

시계열 분석을 이용한 SNMP MIB-II 기반의 회선 이용률 예측 기법

홍 원 태[†] · 안 성 진^{††} · 정 진 옥^{†††}

요 약

이 논문에서는 SNMP MIB-II를 이용하여 회선 이용률을 예측하기 위한 알고리즘을 제안하고자 한다. TCP/IP 프로토콜을 기반으로 하는 인터넷에서 SNMP의 MIB-II를 활용하여 회선의 이용률을 구하고 과거의 회선 이용률을 기반으로 미래의 회선 이용률을 예측하는 방법을 알고리즘화하여 제안한다. 이를 위해 ARIMA 방식 중 차분을 취한 형태의 MA 모델을 예측 기법의 기본 모델로 사용하였다. 예측하기 위해 필요한 전체 시스템과 회선 이용률 예측에 필요한 알고리즘을 제시하였다. 제시된 알고리즘에 대한 실험을 위해서 실존하는 라우터를 대상으로 분석 결과를 제시하고 검증하였다. 분석은 제안된 알고리즘을 기준으로 예측된 데이터를 얻고 실제 데이터와 비교해 보았다. 과거의 데이터가 평균과 분산에서 벗어나는 특이값이 적을 때 분석은 정확성을 띠었다. 회선 이용률 예측 알고리즘은 망 관리자가 미래의 전체 TCP/IP 통신 네트워크 상태를 예상하여 증설을 가능하게 함으로써 회선 용량 계획과 비용 절감측면에서 볼 때 네트워크 관리에 많은 도움을 줄 것으로 기대된다.

Forecasting Technique of Line Utilization based on SNMP MIB-II Using Time Series Analysis

Won-Taek Hong[†] · Seong-Jin Ahn^{††} · Jin-Wook Chung^{†††}

ABSTRACT

In this paper, algorithm is proposed to forecast line utilization using SNMP MIB-II. We calculate line utilization using SNMP MIB-II on TCP/IP based Internet and suggest a method for forecasting a line utilization on the basis of past line utilization. We use a MA model taking difference transform among ARIMA methods. A system for forecasting is proposed. To show availability of this algorithm, some results are shown and analyzed about routers on real environments. We get a future line utilization using this algorithm and compare it to real data. Correct results are obtained in case of being few data deviating from mean value. This algorithm for forecasting line utilization can give effect to line capacity plan for a manager by forecasting the future status of TCP/IP network. This will also help a network management of decision making of performance upgrade.

1. 서 론

인터넷이나 망이란 개념은 WWW(World Wide

Web) 서비스의 등장과 함께 전산 관련 분야에 종사하는 사람들뿐만 아니라 일반 사용자에게도 널리 퍼지게 되었다. 이와 동시에 인터넷에 접속하여 사용하는 사용자들의 수는 급격히 증가하고 있다. 이러한 환경의 변화로 인해 통신망에서의 트래픽 양이 급격히 증가하고 있다. 이는 망의 고속화에 대한 요구를 촉발시켰으

† 준 회원 : 성균관대학교 전기전자 및 컴퓨터공학부
†† 중신회원 : 성균관대학교 컴퓨터교육과 교수
††† 중신회원 : 성균관대학교 전기전자 및 컴퓨터공학부 교수
논문접수 : 1999년 1월 18일, 심사완료 : 1999년 8월 5일

면, 네트워크 관리에 대한 요구를 함께 증가시키고 있다.

일반적으로 통신 장비에 대한 관리를 수행하기 위해서 인터넷에서는 SNMP (Simple Network Management Protocol)을 사용하고 있다. SNMP는 관리 시스템과 피관리 시스템간에 관리 정보를 교환하기 위해서 사용되는 프로토콜이다. 피관리 시스템에 상주하는 에이전트(Agent)는 MIB에서 명시하는 변수들을 수집하고 이를 관리자의 요구에 따라 관리 시스템으로 보고하게 된다. 따라서, 관리 행위는 MIB의 관리 항목에 따라 의존적일 수 밖에 없다. 필요한 MIB변수들을 추출한 후 회선 이용률을 구하여 과거와 현재의 회선 상태를 파악할 수 있다. 그러나 현재 사용하는 회선에 대한 이용률을 살펴보는 것은 과거와 현재의 이용률을 파악하는데 도움이 될 수 있지만 단순히 과거의 데이터에만 의존해서는 미래의 회선 용량에 대한 개선 문제를 구체적으로 해결할 수 없다.

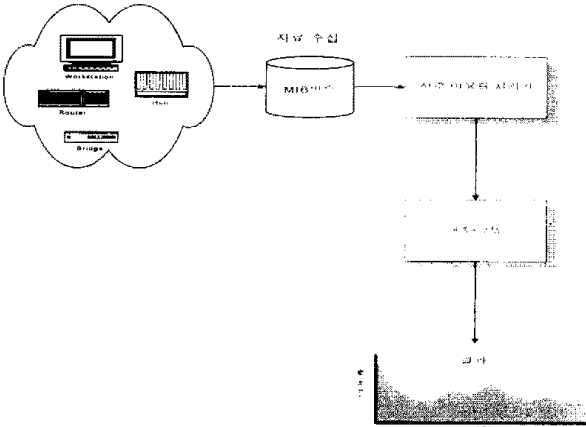
기존의 연구 동향은 관리 정보를 활용한 관리 구조나 성능 관리에 대한 연구[1], SNMP를 이용한 관리 기법 연구[2][3], MIB-II에서 제공하는 정보를 기준으로 한 성능 분석 파라미터 도출에 관한 연구[4][5] 등이 있어 왔다. 특히, [4]에서는 네트워크 관리 데이터를 수집하는 효과적인 기법을 제시하고 있고, [5]에서는 TCP/IP 네트워크 관리를 위한 분석 항목을 정의하고 있다. 또한 시계열 분석 모델을 이용하여 수집된 관리 정보의 특이값을 검출하는 기법에 관한 연구, 누적된 HTTP 패킷을 기반으로 Web 네트워크 트래픽의 양을 예측 기법에 관한 연구가 진행되었다.[6][7]

본 논문에서는 SNMP의 MIB-II를 기반으로 산출된 과거의 회선 이용률을 기본 데이터로 하여 미래의 회선 이용률을 예측하는 알고리즘을 제시하고자 한다. 실제 전화 교환기상의 트래픽 이용률을 예측하는 기법이 제시되기도 했다.[8] 성능 관리 정보를 수집하여 과거의 회선 이용률을 분석하는 연구는 이루어져왔으나, TCP/IP 프로토콜을 기반으로 하는 인터넷에서 SNMP의 MIB-II를 기반으로 시계열 분석을 이용하여 회선 이용률을 예측한다는 것은 새로운 시도이다. 이러한 시도는 SNMP, MIB-II를 기반으로 네트워크의 성능 관리를 하는 시스템에서 네트워크의 운영/관리에 실질적인 도움을 줄 것이다.

2. 회선 이용률 계산 모델과 예측

회선 이용률 예측은 (그림 1)과 같은 단계로 구성된

의 과정을 통해 MIB변수 수집, 회선 이용률 분석기, 예측 시스템 부분의 형태로 나뉘어지고 이 논문에서는 회선 이용률 분석기를 통해 나온 회선 이용률을 기본 데이터로 하여 미래의 회선 이용률을 예측하려 한다. 우선 실존하는 라우터나 워크스테이션으로부터 필요한 MIB변수들을 주기적으로 폴링하여 저장한다. 일정기간 저장된 MIB변수들은 회선 이용률 분석기를 통해 백분율 단위로 표현된다. 이 데이터를 예측 시스템에 적용하면 미래의 회선 이용률을 얻을 수 있게 된다. 본 논문에서 주로 다루려는 부분은 (그림 1)에서 예측 시스템에 해당된다.

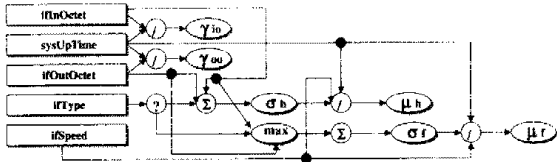


(그림 1) 회선 이용률 예측을 위한 전체 구성

(그림 2)에서는 과거의 회선 이용률을 구하는 전체적인 흐름과 필요한 MIB변수를 보여주고 있다. 회선 이용률을 구하기 위해서는 MIB-II로 부터 필요한 변수들을 추출해야 한다. 이때, 필요한 MIB-II변수는 ifInOctet, ifOutOctet, sysUpTime, ifType, ifSpeed로서 이들 데이터를 일정 간격마다 수집한다. 각 MIB-II변수의 의미는 <표 1>에서 설명되고 있다. (그림 2)에서 γ_{10} 를 구하기 위해서는 ifInOctet와 sysUpTime 관리 정보를 추출하여야 하며 γ_{100} 를 구하기 위해서는 ifOutOctet와 sysUpTime관리 정보를 구해야 한다. 최종적으로 회선 이용률은 ifType에 따라 단방향/반이중 전송 회선인 경우는 μ_h 를, 전이중 회선인 경우는 μ_f 를 구해야 한다.[5][9] 일정 기간 동안 수집된 자료를 기반으로 각 폴링에 대해 이전 값과 현재 값의 차이를 계산하여 누적한 후 분석 시점에서 평균 이용률을 계산하여 반영한다. 이러한 회선 이용률은 회선의 대역폭을 기준으로 한 송신 혹은 수신 회선의 비트의 양으로

회선의 이용 현황을 수치로 환산하여 보여준다.

마지막으로 본 논문에서 제안하는 예측 시스템은 시계열 분석 모델 ARIMA 방식 중 차분을 취한 형태의 MA 모델을 적용하였다. 이것은 MA 모델이 비교적 단기간의 시계열 변동을 잘 나타내준다는 점에 착안하였다.



(그림 2) 회선 이용률 분석 계산 모델

<표 1> 회선 이용률 도출에 필요한 MIB 변수

MIB 변수	의 미
IfInOctet	인터페이스에 수신된 옥텟의 총 수를 나타내는 카운터
IfOutOctet	인터페이스로부터 송신된 옥텟의 총 수를 나타내는 카운터
IfType	물리/링크 계층 프로토콜에 따라 구별되는 인터페이스의 유형
IfSpeed	Bps 단위로 나타나는 인터페이스의 현재 대역폭
SysUpTime	시스템의 네트워크 관리 부분이 마지막으로 재초기화된 이후의 시간을 나타내는 변수

3. MIB-II를 이용한 이용률 예측 기법

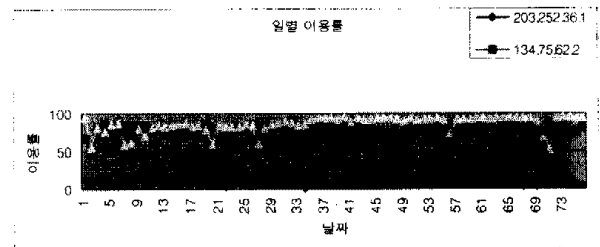
3.1 예측 모델

네트워크 회선의 이용률은 시간의 흐름에 따라 변하는 특징을 나타낸다. 이러한 형태의 자료를 시계열 자료라 하고, 이러한 시계열 자료는 시간의 흐름에 따라 일정한 패턴을 나타내는 경우가 많다. 시계열 자료의 성격을 띠는 회선의 이용률은 관찰되는 시점마다 이산적인 형태의 자료로 얻어질 수 있고, 체계적인 분석 과정을 통해 자료의 성격을 파악할 수 있다.[12] 분석의 주요한 과정은 크게 모형 수립과 예측을 들 수 있다. 모형 수립은 관찰된 시계열 자료의 여러 가지 성질을 조사하고, 특성을 체계화하여 적합한 시계열 모형을 찾는 것이고, 예측은 수립된 모형을 토대로 미래의 값을 제시하는 것이다. 이러한 분석 방법은 시계열 성격을 띠는 네트워크 회선의 이용률에도 적용될 수 있고, 본 논문에서는 이러한 절차를 밟아 미래의 네트워크 회선의 이용률을 예측해 보고자 한다.

시계열이란 시간이 경과함에 따라 관측된 자료들의

집합을 말하는 것이고, 시계열의 분석을 위해서 시계열의 각 항은 서로 비교될 수 있어야 하므로 동일한 시간 간격에 따라야 하고, 시계열을 구성하는 숫자들은 동일한 정의와 방법에 의해 조사된 것이어야 한다. 시간에 따라 변하는 회선 이용률은 이러한 시계열 성격을 만족시키는 자료라 할 수 있다. 회선 이용률의 시계열 분석 적용 여부는 "smartforecast"라는 시계열 분석 소프트웨어 프로그램을 이용하여 실험한 결과를 토대로 하여 모델을 선택하였고, 선택된 모델을 바탕으로 알고리즘을 구성하고 회선 이용률을 기본 데이터로 사용하여 분석을 수행하였다.

본 논문에서는 인터넷을 관리하는 표준 프로토콜인 SNMP를 이용하여 관리 정보를 수집하고 수집된 데이터를 바탕으로 회선의 이용률을 구하여 시계열 데이터를 만든 후 이 데이터로 시계열 모형을 정하고, 향후 회선의 이용률을 예측하려 한다. SNMP를 이용하여 트래픽이 꾸준히 발생하는 라우터를 기준으로 데이터를 수집하고, 이러한 데이터를 기반으로 이용률을 구해 보았다. 실존하는 3개 라우터의 이용률을 약 3달간 관찰한 결과는 다음의 그래프와 같은 결과를 보여준다.



(그림 3) 회선 이용률의 예

이러한 데이터를 기본 데이터로 하여 smartforecast라는 소프트웨어 프로그램을 이용하여 실험한 결과 위의 데이터는 시간이 흐름에 따라 평균과 분산이 변하는 비정상시계열 모형임을 파악할 수 있었다. 그래서 ARIMA 모델중 특정 모델로 규정지어질 수 없었고, 정상적인 시계열 모형으로 변환하기 위해 여러가지 변형을 취해 보았다. 그 중 차분을 취하는 방법으로 실제 데이터를 변화시켜 smartforecast[10]라는 소프트웨어 프로그램을 이용하여 실험한 결과 대체로 2번의 차분을 취한 변형된 형태의 MA(1)모형을 따른다는 것을 파악할 수 있었다.

MA 모형은 다음과 같은 수학적식으로 모형화될 수

있다.

$$Z_t = a_t - \theta a_{t-1}$$

t : 시점

Z_t : 시점 t에서의 이용률을 나타내는 값

θ : 모수

a_t : 백색잡음(오차 또는 잡음)

유한개의 관찰된 시계열 자료로부터 모집단인 시계열 모형에 관한 통계적 추정 및 검정을 하기 위해서는 분석을 단순화시키는 합리적인 가정이 필요하다. 시계열 분석에서 가장 중요한 가정 중의 하나는 정상성을 확보하는 것인데 실제 네트워크 상에서 얻은 회선 이용률은 평균과 분산이 변하는 비정상 시계열 모형이다. 따라서 이러한 데이터는 정상성을 확보하기 위해 실제 데이터에 변형을 취한다. 실제 데이터에 변형을 취한 결과를 수학적식으로 표현한 MA(1)모형에서 Z_t는 실제 데이터가 아닌 2회 차분을 취한 형태의 데이터가 된다. 즉, 원데이터 계열이 Y_t라면, Z_t는

$$\begin{aligned} Z_t &= \nabla^2 Y_t = \nabla Y_t - \nabla Y_{t-1} \\ &= (Y_{t+2} - Y_{t+1}) - (Y_{t+1} - Y_t) \\ &= Y_{t+2} - 2Y_{t+1} + Y_t \end{aligned}$$

이 된다. MA모형이 바르게 식별되고 예측되기 위해서는 모수(θ)가 |θ| < 1의 조건을 만족시켜야 한다는 제약 조건이 필요하다. 이러한 제약 조건은 통계학적 용어로 가역성(invertibility)이라 불린다.[11]

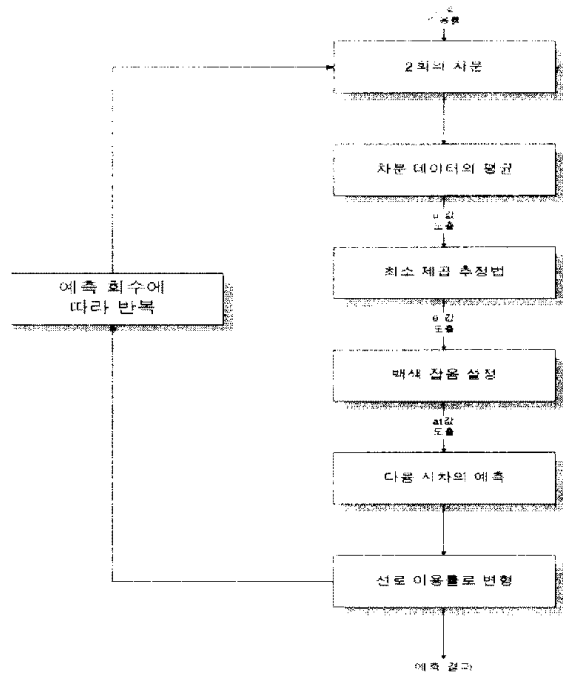
MA(1)모형은 위와 같이 정리될 수 있고, MA(1)모형을 이용하여 한 시차 이후의 예측값을 구하기 위해서는 다음과 같은 식을 이용한다.

$$Z_t(1) = \mu - \theta a_t$$

위의 식에서도 볼 수 있듯이 MA(1)을 이용하여 예측을 하기 위해서는 모수가 되는 θ를 알아내는 것과 각 시점에서의 a_t를 얻어내는 것이 중요하다. (그림 4)에서 제안한 알고리즘의 전체 흐름도를 보여주고 있다.

MA 모형을 나타내는 수식은

$$Z_t = -\theta Z_{t-1} - \theta^2 Z_{t-2} - \theta^3 Z_{t-3} - \dots + a_t$$



(그림 4) 예측 기법의 전체 알고리즘

로 표현이 가능하고, 백색 잡음인 a_t를 최소화시키는 방법으로 최소 제곱 추정법을 사용한다. 최소 제곱 추정법을 사용해서 모수(θ)를 추정하기 위해서는 다음과 같은 제곱합 함수를 최소화해야 한다.

$$S_n(\theta) = \sum_{t=1}^n a_t^2$$

단, a_t는 θ의 함수이다. 이 최소화 문제는 비선형 최적 문제이므로 제곱합 함수를 해석적으로 최소화하는 것은 불가능하다. 따라서 수치해석을 이용하여 최소값을 선택하는 방향을 따르기로 한다. 주어진 시계열 관찰값 Z₁, Z₂, ..., Z_n과 어떤 특정한 값 θ에서의 S_n(θ) 값을 계산하기 위해 MA(1) 모형을 다음과 같이 나타낸다.

$$a_t = Z_t + \theta a_{t-1}$$

여기서 초기값 a₀를 안다면 이 식을 사용해서 a₁, a₂, ..., a_n을 축차적으로 계산할 수 있다. 일반적으로 a₀값은 백색 잡음 과정의 평균인 0으로 놓는다. 따라서 특별한 값 θ에 대해 조건 a₀ = 0 하에서의 오차 제곱합 S_n(θ) = ∑_{t=1}ⁿ a_t²을 구할 수 있다. 사용하는 MA(1)모형에서는 모수가 1개이므로 구간 (-1,1)에서 격자 탐색

방법(grid search method)을 써서 오차 제곱합이 최소가 되는 값을 찾을 수 있다. 좀더 정확한 θ 을 얻기 위해서는 격자의 간격을 줄이면 된다.

다음으로 특정시점의 a_i 는 위의 식을 반복적으로 적용하여 얻을 수 있는데 a_i 식의 직관적인 의미는 예측하고자하는 시점과 가까운 시점의 데이터가 예측값에 높은 가중치를 갖고 적용된다는 것을 의미한다.

3.2 회선 이용률 예측 알고리즘

제시된 MA(1)모델을 기반으로 회선 이용률 예측 알고리즘을 작성할 수 있다. 회선 이용률 예측 알고리즘에서 중심이 되는 것은 θ 값을 알아내는 것이다. 과거의 일정 기간에 수집된 데이터를 바탕으로 회선 이용률을 구한 후 이 데이터를 기반으로 예측하고자 하는 회선의 이용률을 구해 본다. 실제 SNMP를 기반으로 회선 이용률에 적용하기 위한 데이터(MIB 변수)를 폴링해 보면 네트워크 사정에 따라 실제 데이터가 제대로 도달하지 않는 경우를 볼 수 있다. 이러한 경우 데이터를 기반으로 이용률을 구하면 실제 회선의 이용률과는 다른 값을 나타내게 되는데 이러한 수치는 해당 회선의 특성을 제대로 반영한다고 할 수 없다. 따라서 이러한 경우의 회선 이용률은 배제되었다고 가정하였다. 이런 점들을 가정하여 미래의 회선 이용률을 예측하는 알고리즘은 다음과 같다.

[제안한 알고리즘]

단계 1) 초기 MIB변수 폴링

관련 MIB 변수 집합 $V\{ifInOctet, ifOutOctet, sysUpTime, ifSpeed, ifType\}$ 을 폴링한다.

단계 2) 기준 값과 회선 형태 설정

μ_{max} 에 ifSpeed를 설정하고 L_{type} 에 ifType값을 설정한다.

단계 3) 비교 값 설정

B_{io} 에 ifInOctet, B_{oo} 에 ifOutOctet, 그리고 B_t 에 sysUpTime을 설정한다.

단계 4) 다음 MIB변수 폴링

다음 폴링 변수 집합 $V'\{ifInOctet, ifOutOctet, sysUpTime\}$ 을 t_p 시간 간격으로 폴링한다.

단계 5) 회선 형태 비교

L_{type} 의 값이 단 방향 통신 회선에 속하거나 반이중 통신 회선에 속하는 경우는 (단계 6)로 가서 σ_h 를 계산하고, 전이중 통신 회선에 속하는 경우는 (단계 7)로 가서 σ_f 를 계산한다.

단계 6) 단 방향/반이중 통신 회선에 대한 바이트 양 누적

인터페이스에 유출입되는 바이트 양을 누적한다.

$$\begin{aligned} \delta_{hi}(i) &= \text{ifInOctet}(i) B_{io} \\ \delta_{ho}(i) &= \text{ifOutOctet}(i) B_{oo} \\ \sigma_h &= \sum_{i=1}^{n_h} (\delta_{hi}(i) + \delta_{ho}(i)) \end{aligned}$$

단계 7) 전이중 통신 회선에 대한 바이트 양 누적
각 i번째 폴링에 대해 인터페이스에 유출입되는 바이트 양 중에서 최대치를 누적한다.

$$\begin{aligned} \delta_{fi}(i) &= \text{ifInOctet}(i) B_{io} \\ \delta_{fo}(i) &= \text{ifOutOctet}(i) B_{oo} \\ \sigma_f &= \sum_{i=1}^{n_f} \max(\delta_{fi}(i), \delta_{fo}(i)) \end{aligned}$$

단계 8) 시간 변화량 누적

각 i번째 폴링에 대해 지나온 시간 값을 누적하고 종료할 경우는 (단계 9)로 가고 아니면, (단계 3)로 이동한다.

$$\begin{aligned} \delta_{tp}(i) &= \text{sysUpTime } B_t \\ \sigma_t &= \sum_{i=1}^{n_t} \delta_{tp}(i) \end{aligned}$$

단계 9) 회선 이용률 계산

L_{type} 의 값이 단 방향 통신 회선에 속하거나 반이중 통신 회선에 속하는 경우는 첫번째 식으로 계산하고, 전이중 통신 회선에 속하는 경우는 두번째 식으로 계산한다.

$$\begin{aligned} \mu_{ht2} &= \frac{8\sigma_h}{\sigma_t \mu_{max}} \\ \mu_{ft2} &= \frac{8\sigma_f}{\sigma_t \mu_{max}} \end{aligned}$$

단계 10) 2차 차분 데이터

단계 1)에서 단계 9)을 반복하여 얻은 회선 이용률

(Y_t)을 바탕으로 원계열의 2차 차분 데이터(Z_t)를 얻는다.

$$\begin{aligned} Z_t &= \nabla^2 Y_t = \nabla Y_t - \nabla Y_{t-1} \\ &= (Y_{t+2} - Y_{t+1}) - (Y_{t+1} - Y_t) \\ &= Y_{t+2} - 2Y_{t+1} + Y_t \end{aligned}$$

단계 11) 차분 데이터의 평균

단계 10)을 통하여 얻은 차분 데이터(Z_t)을 바탕으로 일정 기간 동안의 차분 데이터의 평균(\bar{Z})을 구한다.

$$\mu = \bar{Z} = \sum_{t=1}^n \frac{Z_t}{n}$$

단계 12) 모수 θ 추정

다음과 같은 제곱합 함수를 최소화하는 모수 θ 를 추정하기 위해서 위에서 제시된 최소 제곱 추정법을 사용한다.

$$S_n(\theta) = \sum_{t=1}^n a_t^2$$

단계 13) 차분 데이터의 예측

단계 11)에서 구한 θ 를 기준으로 MA모형을 이용하여 한 시차 이후의 차분 데이터를 예측한다.

$$Z_t(1) = \mu - \theta a_t$$

여기서 a_t 는 다음과 같이 표현될 수 있다.

$$a_t = Z_t + \theta Z_{t-1} + \theta^2 Z_{t-2} + \theta^3 Z_{t-3} + \dots$$

단계 14) 원 데이터로 변형

단계 12)에서 얻어진 데이터를 바탕으로 단계 10)의 역과정을 통해 차분하기 이전의 데이터를 얻는다.

단계 15) 다음 시차의 예측

단계 11)에서 단계 14)의 과정을 반복함으로써 다음 시차의 데이터를 예측할 수 있다.

4. 실험 및 결과

본 논문에서 제시한 회선 이용률 예측 알고리즘의 적용성을 실험하기 위하여 실제 존재하는 라우터를 기준으로 필요한 MIB변수를 수집하고 이용률을 구현후

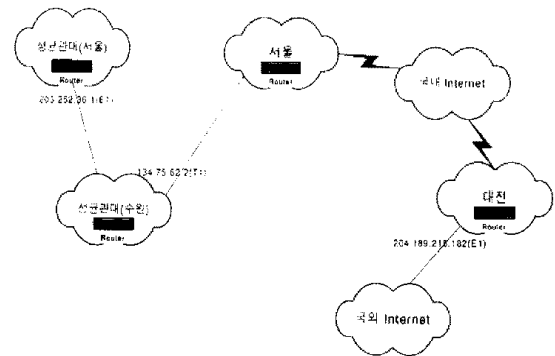
예측해 보았다. 분석 결과 제안한 알고리즘을 바탕으로 예측된 데이터와 실 데이터와는 상이한 점도 있지만 미래의 경향을 유사하게 반영하는 데이터를 산출해 내고 있다. 제안한 알고리즘을 사용하여 과거 회선 이용률을 기반으로 미래 회선 이용률의 증감 추세를 알아낼 수 있다. (그림 5)에서 제시된 라우터들을 주기적으로 모니터링하여 관리 정보를 수집하고 이를 제시한 알고리즘에 따라 회선 이용률을 예측해 보았다. 이 때, 설정된 실험 환경은 다음과 같다.

4.1 실험 환경

(1) 성균관 대학교 자연 과학 캠퍼스와 홍릉 사이의 회선은 134.75.62.2이고, 성균관 대학교 자연 과학 캠퍼스와 성균관 대학교 인문사회 과학 캠퍼스 사이의 회선은 203.252.36.1이고, 대전과 해외망사이의 회선은 204.189.216.182이다.

(2) 각각의 회선에서 Unix의 crontab을 이용하여 회선 대역폭을 고려한 주기적인 폴링을 시행한다. 기본적인 폴링 주기는 30분 또는 1시간으로 할 수 있다.

(3) 폴링 누적 데이터를 수집하여 회선 이용률을 추정한다.



(그림 5) 실험 대상 네트워크 구성도

4.2 결과 분석

실제로 제시된 알고리즘을 바탕으로 성균관 대학교 자연과학 캠퍼스와 홍릉 연구전산망 사이의 회선(134.75.62.2), 성균관 대학교 자연과학 캠퍼스와 성균관 대학교 인문사회과학 캠퍼스 사이의 회선(203.252.36.1), 대전의 연구전산망과 해외망사이의 회선(204.189.216.182)에 대한 회선 이용률의 예측 결과를 (그림 6), (그림 7), (그림 8)에서 차례로 표현하였다.

(그림 6)에서는 134.75.62.2의 예측 결과를 보여주고

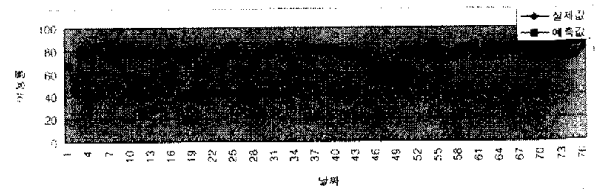
있다. 실제 데이터와 예측 데이터를 비교해 보았을 때, 유사한 경향을 나타낸다. 예측된 데이터와 실제 회선 이용률 데이터를 비교해 보면 이용률의 차이를 보이기 는 하지만 실제 데이터가 증가 추세를 보이는 것처럼 예측 데이터는 증가 추세를 보인다는 것을 관찰할 수 있다. 이 예측된 데이터의 의미는 앞으로 회선 이용률이 증가추세를 보인다는 것을 의미한다.

(그림 7)에서는 203.252.36.1의 예측 결과를 보여주고 있다. 이 회선의 특징은 비교적 이용률이 낮고, 이용률이 평균보다 큰 폭으로 변화하는 경우가 발생한다는 것이다. 실제 데이터와 예측 데이터를 비교해 보았을 때, 이용률 상의 큰 차이는 없지만 예측한 기간의 실제 데이터가 비교적 변화가 심한 편인데 반해 예측된 데이터는 일정한 이용률로 수렴해 가고 있다. 실제 데이터의 변화가 비교적 심한 환경에서 예측된 데이터의 의미는 등락폭을 감안할 때 일정한 이용률로 수렴해 가고 있다는 것을 보여주고 있다.

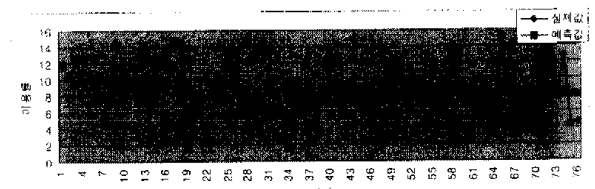
(그림 8)에서는 204.189.216.182의 예측 결과를 보여 주고 있다. 실제 데이터와 예측 데이터를 비교해 보았을 때, 유사한 경향을 나타내지만 이용률의 차이를 보인다. 예측된 데이터가 100%를 넘긴 경우가 발생하는 데 이러한 현상을 실제 회선 이용률 데이터에 적용해 해석하면 회선 증설을 의미하는 대목이다. 이 예측된 데이터는 앞으로 회선 이용률이 꾸준한 증가 추세를 보일 것이라는 것을 의미한다. 대전 지역의 연구 전산망 가입자는 204.189.216.182 라우터를 통해서 해외 Internet에 접속하게 된다. 따라서, 대전 지역 이남의 가입자는 이 라우터가 해외로 접속하는 인터페이스가 된다. 회선 이용률의 예측 결과 앞으로의 회선 이용률이 상당할 것으로 보아 증속의 필요성이 요구되고 있다.

(그림 9)에서는 204.189.216.182의 시간대별 데이터에 대한 예측 결과를 보여주고 있다. 실제 데이터와 예측 데이터를 비교해 보았을 때, 초기에는 증가추세를 반영하지만, 실제 데이터의 급격한 감소 추세를 반영하지 못하는 구간이 발생하는 것을 볼 수 있다. 이것은 시간대별 데이터의 경우 하루의 일정한 시간대에 증가하거나 감소하는 경향이 있는데 본 알고리즘은 날짜에 따른 회선 이용률을 기준으로 예측 모델을 선택하였기 때문에 발생하는 현상으로 볼 수 있다. 즉, 기본 단위를 하루로 예측한 (그림 8)과 시간 단위로 예측한 (그림 9)는 기본적으로 과거 데이터의 특성이 다르다. 이러한 시간대별 자료를 이용한 예측 기법은 주로 장기

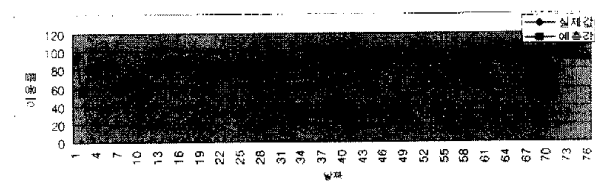
적인 예측보다는 특정 요일의 회선 이용률을 같은 간격인 시간대별로 예측하는 단기 예측에 효과적으로 쓰일 수 있다.



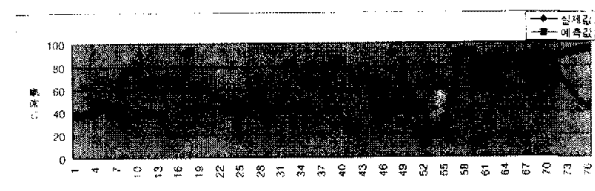
(그림 6) 134.75.62.2의 회선 이용률 예측 결과



(그림 7) 203.252.36.1의 회선 이용률 예측 결과



(그림 8) 204.189.216.182의 회선 이용률 예측 결과



(그림 9) 204.189.216.182의 시간대별 회선 이용률 예측 결과

실제 데이터와 예측 데이터간의 차이인 예측 오차(E_t)의 크기에 따라 예측의 질을 평가할 수 있는데 <표 2>에서는 위에서 예측된 데이터를 기본값으로 MAE(mean absolute error), MSE(mean squared error), MAPE(mean absolute percentage error)를 보여주고 있다. 참고로 MAE, MSE, MAPE는 다음과 같이 정의된다. 평가 척도를 사용하여 예측한 데이터의 질을 평가할 때, <표 2>에서 볼 수 있듯이 134.75.62.2와 203.252.36.1의 MAE와 MSE는 204.189.216.182의 MAE와 MSE에 비해 상

대직으로 작음값이다. 이것은 실제값과 예측값의 편차가 크지 않음을 알려준다. 반면 204.189.216.182는 실제 데이터가 크게 감소하는 부분이 발생하여 MAE와 MSE의 수치가 커지는 현상이 발생한다. 반면 203.252.36.1는 (그림 7)에서 볼 수 있듯이 실제 이용률에 대해 상대적으로 예측 오차가 큰 점이 반영되어 MAPE가 크게 나타난다.

$$MAE = \frac{1}{n} \sum_{i=1}^n |E_i|$$

$$MSE = \frac{1}{n} \sum_{i=1}^n E_i^2$$

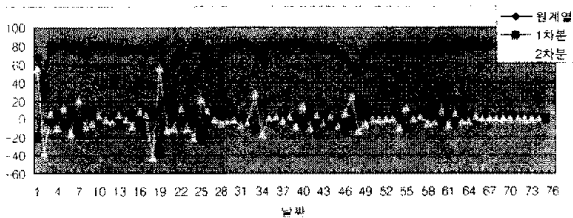
$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|E_i|}{Z_i}$$

〈표 2〉 예측 데이터의 평가

회선 IP	MAE	MSE	MAPE
134.75.62.2	6.42	57.50	0.0859
203.252.36.1	2.81	11.26	0.5072
204.189.216.182	13.08	351.41	0.1863

4.3 데이터 차분에 따른 원계열의 변형

(그림 10)는 134.75.62.2의 원 계열, 1차분, 2차분 데이터를 그래프화하여 보여주고 있다. (그림 10)에서 보듯이 1차분, 2차분을 취한 후의 데이터는 원계열에 비해 작은 대역에서 데이터의 추이가 변하고 있음을 알 수 있다. 이렇게 차분을 취함으로써 정상성을 확보하고 최소 제곱 추정법 시 계산상의 부하를 줄일 수 있는 장점이 있다. 즉, 시계열 분석을 위해서는 정상성을 확보하는 것이 필요한데 실제 네트워크 상에서 얻은 회선 이용률은 비정상적인 시계열 모형이므로 이러한 데이터는 2회 차분을 취한 형태의 데이터를 얻어 적용함으로써 정상성을 확보할 수 있었다.



(그림 10) 원 계열 데이터와 차분 계열 데이터의 비교

5. 결 론

본 논문에서는 인터넷 상에서 효율적인 네트워크 관

리를 위해 회선 이용률 예측 알고리즘을 설계했다. 회선상에서 발생하는 트래픽은 정기간의 자료를 기반으로 향후 예측이 가능하다. 예측을 하기 위해 회귀 분석과 시계열 분석을 적용할 수 있지만, 본 논문에서는 시간의 흐름에 따라 변하는 특징을 나타내는 회선 이용률을 기본 시계열 자료로 하여 회선 이용률을 예측하였다. 제시된 알고리즘은 기본 자료가 되는 각각의 회선 이용률이 평균 회선 이용률에서 크게 벗어나지 않는다는 가정이 필요하다. 이러한 환경을 만족시킬 때 제시된 알고리즘은 미래 회선 이용률의 예측을 가능하게 해준다. 실존하는 라우터를 대상으로한 실험을 통해 회선 이용률의 예측 결과를 확인해 보았다. 만약 실제 회선상의 응답이 늦어 MIB데이터가 전송되지 않는 경우나 평균 회선 이용률에서 크게 벗어나는 최대 값 데이터가 발생하는 경우에는 새로운 대응을 해야 할 것이다.

회선 이용률을 예측하기 위해서는 과거의 회선 이용률을 구하는데 필요한 MIB변수를 수집해야 하고, 누적된 데이터를 바탕으로 주기적으로 회선 이용률을 구해야 하는 사전 처리가 필요하다. 이러한 과거 회선 이용률은 ARIMA 모형 중 특정 모델로 규정될 수 없는 비정상적인 시계열 모형이므로 정상적인 시계열 모형으로 변환하기 위해 여러 가지 변형을 취해 보았다. 실험을 통해 여러 가지 변형 중 대체로 차분을 2회 실시한 변형된 형태의 MA(1)모형을 따른다는 것을 파악할 수 있었다. 이러한 검증은 smartforecast라는 소프트웨어를 사용하여 확인하였다. 변형된 형태의 MA(1)모형을 실제 과거 회선 이용률에 적용하여 예측한 결과 과거 데이터의 경향성을 유지해가면서 미래의 회선 이용률을 보여주는 것을 확인할 수 있었다.

TCP/IP 네트워크는 통신 환경의 변화와 함께 급격한 증가 추세에 있다. 이러한 흐름은 효율적인 네트워크 관리의 필요성을 대두시켰고, 네트워크 관리자는 네트워크 관리 시스템을 통해 보다 안정적인 환경을 제공하고자 한다. 본 논문에서는 ARIMA모형을 적용하여 예측함으로써 중장기 예측이 아닌 비교적 단기 예측에서 사용될 수 있는 효율적인 예측 방법론을 제시하였다. 본 논문 결과는 SNMP, MIB-II를 기반으로 네트워크의 성능 관리를 하는 시스템에서 적합한 회선 용량 계획을 통한 안정적인 네트워크 환경을 유지하는데 많은 도움을 줄 것이다. 또한 본 논문에서 제안된 알고리즘을 바탕으로 예측된 결과는 망 관리자가 회선

용량에 대한 현황을 파악하고 포화 시점을 예측함으로써 용량 계획에 대한 지표로 활용할 수 있다.

참 고 문 헌

[1] Takayuki Kushida, "The Traffic Measurement and the Empirical Studies for the Internet," GLOBECOM'98, pp.1142-1147, 1998

[2] S. Amarnath and Anurag Kumar, "A New Technique for Link Utilization Estimation in Packet Data Networks using SNMP variables," GLOBECOM'97, pp.212-216, 1997

[3] Sang Chul Shin, Seong Jin Ahn, Jin Wook Chung, "Design and Implementation of SNMP based Performance Parameter Extraction System," APNOMS, pp.137-146, 1997

[4] Sang Chul Shin, Seong Jin Ahn, Jin Wook Chung, "A New Approach to Gather Network Management Data Periodically," ITC-CSCC, pp.529-532, July 1997

[5] 안성진, "TCP/IP 망 관리를 위한 시스템 분석 파라미터 계산 알고리즘", 성균관대학교 박사학위논문, pp.54-71, 1998

[6] Cynthia S. Hood, Chuanyi Ji "Beyond Thresholds: An Alternative Method for Extracting Information from Network Measurements," GLOBECOM'97, pp.487-491, 1997

[7] J. Judge, H. W. P. Beadle, J. Chicharo, "Sampling Http Response Packets for Prediction of Web Traffic Volume Statistics," GLOBECOM'98, pp.2617-2622, 1998

[8] 이강원, 김태윤, "효율적인 통신망 관리를 위한 예측 시스템 설계", 정보과학회논문지 25권 1호, pp.76-82, 1997

[9] William Stallings, "SNMP, SNMPv2 and RMON: Practical Network Management", Addison-Wesley Publishing Company, 1996

[10] Charles N. Smart, "SmartForecasts Version 3 User's Guide," Smart Software Inc, 1992

[11] William W. S. Wei, "Time Series Analysis Univariate and Multivariate Methods", Addison-Wesley Publishing Company, 1994

[12] John Blommers, "Practical Planning for Network Growth", Prentice Hall, 1996



홍 원 택

e-mail : wthong@songgang.skku.ac.kr
 1998년 성균관대학교 정보공학과 졸업 (학사)
 1998년~현재 성균관대학교 전기 전자 및 컴퓨터공학부 대학원 석사과정

관심분야 : 네트워크 관리, 네트워크 보안



안 성 진

e-mail : sjahn@songgang.skku.ac.kr
 1988년 성균관대학교 정보공학과 졸업 (학사)
 1990년 성균관대학교 대학원 정보공학과 졸업 (석사)
 1990년~1995년 한국전자통신연구원 연구 전산망 개발실 연구원

1996년 정보통신 기술사 자격 취득
 1998년 성균관대학교 대학원 정보공학과 졸업 (박사)
 1999년~현재 성균관대학교 컴퓨터교육과 전임강사
 관심분야 : 네트워크 관리, 트래픽 분석, Unix 네트워크



정 진 욱

e-mail : jwchung@songgang.skku.ac.kr
 1974년 성균관대학교 전기공학과 학사
 1979년 성균관대학교 대학원 전자공학과 석사
 1991년 서울대학교 대학원 계산통계학과 박사

1982년~1985년 한국과학기술 연구소 실장
 1981년~1982년 Racal Milgo Co. 객원연구원
 1985년~현재 성균관대학교 전기전자 및 컴퓨터공학부 교수
 관심분야 : 컴퓨터 네트워크, 네트워크 관리, 네트워크 보안