2001년 한국정보처리학회 추계 학술발표논문집 제8권 제2호

시계열 분석을 이용한 실시간 네트워크 트래픽 예측 시스템의 설계

°정상준*, 권영헌**, 최혁수*, 김종근*
* 영남대학교 컴퓨터공학과
** 세경대학 컴퓨터정보통신과

Design a Realtime Network Traffic Prediction System based on Timeseries Analysis

°Sangjoon Jung*, Younghun Kwon**, Hycksu Choi*, Chonggun Kim*
* Dept. of Computer Engineering, Yeungnam University
** Dept. of Computer Information & Network, Seakyung College

요 약

서브네트워크에서 실시간으로 통신 트래픽을 감시하고, 트래픽 정보를 바탕으로 시계열 분석을 이용해 트래픽의 변화추이를 예측할 수 있는 시스템을 설계 및 구현한다. SNMP를 이용한 MIB-II 정보를 바탕으로 하는 분석 방법은 누적 데이터를 기본으로 하는 관리 방법으로 이상 징후의 판단이 실시간 감시에는 적합하지 않은 점이 있다. 따라서, 본 논문에서는 실시간 트래픽 감시를 위해 서브네트워크에 들어오거나 나가는 트래픽의 양을 측정하여 분석하고, 이 정보를 바탕으로 특정 시점 이후의 트래픽 추이를 시계열 분석 방법을 이용하여 미래의 트래픽 양을 예측하는 알고리즘을 시스템으로 구현한다. 예측 알고리즘으로는 AR, MA, ARMA, ARIMA 모델 중에 평균 제곱 오차를 최소로 가지는 알고리즘을 선택하여 예측하도록 설계한다. 개발되는 시스템을 망 관리자가 전체 통신 네트워크의 부하 상태를 예상할 수 있게 하여 신속하고 예방적인 대응을 할 수 있다.

1. 서론

인터넷을 선두로 하여 전세계적으로 다양한 통신망이 구축, 운용되고 있으며, 이들 통신망을 이용하여 다양한 종류의 통신이 서비스 되고 있다. 이러한 통신 서비스에 대하여 가입자들은 고품질의 서비스를 요구하고 있으며, 이러한 사항을 충족하기 위해서는 통신망 운용의 관리가 필수적이다[1-2]. 인트라넷에서 네트워크를 효과적으로 관리하기 위해서는 다양한 관리 대상이 있으나, 사용자가 신뢰할 수 있는 성능을 제공하기 위해서는 네트워크 트래픽의 추이를 분석하고 대응하는 것이 중요하다. 이러한 시스템은 인트라넷에서의 특정 응용 서비스가 중지되었는지를 관리 시스템이 자동으로 파악함으로써 중단 없는 서비스 구조를 가질 수 있으며, 네트워크 트래픽 분석을 통해 네트워크의 확장 및 축소의 필요성과 그 규모를 결정할 수 있게 된다.

기존의 연구로는 MIB(Management Information Base) 정보를 활용한 관리 구조나 성능 관리에 대한 연구[6-8], SNMP를 이용한 관리 기법 연구[7], MIB-II에서 제공하는 정보를 기준으로 한 성능 분석 파라미터 도출에 관한 연구[8] 등이 있다. 효율적인 성능 관리를 수행하기 위해 시계열 분석 모델을 이용하여 수집된 정보의 특이값을 검출하는 기법에 관한 연구 및 누적된 HTTP 패킷을 기반으로 Web 네트워트래픽의 양을 예측하는 기법에 관한 연구가 활발히 진행되고 있으며, 실제 전화 교환기 상의 트래픽 이용율 예측 알고리즘이 제시되기도 하였다[9].

본 논문에서는 실시간 네트워크 관리 체계 하의 사용자 노드에서 보다 빠른 성능 분석을 위한 실시간 패킷 분석 시스템을 제안하고 구현한다. 트래픽 추이를 나타내고 미래의 트래픽을 예측하는 알고리즘은 시계열 모델[4-6]을 이용하여 실제 패킷을 분석하고 예측한다.

2. 관련 연구

망관리에 있어서 성능 관리와 함께 구성 관리를 하기 위해서는 현재 시점의 트래픽 양을 측정하고 분석하는 것이 중요하며 정확한 예측 알고리즘을 제시한다면 미래 시점의 트래픽을 예측하는 것 또한 매우 중요하다.

2.1 망 관리 시스템의 체계

일반적으로 통신 장비에 대한 관리를 수행하기 위해서 인터넷에서는 SNMP(Simple Network Management Protocol)을 사용하고 있다[1-3]. SNMP를 이용하는 시스템은 관리 정보를 모으기 위해서는 관리국이 에이전트에 주기적으로 정보를 요청(Polling)하여 에이전트가 수집한 데이터를 관리국에 제공함으로써, 네트워크 관리가 수행된다. 따라서, 네트워크 상에서 관리국에 의한 폴링은 통신량의 증가를 가져오고, 때로는 폴링을 위한 요구 횟수가 많아져 정보 전송 시간이 지연되는 경우도 발생한다. 따라서, 실제 요구된 시간에 에이전트로부터 정보를 얻지 못하는 단점이 발생할 수도 있다.

기존의 네트워크 관리 시스템이 가지는 폴링(Polling)의 요청으로 인해 통신량이 증가하는 단점을 보완하기 위해서는 요구 메시지를 최소화하는 방법으로 네트워크의 관리가 이루어져야 한다. 네트워크 관리국에서 실시간으로 네트워크 내의 패킷들을 받아들여 어떤 정보가 흘러가는가에 대한 감시를 수행하면 위에서 언급한 통신 요구 시간 및 수집 시간을 최소화할 수 있다[8][10].

2.2 시계열 자료 예측 모형

성능 분석을 수행하기 위해서는 시간의 흐름에 따라 패킷의 양이 변화하는 특징을 가진다. 이러한 형태의 자료를 시계열 자료라 하고, 이러한 시계열 자료는 시간의 흐름에 따라 일정한 패턴을 나타내는 경우가 많다[4-6]. 즉, 관찰되는 시점에 따라 이산적인 형태의 자료를 얻을 수 있고, 체계적인 분석 과정을 통해 자료의 성격을 파악할 수 있다. 시계열 모형은 다음과 같다.

(1) AR 모형

현시점 t에서의 시계열 Z_t는 p개의 과거값들의 가중합과 이들로 설명되지 않는 부분인 오차항 a_t의 선형결합으로 표현된다. 자기회귀모형은 시계열 자체에 대한 회귀 형태를 취하는 모형으로 일반 p차 AR과정을 따른 $\{Z_t\}$는 다음과 같이 나타낸다.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t$$

(2) MA 모형

이동평균 모형은 시계열 $\{Z_t\}$가 시계열 자체에 대한 회귀형태를 띠고 있는 자기회귀 과정과는 달리 현재와 과거의 백색잡음들의 가중선형결합으로 표현되는 모형으로, 일반 q차 MA과정은 다음과 같이 나타낸다.

$$Z_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

(3) ARMA 모형

어떤 시계열 데이터의 현재값 Z_t가 자신의 과거값들 Z_{t-1}, Z_{t-2}, Z_t와 오차항 a_t, 그리고 과거의 오차항들 a_{t-1}, a_{t-2}, $\cdots$, a_{t-q}에 의해 나타낼 수 있을 때, 이 시계열 데이터를 자기회귀 이동평균 모형이라 한다. 이때 가장 긴 AR의 차수 p와 MA의 차수 q를 ARMA모형의 차수라 하고 ARMA(p, q)로 나타낸다.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \cdots - \theta_q a_{t-q}$$

(4) ARIMA 모형

시계열 $\{Z_t\}$의 d차 차분한 시계열 $\{W_t = (1-B)^d Z_t\}$이 AR차수가 p, MA 차수가 q인 ARMA(p, q) 모형을 갖는다면 시계열 $\{Z_t\}$는 차수가 (p, d, q)인 자기회귀누적이동평균(ARIMA) 모형을 갖는다고 한다.

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \cdots + \phi_p Z_{t-p} + u_t - \theta_1 u_{t-1} - \theta_2 u_{t-2} - \cdots - \theta_q u_{t-q}$$

$\phi(L)Z_t = (L)u_t$에서

단, $\phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$
$\theta(L) = 1 - \theta_1 L - \cdots - \theta_q L^q$

즉, $\phi(L)\nabla^d Z_t = (L)u_t$

3. 실시간 네트워크 트래픽 예측 모델

3.1 패킷 분석 및 예측 시스템

시스템의 전체 구조는 트래픽 캡쳐기와 트래픽 분석기로 구성되어 있다. 그림 2는 전체 시스템의 구조를 보여준다. 그림 1과 같이 캡쳐기와 분석기는 하나의 시스템에 별도의 과정을 수행하도록 구성할 수 있으며, 또는 별도의 시스템에 탑재하여 독립적으로 수행될 수도 있다.

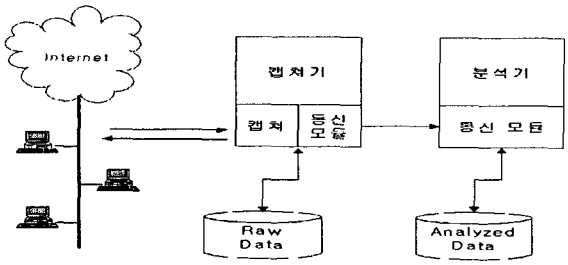

그림 1 분석 및 예측 시스템의 전체 구조

트래픽 캡처기는 네트워크를 모니터링 하기 위해 특정 서브네트워크에 들어오는 모든 패킷들을 감지하여 분석할 수 있으며, 프로토콜별 분석과 패킷 헤더의 분석을 통하여 네트워크의 상태 등을 알 수 있고, 서비스되는 패킷들의 양을 분석하여 패킷 폭주로 인한 장애를 알 수 있다. 캡처기의 핵심은 패킷 드라이버이며 패킷 드라이버에 의해 네트워크에 전송되고 있는 패킷을 수집한다. 그림 2는 캡처기 내부의 패킷 드라이버 구조를 보여주며, 상위 어플리케이션에서 패킷의 정보를 수집하여 저장한다.

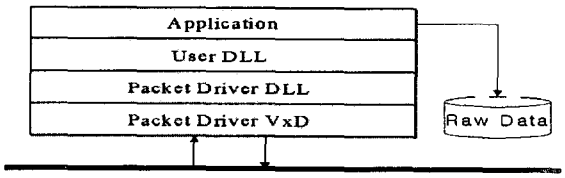

그림 2 캡처기의 패킷 드라이버 구조

트래픽 분석기는 월별, 일별, 시간별로 기초적인 분석을 수행하여 저장하며 이를 토대로 AR, MA, ARMA, ARIMA 모형을 적용하여 월, 일, 시간별 예측 데이터를 생성하는 과정을 수행한다. 캡처기와 분석기는 별도의 시스템에 탑재될 수 있으며 통신 모듈에 의한 통신을 통하여 저장된 데이터를 전송할 수 있다. 수집되는 자료의 저장 구조는 그림 3과 같다

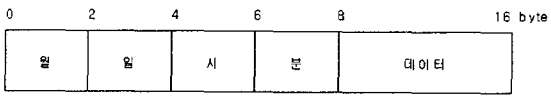

그림 3 수집되는 데이터의 저장 구조

캡처기에서 받아들여지는 패킷의 데이터는 DB에 저장되며 이 정보를 통신 모듈을 통하여 분석기에 넘겨진다. 전송된 파일을 읽어들여 기초 분석을 통하여 DB에 저장하고, DB에 저장된 데이터를 기초하여 패킷별 분석과 예측 모형별 알고리즘을 적용하여 그래프 또는 텍스트로 출력하게 된다. 전체 시스템 및 부석 시스템의 구조를 그림 4에서 보이고 있다.

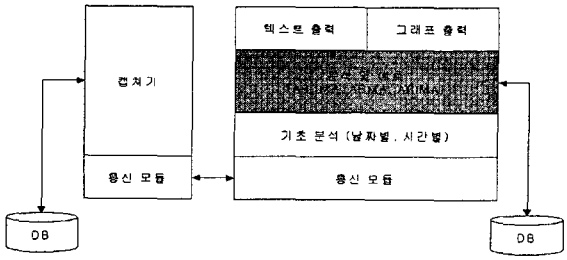

그림 4 실시간 트래픽 예측 시스템의 구조

분석기로 넘겨진 데이터는 먼저 기초 분석을 수행하는데 패킷별, 날짜별, 시간별 분석을 통해 별도의 파일로 구성하여 상위 분석 모듈에 넘겨준다. 분석 및 예측 모듈에서 이 파일을 근거로 하여 월별, 일별, 시간별 예측이 가능하다

3.2 예측 알고리즘

본 연구에서는 ARIMA 모형의 근간인 자기회귀(AutoRegressive ; AR모형) 방법과 이동평균(Moving Average ; MA모형)방법, 두 방법의 합인 자기회귀이동평균(ARMA모형) 방법, 그리고, 자기회기누적이동평균(ARIMA) 방법 등을 모두 적용하여, 트래픽 특성을 잘 나타내며 추이를 가장 잘 설명하는 모델을 최소제곱 추정법에 근거하여 기본 모델을 찾아내고, 이 모델에서의 자기 상관함수를 도출해 내고, 도출된 함수에 측정값을 입력하여 특정 시간 이후의 트래픽 양을 예측한다. 모형의 적합성 판별은 각 모형의 예측값과 실제 데이터를 비교하여 평균제곱오차(Mean Square Error)를 최소화하는 모형으로 한다. 모델을 적용하기 위한 모수의 계산 과정은 그림 5와 같이 수행한다.

Step 1. X_t(현재의 트래픽)값 측정 (t는 현재 시점)
Step 2. X_t와 X_{t-1}간의 상관계수를 구한다.
Step 3. p 또는 q 또는 p, q의 값을 설정한다.
Step 4. 모수를 추정한다.

$$(\phi_1, \phi_2, \cdots, \phi_p, \theta_1, \theta_2, \cdots, \phi_q$$

Step 5. 각 모형에 적용하여 X_{t+1}째의 값을 예측한다.
Step 6. 각 모형의 예측값과 실제 데이터를 비교하여 평균제곱오차(MSE)를 최소화 하는 모형을 선택한다.

그림 5 트래픽 예측 알고리즘 계산 과정

4. 실시간 네트워크 트래픽 예측 시스템

4.1 구현 환경

트래픽 분석용 예측 모형을 위한 모듈들을 개발하였으며 동작 확인을 위해 초기 단계의 예측 시스템을

구현해 보았다. 본 시스템은 Windows98을 운영체제로 하는 IBM 호환 PC(Intel Pentium III 450MHz)에서 MS Visual C++ 6.0을 이용하여 구현한다.

4.2 결과

네트워크의 속도가 갑지기 느려질 경우 이때 네트워크의 프로토콜 별 트래픽을 분석하거나 어떤 네트워크 프로토콜의 사용이 많은지를 알 수 있으며, 특정 시점 이후의 트래픽 양을 예측할 수 있다. 본 시스템은 예측 기능을 포함하고 있으며 그림 6, 그림 7과 같이 텍스트 출력 화면과 그래프 출력 화면으로 구성하였다. 사용자는 일 단위 또는 월 단위로 원하는 기간을 선택하여 예측 결과를 화면으로 볼 수 있다.

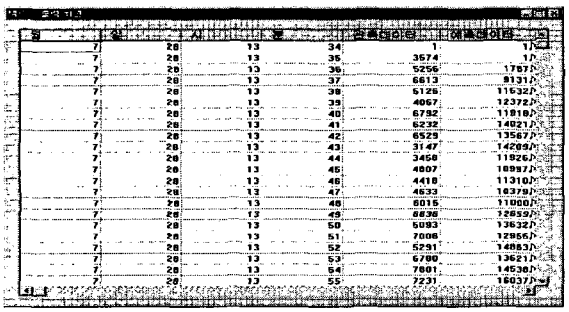

그림 7 예측 결과의 텍스트 출력 화면

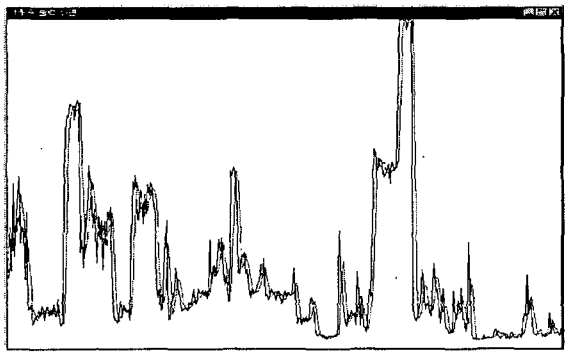

그림 8 예측 결과의 그래프 출력 화면

4.3 결과 및 고찰

실시간으로 트래픽을 측정하여 시간의 흐름에 따라 변하는 특징을 나타내는 시계열 분석을 적용하여 트래픽의 예측을 가능하게 하였다. 시계열 예측 알고리즘을 실제 시스템에 적용한 결과 예측의 오차는 다소 차이가 있었지만, 예측의 결과가 과거의 데이터의 경험성을 유지해가면서 변화한다는 것을 알 수 있었다.

5. 결론

본 논문에서 설계하고 구현한 결과는 실시간 트래픽 분석 및 예측을 목적으로 하고 있다. 본 시스템은 각 노드의 부하 여부를 감시하여, 비정상적인 트래픽의 폭주, 네트워크의 다운 등과 같은 비정상적인 작동을 발견하게되면 분석 모듈의 작동에 의해 해킹을 비롯한 네트워크 장애를 감지할 수 있으며, 예측 모듈에 의해 과거값과 현재의 측정된 데이터를 바탕으로 특정 시점 이후의 트래픽을 예측할 수 있다.

추후 연구과제로는 예측 기능을 보완하여 보다 높은 예측 과정을 수행하여 네트워크 관리를 하며, 단순히 네트워크 양을 공지하는데 그치지 않고, 실제 성능 관리를 수행할 수 있는 시스템으로 발전해 갈 것이다.

[참고문헌]

[1] William Stallings, "SNMP, SNMPv2, SNMPv3, and RMON 1 and 2", Addison Wesley, 1999.

[2] Mark A. Miller, "Managing Internetworks SNMP", M&T books, 1998.

[3] Kyung Hyu Lee, "An Agent-Manager Scheme for the Integrated Transfort Network Management", IEEE International Conference on Communications, pp.1017-1021, 1999,6.

[4] 최기헌, 이종협, "SAS/ETS를 이용한 시계열 분석과 그 응용", 자유아카데미, 1994.

[5] Yantai Shu, Zhigang Jin, Lianfang Zhang, Lei Wang, "Traffic Prediction Using FARIMA Models", IEEE International Conference on Communications, pp.891-895. 1999,6.

[6] 홍원택, 안성진, 정진욱, "시계열 분석을 이용한 SNMP MIB-II 기반의 회선 이용률 예측 기법", 한국정보처리학회 논문지 제6권 제9호,

[7] 안성진, 정진욱, "SNMP MIB-II를 이용한 인터네트 분석 파라미터계산 알고리즘에 관한 연구", 한국정보처리학회 논문지 제5권 제8호, pp.2102-2116, 1998,8.

[8] 김동수, 정태명, "실시간 네트워크 관리를 위한 SNMP 확장에 관한 연구", 한국정보처리학회 논문지 제6권 제2호, pp.449-458, 1999,2.
pp.2470-2478, 1999,9.

[9] 이강원, 김태윤, "효율적인 통신망 설계를 위한 예측 시스템 설계", 한국정보과학회 논문지, 제25권 제1호, pp.76-82, 1998,1.

[10] 정상준, 권영헌, 최혁수, 이정협, 김종근, "실시간 망 관리를 위한 패킷 분석 시스템의 설계 및 구현", 한국멀티미디어학회 춘계학술발표대회 논문집. 2001,5.