[7] KIM, Y., WILLEMAIN, T. R., HADDOCK, J. and RUNGER, G. (1993c), The Threshold Bootstrap: A New Approach to Simulation Output Analysis, In Proceedings of the 1993 Winter Simulation , Conference, G. Evans, M. Mollaghasemi, E. Russell and W. Biles, eds. IEEE Press: Piscataway, NJ. 498-502.

[8] Paxson, V.(1997), Fast, Approximate Synthesis of Fractional Gaussian Noise for Generating Self-Similar Network Traffic, Computer Communication Review, V. 27 N.5, pp. 5-7

[9]Upkar Varshney, Andy Snow, Matt McGivern, and Christi Howard(2002), Voice Over IP, Communication of the ACM Vol. 45, No. 1