

SARIMA 모델을 기반으로 한 선로 이용률의 동적 임계값 학습 기법

조 강 홍[†] · 안 성 진^{**} · 정 진 욱^{***}

요 약

이 논문에서는 네트워크의 QoS에 가장 큰 영향을 미치는 네트워크 선로 이용률의 과거 데이터를 기반으로 단기간 예측과 계절성(seasonality) 예측에 적합한 계절자기회귀이동평균(SARIMA : seasonal ARIMA) 모델을 적용하여 네트워크 특성을 고려한 동적인 임계값을 학습하는 알고리즘을 제시하였다. 이 기법을 통해 선로 이용률의 임계값은 네트워크 환경과 시간에 따라 동적으로 변경되며, 확률을 근거로 그 신뢰성을 제공할 수 있다. 또한, 실제 환경을 통하여 제시한 모델의 적합성 여부를 평가하였으며, 알고리즘의 성능을 실험하였다. 네트워크 관리자들은 이 알고리즘을 통하여 고정 임계값이 가지는 단점을 극복할 수 있을 것이며, 관리 행위의 효율성을 높일 수 있을 것이다.

Learning Algorithm of Dynamic Threshold in Line Utilization based SARIMA model

Kang Hong Cho[†] · Seong Jin Ahn^{**} · Jin Wook Chung^{***}

ABSTRACT

We applies a seasonal ARIMA model to the timely forecasting in a line utilization and its confidence interval on the base of the past data of the line utilization that QoS of the network is greatly influenced by. And this paper proposes the learning algorithm of dynamic threshold in line utilization using the SARIMA model. We can find the proper dynamic threshold in timely line utilization on the various network environments and provide the confidence based on probability. Also, we have evaluated the validity of the proposed model and estimated the value of a proper threshold on real network. Network manager can overcome a shortcoming of original threshold method and maximize the performance of this algorithm.

키워드 : 예측, ARIMA, SARIMA, 선로 이용률, SNMP, MIB-II

1. 서 론

인터넷과 WWW(World Wide Web)의 등장은 사용자들이 네트워크를 이용한 다양한 서비스를 삶의 일부분으로 생각할 정도로 바꾸어 놓았다. 네트워크 사용자들의 급속한 증가와 다양한 응용 서비스들이 발생시키는 트래픽은 네트워크 상의 성능 저하 발생 확률을 크게 급증시키고 있다. 결과적으로 네트워크의 Quality of Service(QoS)는 가장 중요한 문제로 대두되고 있는 상태이며 네트워크 상의 선로 이용률 측정은 네트워크 관리를 위한 가장 기본적인 QoS 파라미터가 된다. 특정 시간대의 트래픽 폭주는 네트워크 상의 응답 시간을 크게 증가시키며, 네트워크의 QoS를 떨어뜨리는 가장 중요한 요소가 된다.

이와 같은 네트워크의 QoS가 적합한 수준인가 그렇지 않은가를 결정하기 위하여 대부분의 경우, 해당 QoS에 적합한 임계값(threshold)을 두어 판단한다. 즉, 특정 QoS의 계산값이 정해진 임계값을 넘어섰거나 그 밑으로 내려갔을 경우, 서비스의 장애로 구분한다. 네트워크 상의 선로 이용률도 마찬가지로 일반적으로 80%~90%를 임계값으로 설정하여 이 이상을 넘었을 경우, 트래픽이 폭주한 상태로 판단하게 된다.

대부분의 경우, 이와 같이 고정된 임계값을 설정하여 선로를 관리하는데, 여기에는 몇 가지 문제점을 가진다. 첫째는 선로 이용률을 위한 적절한 임계값을 찾는 것이 쉽지 않다는 것이다. 임계값은 특정 네트워크 종류, 구성, 특성에 따라 서로 달라질 수 있기 때문에 고정된 임계값을 위한 기준을 찾기가 어렵다. 둘째는 한 시점에서 적절한 임계값을 찾았다 할지라도, 계속해서 동적으로 변하는 환경에서 다음 시점에 그 기준을 적용시키는 것은 적합하지 않다. 짧은 시점을 볼 때는 매 시간마다 임계값이 변화해야 하며, 길게는 네트워크 환경

[†] 준 회원 : 성균관대학교 전기전자 및 컴퓨터공학부

^{**} 종신회원 : 성균관대학교 컴퓨터교육과

^{***} 종신회원 : 성균관대학교 전기전자 및 컴퓨터공학과 교수

논문접수 : 2002년 1월 8일, 심사완료 : 2002년 8월 19일

이 변함에 따라 전체적으로 변해야 한다. 예를 들면, 일반적으로 회사의 경우 하루를 기준으로 낮 시간에는 트래픽의 양이 많지만, 밤 시간에는 트래픽의 양이 상대적으로 굉장히 적다. 이런 경우에 동일한 임계값을 적용하는 것은 아무런 의미를 가지지 못하게 된다. 셋째는 적절한 임계값을 찾았다 할지라도 고정된 임계값을 사용할 경우, 정확하게 네트워크 상태 변화를 감지하지 못하는 단점을 가진다. 예를 들면, 임계값을 넘지 않는 범위 내에서 네트워크 상태가 시간에 따라 급증하거나 급감할 경우, 네트워크 상태 변화를 전혀 감지할 수 없게 된다. 네트워크 관점에서 볼 때, 이와 같은 경우는 중요한 사건이 발생한 시점이며 이를 감지할 수 있어야 한다.

결론적으로 현재 사용하고 있는 고정 임계값은 사용하기 간단하지만 정확성과 효율성 측면에서 여러 가지 단점을 가지고 있으며, 이를 해결하기 위해서는 네트워크 환경과 시간을 고려하여 임계값을 학습하고, 이에 따라 동적으로 변경할 수 있는 메커니즘이 필요하다. 본 논문과 관련된 예측 알고리즘들을 사용하는 여러 연구들이 네트워크 관리의 다양한 분야에서 진행되어 왔다.

선로 이용률과 관련된 논문은 SNMP의 MIB을 이용하여 현재 또는 과거의 선로 이용률을 분석하고 앞으로의 선로 용량의 개선 문제를 해결하는 성능 관리 분야에서 주로 연구가 있어 왔다[1, 2]. 또한, 예측 모델을 이용한 장기간의 네트워크 트래픽 증가 추세에 대한 연구도 진행되었다[3]. 그러나, 최근에 들어서는 장애 탐지를 위한 연구들이 많이 이루어지고 있는데 선로 이용률은 다른 QoS 파라미터들에게 영향을 끼치는 중요한 요소이기 때문에 장애 관리 측면에서도 그 기능을 수행할 수 있다. 장애 관리 측면에서는 기본적으로 임계값(threshold)을 기준으로 해당 QoS가 임계값을 위반했는지를 탐지하는 분야에 대한 연구가 많이 이루어지고 있다. 그러나, 이는 결국 장애가 발생했을 경우를 탐지하는 방법으로, 네트워크 관리자가 정상적인 서비스를 제공하기 위해 대응해야 하는 시간을 제공할 수 없다. 따라서, 과거의 데이터를 기반으로 예측 기법을 사용하여 앞으로 발생할 장애를 탐지하는 것이 더 효과적이라고 할 수 있다. 이와 관련된 연구로써 [4]는 확률에 근거하여 선로 이용률이 고정 임계값을 넘어설 시점과 확률을 예측하는 기법을 제시하고 있고, [5, 6]은 MIB 데이터를 통해 베이저안 네트워크(Bayesian Network)의 알려지지 않은 장애를 탐지하는 기법을 제시하고 있고, [7, 8]은 시계열 모형을 기반으로 적용형 임계값을 구하여 장애를 탐지하는 기법을 기술하고 있다. 또한, 서비스 관리(service management)에서도 예측을 기반으로 연구가 다양하게 진행중인데, [9]는 웹 서버의 초당 HTTP 접속 횟수를 기준으로 시계열을 통해 웹 서버의 장애 발생 시점과 그 확률에 대한 접근 방법을 제시하고 있으며, [10]은 누적된 HTTP 패킷을 기반으로 웹 트래픽의 양을 예측하는 기법에 대해 기술하고 있다.

이 논문에서는 계절자기회귀이동평균(SARIMA : seasonal

ARIMA) 모형을 적용하여 네트워크 특성을 고려한 선로 이용률의 동적 임계값을 학습하는 알고리즘을 제시하였다. 이 기법을 통해 하루를 기준으로 선로 이용률의 임계값은 네트워크 환경과 시간에 따라 동적으로 변화되며, 네트워크 선로 이용률을 근거로 네트워크 상태의 미묘한 변화까지 감지할 수 있다.

2. 선로 이용률 트래픽의 ARIMA 모형

선로의 이용률은 시간의 흐름에 따라 민감하게 변하는 특성을 가진다. 즉, 하루를 기준으로 각 시간대별로, 일주일을 기준으로 각 요일별로, 한 달을 기준으로 각 일별로 선로의 이용률은 일정한 형태의 패턴을 가지게 된다. 이런 시계열 자료의 성격을 띠는 선로의 이용률은 관찰되는 시점마다 이산적인 형태의 자료로 얻어질 수 있고, 여러 가지 예측 방법을 이용하여 앞으로의 선로 이용률의 추정값을 예측할 수 있다. 이 논문에서는 다양한 예측 방법 중, 단기 예측에 적합하고 그 적합성이 입증된 Box-Jenkins의 ARIMA 모형을 이용하여 선로 이용률을 예측한다. 즉, 이전의 정상적인 트래픽 데이터 패턴을 분석하여 현재 네트워크가 어떤 선로 이용률 트래픽 특성을 가지는지 그 모형을 정하고, 이를 통해 다음 시점의 선로 이용률 임계값의 범위를 예측할 수 있다.

(그림 1)은 각 시간대별로 수집된 일반적인 네트워크 상의 선로 이용률의 예를 보여주고 있다. 1999년 3월 첫째 주와 둘째 주 동안 대전의 연구전산망과 해외 망 사이의 회선으로부터 수집된 시계열 자료이며 이는 관련 MIB 정보를 폴링하여 얻은 관리 정보를 선로의 대역폭에 대한 백분율로 나타낸 값이다. 각 데이터를 10분 간격으로 폴링하였으며 총 시계열 자료는 908개이다.

(그림 1) 선로 이용률의 예

관측된 시계열 자료를 ARIMA 모형에 적용하기 위해서는 다음의 모형의 식별(model identification), 모수 추정(estimation of parameters), 잔차 분석(diagnostic checking)의 단계를 거쳐 적합한 모형을 찾는다[11, 12].

모형의 식별은 관찰된 시계열 자료의 여러 가지 성질을 조사하여 적합한 시계열 모형을 찾는 과정이다. (그림 1)의 그래프에서 보는 바와 같이 시간대별 선로 이용률 데이터는 하루인 24시간을 주기로 반복되는 패턴, 즉 계절성을 나타내고 있으며, 시간의 흐름에 따라 평균과 분산이 변하는 비정상시계열 모형임을 파악할 수 있다. 모형을 식별하기 위해 값의 상관관계를 나타내는 ACF(Autocorrelation Function)와 PACF(Partial Autocorrelation Function)를 계산하는 것이 필요하며 (그림 2)는 그 결과 그래프이다. (그림 2)의 ACF는 지수적으로 감소하고 PACF는 lag 1 이후에 절단 형태를 띠며, 하루 주기인 24개마다 반복되는 것으로 미루어 보아 (그림 1)의 선로 이용률은 ARIMA(1, 0, 0)(1, 1, 0)₂₄ 모형을 따르는 것으로 분석할 수 있다. 이는 다양한 여러 모형을 적용시킨 후에 ARIMA(1, 0, 0)(1, 1, 0)₂₄ 모형을 가장 유사한 형태로 식별한 결과이다.

(그림 2) ACF와 PACF

AR(1)의 주기가 24인 계절성(seasonality)과 계절 차분(seasonal difference)을 가지는 모형은 다음의 식 (1)과 같다.

$$(1 - \phi B^{24})(1 - \phi B)(1 - B^{24})Z_t = a_t$$

$$Z_t = \phi Z_{t-1} + Z_{t-24} + \phi Z_{t-24} - \phi Z_{t-25} - \phi \phi Z_{t-25} - \phi Z_{t-48} + \phi \phi Z_{t-49} + a_t \quad (1)$$

여기서 a_t 백색 잡음으로서 평균 0과 분산 σ_a^2 를 가지는 독립이고 동일한 분포(iid: independent, identically distributed)의 확률 변수이다. ϕ 는 자기회기 모형(autoregressive)의 모수이고, ϕ 는 계절자기회귀 모형(seasonal autoregressive)의 모수이다.

모수 추정에는 ϕ , ϕ 각각의 값을 추정하는 단계이며, 최소제곱추정법(least squares method)을 이용해서 적합한 값을 추정할 수 있다. 위에서 나타난 시계열 자료에 대한 모수 추정값은 $\phi = 0.728$, $\phi = -0.395$ 이다.

잔차 분석은 선택한 시계열 모형이 주어진 시계열에 적합한지 그렇지 않은지를 판단하는 과정으로 a_t 가 완전한 임의성, 즉 난수성을 갖는지를 파악함으로써 적합성을 판단할 수 있다. 간단하게는 잔차의 정규성, 독립성, 등분산성 등을 통해 검토해 볼 수 있고, 좀더 정확하게는 ACF의 값이 거의 0에 근접해 있는지를 보여주는 Ljung-Box 통계와 χ^2 검정을 통해 그 모형의 적합성을 분석할 수 있다. 본 시계열 데이터에서 a_t 의 분석 결과, 자유도는 18이고 95% 유의 수준에서 χ^2 검정 통계량 값은 36.415로서, Ljung-Box 통계값이 33.156이므로 잔차 분석의 결과가 유의함을 알 수 있다. 이와 같이 선로 이용률에 대한 적합한 ARIMA 모형을 찾은 후에 이 모형을 기반으로 앞으로의 선로 이용률에 대한 예측값을 추정할 수 있게 된다.

3. 선로 이용률의 동적 임계값 학습 알고리즘

선로 이용률의 계산을 위한 접근 방법은 MIB-II에서 제공하는 정보를 기준으로 하여 성능 분석 파라미터를 도출하는 여러 가지 방법에서 연구가 진행되었다[1, 2]. MIB-II의 정보를 이용한 선로 이용률의 계산은 간단하며, 쉽게 데이터를 얻을 수 있다는 장점을 가지기 때문에 많이 사용된다. 선로 이용률을 계산하기 위해서는 선로의 타입을 통해 해당 선로의 특성을 구분한 하는 것이 필요하고 인터페이스 그룹의 ifInOctet, ifOutOctet, sysUpTime, ifSpeed 등을 사용하여 해당 선로의 입력 및 출력 이용률을 구분하여 계산할 수 있다. 먼저 입력 이용률일 경우에는 다음의 식 (2-1)과 같이 계산된다.

$$Z_n = \frac{(IfInOctet_n - IfInOctet_{n-1})}{IfSpeed \times (sysUpTime_n - sysUpTime_{n-1})} \quad (2-1)$$

출력 이용률일 경우에는 다음의 식 (2-2)와 같이 계산된다.

$$Z_n = \frac{(IfOutOctet_n - IfOutOctet_{n-1})}{IfSpeed \times (sysUpTime_n - sysUpTime_{n-1})} \quad (2-2)$$

식 (2)를 이용하여 하루를 기준으로 한 시간대별 선로 이용률을 계산할 수 있으며, n 시점까지의 과거 데이터를 이용하여 $n+1, n+2, \dots, n+l$ 시간의 선로 이용률을 예측한다. 시점이 n 이고 선시차가 l 인 시계열 Z_1, Z_2, \dots, Z_n 의 최소 평균 제곱오차 예측(minimum mean square error forecast)의 일반식은

$$\hat{Z}_n(l) = E(Z_{n+l} | Z_n, Z_{n-1}, \dots) \quad (3)$$

이다.

위의 식 (3)에 ARIMA(1,0,0) (1, 1, 0)₂₄ 모델을 적용할 경우, 각 시점 $k = 1, 2, 3, \dots, l$ 에서의 예측값은 다음과 같다.

$$\begin{aligned} \widehat{Z}_n(k) &= \phi \widehat{Z}_{n-1}(k) + \widehat{Z}_{n-24}(k) + \phi \widehat{Z}_{l-24}(k) \\ &\quad - \phi \widehat{Z}_{l-25}(k) - \phi \phi \widehat{Z}_{l-25}(k) \\ &\quad - \phi \widehat{Z}_{l-48}(k) + \phi \phi \widehat{Z}_{l-48}(k) \end{aligned} \quad (4)$$

정규 분포(normal process)에서, 예측값 $\widehat{Z}_n(k)$ 의 $(1 - \alpha)$ 100%, 즉 신뢰 구간을 계산하기 위해서는 먼저 모수 ϕ_j 값을 구하는 것이 필요하고 이를 이용하여 식 (5)에서 신뢰 구간을 계산할 수 있다.

$$\begin{aligned} \phi_j &= \phi^j, & 1 \leq j \leq 23 \\ \phi_j &= \phi \phi_{j-1} + (1 + \phi) \phi_{j-24} + \phi(1 + \phi) \phi_{j-25}, & 24 \leq j \leq 47 \\ \phi_j &= \phi \phi_{j-1} + (1 + \phi) \phi_{j-24} + \phi(1 + \phi) \phi_{j-25} \\ &\quad - \phi \phi_{j-48} + \phi \phi \phi_{j-49}, & j \geq 48 \end{aligned}$$

$$\widehat{Z}_n(k) \pm N_{\frac{\alpha}{2}} \left[1 + \sum_{j=0}^{l-1} \phi_j^2 \right]^{\frac{1}{2}} \sigma_a \quad (5)$$

여기서, $N_{\frac{\alpha}{2}}$ 는 $P(N > N_{\frac{\alpha}{2}}) = \frac{\alpha}{2}$ 인 표준정규분포값이다. 따라서, 시점 k 에서의 예측값의 범위는

$$\begin{aligned} \widehat{Z}_n(k) - N_{\frac{\alpha}{2}} \left[1 + \sum_{j=0}^{l-1} \phi_j^2 \right]^{\frac{1}{2}} \sigma_a &\leq \widehat{Z}_n(k) \\ &\leq \widehat{Z}_n(k) + N_{\frac{\alpha}{2}} \left[1 + \sum_{j=0}^{l-1} \phi_j^2 \right]^{\frac{1}{2}} \sigma_a \end{aligned}$$

이다.

결과적으로 이 신뢰 구간이 예측한 선로 이용률에 대한 허용 여부를 판단하는 상위, 하위 임계값이 될 수 있다.

$$\begin{aligned} \overline{\lambda_{high}(k)} &= \widehat{Z}_n(k) + N_{\frac{\alpha}{2}} \left[1 + \sum_{j=0}^{l-1} \phi_j^2 \right]^{\frac{1}{2}} \sigma_a \\ \overline{\lambda_{low}(k)} &= \widehat{Z}_n(k) - N_{\frac{\alpha}{2}} \left[1 + \sum_{j=0}^{l-1} \phi_j^2 \right]^{\frac{1}{2}} \sigma_a \end{aligned} \quad (6)$$

그러나, 단지 이 임계값만으로는 적절한 고정 기준치를 넘어설 때 적합하지 않을 수 있기 때문에, 현재 절대적으로 사용하는 상위 고정 임계값($\lambda_h(k)$), 하위 고정 임계값($\lambda_l(k)$)과 비교하여 그 단점을 보완하는 것이 필요하다. 따라서, 최종 임계값은

$$\begin{aligned} \text{if } \overline{\lambda_{high}(k)} > \lambda_h(k) \quad &\lambda_{high}(k) = \lambda_h(k) \\ \text{else } \overline{\lambda_{high}(k)} < \lambda_h(k) \quad &\lambda_{high}(k) = \overline{\lambda_{high}(k)} \\ \text{if } \overline{\lambda_{low}(k)} > \lambda_l(k) \quad &\lambda_{low}(k) = \lambda_l(k) \\ \text{else } < \lambda_l(k) \quad &\lambda_{low}(k) = \overline{\lambda_{low}(k)} \end{aligned} \quad (7)$$

이다.

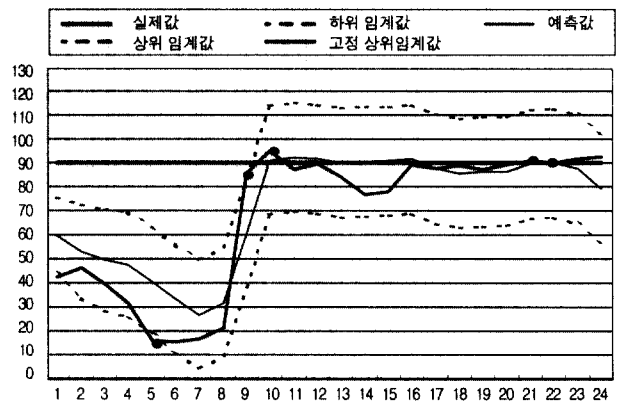
4. 실험 및 평가

위에서 제시한 알고리즘의 적용과 그 타당성을 평가하기 위하여 식 (1)에서 식별한 ARIMA(1,0,0) (1, 1, 0)₂₄ 모델을 이용하여 데이터를 수집한 마지막 날의 다음 날에 대한 선로 이용률 예측값을 계산하였다. 예측값의 개수는 하루를 기준으로 각 시간대별 24개 예측값(0시~23시)이며, 95%의 신뢰성을 가지고 예측하였다.

<표 1>은 $\widehat{Z}_n(k) \pm N_{\frac{\alpha}{2}} \left[1 + \sum_{j=0}^{l-1} \phi_j^2 \right]^{\frac{1}{2}} \sigma_a$ 즉, 예측값과 신뢰 구간으로 계산된 상위 임계값과 하위 임계값을 나타내며 (그림 3)은 실제값, 예측값 그리고 이를 기반으로 한 상위 하위 동적 임계값과 고정 임계값의 그래프를 나타낸다.

<표 1> 예측 데이터

시	하한값	$Z_n(l)$	상한값	시	상한값	$Z_n(l)$	하한값
0	43.927	59.583	75.240	12	67.206	89.963	112.719
1	33.686	53.032	72.378	13	67.705	90.462	113.220
2	28.507	49.537	70.568	14	67.923	90.682	113.440
3	25.691	47.557	69.422	15	68.728	92.487	114.246
4	18.194	40.487	62.780	16	64.736	87.495	110.254
5	10.692	33.206	55.721	17	62.808	85.567	108.326
6	4.123	26.754	49.385	18	63.463	86.222	108.981
7	9.148	31.840	54.531	19	63.696	86.455	109.214
8	37.773	60.496	82.220	20	66.653	89.412	112.171
9	68.203	90.944	113.684	21	67.113	89.872	112.632
10	69.536	92.285	115.034	22	64.635	87.394	110.153
11	68.866	91.620	114.374	23	56.506	79.265	102.025



(그림 3) 예측 그래프

먼저 (그림 3)에 나타난 예측값과 실제값을 분석해 보면 그래프에서 보는 바와 같이 실제 데이터와 예측 데이터와의 패턴은 거의 동일하며, 각 시간대별 데이터 값 자체도 큰 차이를 보이지 않는다는 것을 알 수 있다. 특히, 그래프에서 보는 바와 같이 7시에서 8시 사이에 선로 이용률이 급증하고 있는데 본 모델에서는 이 시점을 정확하게 예측하고 있는 것을 파악할 수 있다.

좀 더 정확한 예측 데이터의 평가를 위해서는 실제값(Z_{n+k})과 예측값($\hat{Z}_n(k)$) 사이의 차이인 예측 오차(e_k)의 크기로 예측의 정확성을 평가하는 여러 가지 방법이 존재한다. 그 중에서도 MAE(Mean Absolute Error), MSE(Mean Squared error), MAPE(Mean Absolute Percentage Error) 등이 주로 사용되며 각각의 계산식은 다음과 같다.

$$\begin{aligned} MAE &= \frac{1}{l} \sum_{k=0}^l |Z_{n+k} - \hat{Z}_n(k)|, \\ MSE &= \frac{1}{l} \sum_{k=0}^l |(Z_{n+k} - \hat{Z}_n(k))|^2, \\ MAPE &= \frac{1}{l} \sum_{k=0}^l \left| \frac{Z_{n+k} - \hat{Z}_n(k)}{Z_{n+k}} \right| \end{aligned} \quad (7)$$

<표 1>의 예측값에 대한 각각의 예측 평가 계산값은 MAE = 8.279, MSE = 109.26, MAPE = 23.308이다. 이 값은 다른 모형들을 적용시켰을 때보다 낮은 수치를 나타내는 값이기 때문에 본 실측 데이터에 대한 ARIMA(1, 0, 0) (1, 1, 0)₂₄ 모형의 적합성 여부를 평가할 수 있다.

(그림 3)에서 점으로 표시된 부분이 선로 이용률이 임계값을 넘어선 시간이다. 먼저 5시의 실측 데이터의 경우는 동적 하위 임계값 이하로 값이 측정된 경우이다. 이와 같은 경우는 실제로 사용자가 적어 트래픽이 낮을 수도 있지만 기존에 정상일 경우의 데이터를 기반으로 파악된 트래픽 패턴을 벗어난 경우이기 때문에 실제로 특정 네트워크 구간이 다운되거나 문제가 발생하여 트래픽의 패턴이 변화된 경우일 확률이 높다. 따라서, 이와 같이 트래픽의 패턴이 낮게 변화되는 때는 하위 고정 임계값의 경우 아무 근거없이 적절하게 임계값을 적용하기 힘들기 때문에 찾아내기 어렵다. 9시의 경우는 실측 데이터의 값이 기존의 증가 패턴을 훨씬 뛰어넘는 경우로 갑작스럽게 트래픽의 양이 증가된 경우이다. 실제로 이 경우는 ISP 또는 관리자의 입장에서 중요한 시점이며 반드시 발견해야 하는 임계값 위반의 경우이다. 그러나, 상위 고정 임계값을 적용할 경우에는 임계값을 넘지 않는 범위에서 증가했기 때문에 발견할 수 없지만, 동적 임계값의 경우에는 과거의 데이터를 근거로 그 패턴을 파악했기 때문에 그 시점을 파악할 수 있다. 그리고, 그 외의 10시, 21시 그리고 22시의 경우에는 상위 고정 임계값을 넘은 경우이다.

본 실험을 통해 본 논문에서 제시하는 알고리즘의 예측의 정확성을 검증하였고 동적 임계값을 적용하였을 경우, 기존의 고정 임계값이 찾아내지 못하는 트래픽의 중요 임계 시점을 추가적으로 발견해 낼 수 있었다.

5. 결 론

네트워크의 사용자가 급증하고 네트워크의 규모가 복잡해

짐에 따라 네트워크의 트래픽은 크게 증가하고 있으며 그 형태도 다양해지고 있다. 네트워크 사용자들의 급속한 증가와 다양한 응용 서비스들이 발생시키는 트래픽은 네트워크 상의 성능 저하 발생 확률을 크게 급증시키고 있다. 결과적으로 네트워크의 Quality of Service(QoS)는 가장 중요한 문제로 대두되고 있는 상태이며 네트워크 상의 트래픽의 양을 나타내는 선로 이용률 측정은 네트워크 관리를 위한 가장 기본적인 QoS 파라미터이다.

현재 이런 선로 이용률이 포화 상태를 넘어섰는지를 파악하는 근거는 고정 임계값을 사용하여 그 이상인지를 판단하는 단순한 방법을 사용하고 있다. 그러나, 이와 같은 고정 임계값은 네트워크 또는 그 환경에 따라 다르게 적용되어야 하기 때문에 적절한 임계값을 찾기 쉽지 않다. 또한, 적절한 임계값을 찾더라도 민감하게 변화되는 네트워크 트래픽의 갑작스런 증가와 감소 등이 발생하는 중요 시점을 찾아내는 것은 쉽지 않다. 이는 관리자의 측면에서는 중요한 부분이며 반드시 파악되어야 한다. 결론적으로 현재 사용하고 있는 고정 임계값은 사용하기 간단하지만 정확성과 효율성 측면에서 여러 가지 단점을 가지고 있으며, 이를 해결하기 위해서는 네트워크 환경과 시간을 고려하여 임계값을 학습하고, 이에 따라 동적으로 변경할 수 있는 메커니즘이 필요하다.

이 논문에서는 계절자기회귀이동평균(SARIMA : seasonal ARIMA) 모형을 적용하여 네트워크 특성을 고려한 선로 이용률의 동적 임계값을 학습하는 알고리즘을 제시하였다. 또한, 실제 환경을 통하여 제시한 모델의 적합성 여부를 평가하였으며, 기존의 고정 임계값과 비교하여 동적 임계값의 성능을 실험하였다. 이 기법을 통해 하루를 기준으로 선로 이용률의 임계값은 네트워크 환경과 시간에 따라 동적으로 학습할 수 있으며, 이를 근거로 네트워크의 선로 이용률의 상태의 미묘한 변화까지 감지할 수 있을 것이다.

또한, 이 논문에서 적용한 선로 이용률 이외의 네트워크의 QoS와 관련된 여러 파라미터들에 본 기법을 적용함으로써 과거 데이터의 패턴에 근거한 동적 임계값의 적용 범위를 확장할 수 있을 것이다.

참 고 문 헌

- [1] 홍원택, 안성진, 정진욱 "시계열 분석을 이용한 SNMP MIB-II 기반의 회선 이용률 예측기법", 정보처리학회, Vol.6, 1999.
- [2] S. Amarnath, Anurag Kumar, "A New Technique for Link Utilization Estimation in Packet Data Networks using SNMP variables," GLOBECOM'97, Vol.1, pp.212-216, 1998.
- [3] Nancy K. Groschwitz, George C. Polyzos, "A Time Series Model of Long-Term NSFNET Backbone Traffic," ICC'94, pp.1400-1404, 1994.

- [4] 조강홍, 안성진, 정진욱 "ARIMA 모델을 이용한 선로 이용률의 입계값 위반 예측 기법", 한국통신학회, Vol.25, No.8A, pp. 1153-1160, 2000.
- [5] Cynthia, S. Hood, Chuanyi Ji, "Intelligent Agents for Proactive Fault Detection," IEEE Internet Computing, Vol.22, pp.65-72, 1998.
- [6] Cynthia S. Hood, Chuanyi Ji, "Proactive Network Fault Detection," INFOCOM'97, Vol.3, pp.1147-1155, 1997.
- [7] Marina Thottan, Chuanyi Ji, "Adaptive Thresholding for Proactive Network Problem Detection," System Management, pp.108-116, 1998.
- [8] Cynthia S. Hood, Chuanyi Ji, "Beyond Thresholds : An alternative Method for Extracting Information from Network Measurements," GLOBECOM'97, pp.487-491 Vol.1, 1997.
- [9] Hellerstein, J. L, Fan Zhang, Shahabuddin, "An Approach to Predictive Detection for Service Management," Integrated Network Management, pp.309-322, 1999.
- [10] J. Judge, H. W. P. Beadle, J. Chicharo, "Sampling HTTP Response Packets for Prediction for Web Traffic Volume Statistics," GLOBECOM'98, pp.2617-2622, 1998.
- [11] William W. S. Wei, "Time Series Analysis : Univariate and Multivariate Methods," Addison-Wesley Publishing Company, 1990.
- [12] C. Chatfield, "The Analysis of Time Series : An introduction," Chapman & Hall, 1989.
- [13] Allan Leinwand, "Accomplishing Performance Management with SNMP," INET'93, pp.CEA-1-CEA-5, 1993.
- [14] 안성진, 정진욱, "SNMP MIB-II를 이용한 인터넷 분석 파라미터 계산 알고리즘에 관한 연구", 정보처리논문지, 제5권 제8호, pp.2102-2116, 1998.

조 강 홍

e-mail : khcho@songgang.skku.ac.kr

1997년 성균관대학교 정보공학과 학사

1999년 성균관대학교 전기전자 및 컴퓨터 공학부 석사

1999년 성균관대학교 전기전자 및 컴퓨터 공학부 박사과정

관심분야 : 네트워크 관리, 트래픽 분석

안 성 진

e-mail : sjahn@comedu.skku.ac.kr

1988년 성균관대학교 정보공학과(학사)

1990년 성균관대학교 대학원 정보공학과 (석사)

1990년~1995년 한국전자통신연구원 연구 전산망 개발실 연구원

1996년 정보통신 기술사 자격 취득

1998년 성균관대학교 대학원 정보공학과(박사)

1999년~현재 성균관대학교 컴퓨터교육과 조교수

관심분야 : 네트워크 관리, 트래픽 분석, Unix 네트워킹

정 진 욱

email : 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