

ARIMA 모델을 이용한 선로 이용률의 임계값 위반 예측 기법

정희원 조강홍*, 안성진** 정진욱*

Prediction Algorithm of Threshold Violation in Line Utilization using ARIMA model

Kang Hong Cho*, Seong-Jin Ahn**, Jin Wook Chung* *Regular Members*

요약

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 앞으로의 시간대별 선로 이용률을 예측하고 그 신뢰 구간을 추정함으로써, 확률에 근거한 선로 이용률 임계값 위반 예측 기법을 제시하였다. 이 기법을 통해 선로 이용률의 임계값 위반 시점을 예측할 수 있으며, 확률에 근거한 신뢰성을 제공할 수 있다. 또한, 제시한 모델의 적합성 여부를 평가하였으며, 실험을 통하여 적절한 수준의 임계값과 임계값 탐지의 기준이 되는 탐지 확률값을 추론함으로써 본 알고리즘의 성능을 최대화하였다.

ABSTRACT

This paper 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 proposes the prediction algorithm of threshold violation in line utilization using the seasonal ARIMA model. We can predict the time of threshold violation in line utilization 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 and a detection probability, it thus appears that we have maximized the performance of this algorithm.

1. 서론

인터넷과 WWW(World Wide Web)의 등장은 사용자들이 네트워크를 이용한 다양한 서비스를 삶의 일부로 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 분

제로 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.

네트워크를 우리가 살고 있는 환경에 비유해 본다면, 선로의 이용률은 자동차의 트래픽 양으로 가정할 수 있을 것이다. 자동차 운전자들이 알고자 하

* 성균관대학교 전기,전자 및 컴퓨터 공학부(khcho@songgang.skku.ac.kr),

** 성균관대학교 컴퓨터교육과 교수

논문번호 : 00040-0207, 접수일자 : 2000년 2월 7일

는 정보는 언제 어느 지점에 체증이 발생할지를 알고 싶어한다. 또 다른 예로 우리가 외출 또는 여행을 하고자할 때, 다음의 날씨를 알고 싶어한다. 따라서, 기상 예보는 중요한 역할을 하고 있으며, 기상에서 중요한 요소들로는 내일의 날씨와 그 확률(예를 들면, 비올 확률 80% 등)과 최고 또는 최저 온도(최저 -5도 ~ 최고 7도)가 될 것이다. 마찬가지로 네트워크 상의 선로 이용률은 동일한 정보를 제공하고 있으며, 네트워크 관리자 입장에서는 언제 성능 저하가 발생하는냐 하는 발생 시점과 어느 정도의 신뢰성을 가지느냐 하는 발생 확률을 알고 싶어할 것이다.

현재의 선로 이용률은 현재 또는 과거의 선로 이용률을 분석하여 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 이용률, 또한 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 연구로서 [4], [5]는 MIB 데이터를 통해 베이지안 네트워크(Bayesian Network)의 알려지지 않은 장애를 탐지하는 기법을 제시하고 있고, [6][7]는 시계열 모형을 기반으로 적응형 임계값을 구하여 장애를 탐지하는 기법을 기술하고 있다. 또한, 서비스 관리(service management)에서도 예측을 기반으로 연구가 다양하게 진행중인데, [8]은 웹 서버의 초당 HTTP 접속 횟수를 기준으로 시계열을 통해 웹 서버의 장애 발생 시점과 그 확률에 대한 접근 방법을 제시하고 있으며, [9]는 누적된 HTTP 패킷을 기반으로 웹 트래픽의 양을 예측하는 기법에 대해 기술하고 있다.

이 논문에서는 과거의 데이터를 기반으로 한 시계열 모형 중 단기간 예측과 계절성(seasonality) 예측에 적합한 seasonal ARIMA(SARIMA) 모델을

적용하여 하루 동안의 시간대별 선로 이용률을 예측하고, 이에 대한 신뢰 구간을 추정함으로써 선로 이용률 임계값에 대한 위반 여부를 탐지하는 기법을 제시한다. 결과적으로, 이 기법을 이용하여 임계값 위반(threshold violation)이 발생하는 시점 분석할 수 있으며 통계학의 정규분포를 이용하여 임계값을 위반할 확률을 계산할 수 있다.

II. 선로 이용률의 ARIMA 모델

선로의 이용률은 시간의 흐름에 따라 민감하게 변하는 특징을 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 언어될 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.

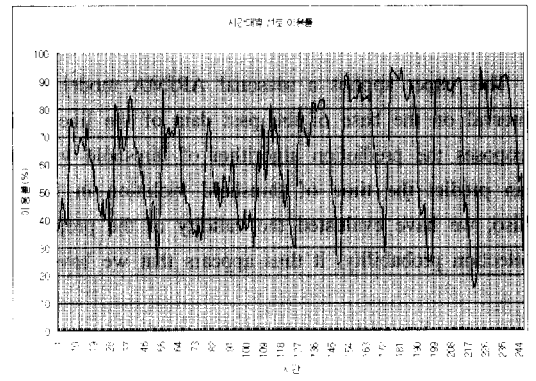


그림 1. 선로 이용률의 예

그림 1은 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.

관측된 시계열 자료를 ARIMA 모형에 적용하기

위해서는 모형의 식별(model identification), 모수 추정(estimation of parameters), 잔차 분석(diagnostic checking)의 단계를 거쳐 적합한 모형을 찾는다^{[10][11]}.

모형의 식별은 관찰된 시계열 자료의 여러 가지 성질을 조사하여 적합한 시계열 모형을 찾는 과정이다. 그림 1의 그래프에서 보는 바와 같이 시간대별 선로 이용률 데이터는 하루인 24시간을 주기로 반복되는 패턴을 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 2와 3은 그 결과 그래프이다. 그림 2의 ACF는 지속적으로 감소하고 그림 3의 PACF는 lag 1 이후에 절단 형태를 띠며, 하루 주기인 24개마다 반복되는 것으로 미루어 보아 그림 1의 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 유사형 형태로 식별한 결과이다.

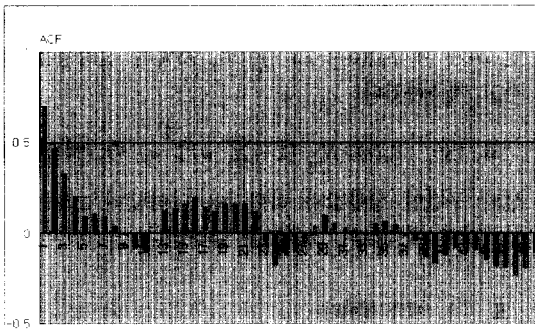


그림 2. ACF

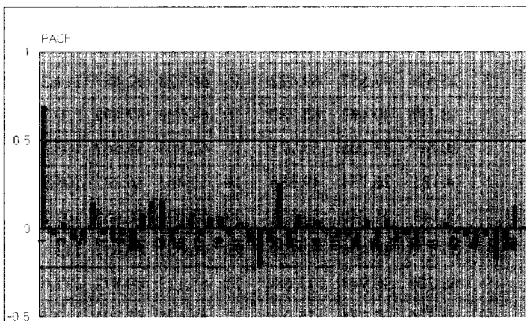


그림 3. PACF

AR(1)의 주기가 24인 계절성(seasonality)과 계절차분(seasonal difference)을 가지는 모형은 다음의 식 1과 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(autogressive)의 모수이고, ϕ 는 계절자기회기 모형(seasonal autogressive)의 모수이다.

모수 추정은 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 적합하지 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, 좀 더 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.

III. 선로 이용률의 일계값 위반 예측 알고리즘

선로 이용률의 계산을 위한 접근 방법은 MIB-II에서 제공하는 정보를 기준으로 하여 성능 분석 파라미터를 도출하는 여러 가지 방법에서 연구가 진행되었다^{[11][12]}. MIB-II의 정보를 이용한 선로 이용률의 계산은 간단하며, 쉽게 데이터를 얻을 수 있다는 장점을 가지기 때문에 많이 사용된다. 선로 이용률을 계산하기 위해서는 인터페이스 그룹의 ifInOctet, ifOutOctet, sysUpTime, ifSpeed 등이 필요하며 이

를 계산하는 공식은 다음과 같다^{[12][13]}.

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

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

$$\hat{Z}_n(k) = \phi \hat{Z}_{n-1}(k) + \hat{Z}_{n-24}(k) + \theta \hat{Z}_{t-24}(k) - \phi \hat{Z}_{t-25}(k) - \phi \theta \hat{Z}_{t-25}(k) - \theta \hat{Z}_{t-48}(k) + \phi \theta \hat{Z}_{t-49}(k) \quad (4)$$

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

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

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

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

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

이다.

예측값 $\hat{Z}_n(k)$ 과 신뢰 구간이 계산된 후에, n 시점에서 산출된 선시차 $k=1, 2, 3, \dots, l$ 의 예측값 $\hat{Z}_n(k)$ 이 그 시점의 선로 이용률 임계값인 $\lambda_n(k)$ 을 넘어설 확률을 계산할 수 있다. 이를 $\hat{P}_n(k)$ 라고 하면,

$$\begin{aligned} \hat{P}_n(k) &= P(\hat{Z}_n(k) \geq \lambda_n(k)) \\ &= P(S_n(k) \geq \frac{\lambda_n(k) - \hat{Z}_n(k)}{\sigma_n(k)}) \end{aligned} \quad (6)$$

이다. 여기서 $S_n(k)$ 은 정규분포로 표준화된 확률 변수이다.

결과적으로, 선시차 $k=1, 2, 3, \dots, l$ 의 예측값 $\hat{Z}_n(k)$ 와 확률 변수 $\hat{P}_n(k)$ 값을 통해 하루 동안의 각 시간에 따른 선로 이용률을 예측할 수 있고, 이를 이용하여 각 시간대의 선로 이용률이 임계값을 위반할 확률을 계산할 수 있다.

IV. 실험 및 평가

위에서 제시한 알고리즘의 적용과 그 타당성을 평가하기 위하여 식1에서 식별한 ARIMA(1,0,0)(1,1,0)24 모형을 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 1은 $\hat{Z}_n(k) \pm N_{\frac{\alpha}{2}} \left[1 + \sum_{j=0}^{l-1} \phi_j^2 \right]^{\frac{1}{2}} \sigma_a$ 즉, 예측값과 신뢰 구간인 상한값과 하한값을 나타내며 그림 4는 실제값과 예측값의 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

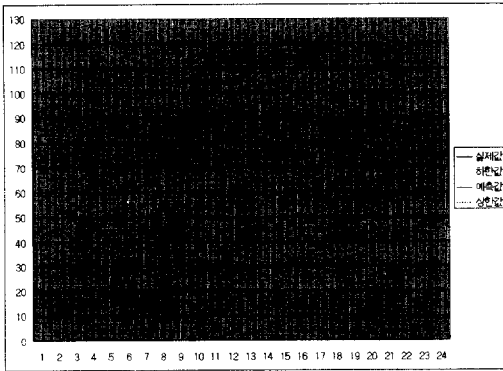


그림 4. 예측 그래프

그림 4의 예측 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. 또한, 실제 데이터 값은 신뢰 구간인 상한값과 하한값의 범위에 모두 포함되고 있다는 것을 알 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.

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

표 1의 예측값에 대한 각각의 예측 평가 계산값은 MAE=8.279, MSE=109.26, MAPE=23.308 이다. 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.

예측된 데이터를 기반으로 하여 선로 이용률 임계값 위반 예측 알고리즘을 적용하기 위해서는 적절한 선로 이용률 임계값 $\hat{\lambda}_n(k)$ 과 임계값 위반을 탐지하는 기준이 되는 탐지 확률 $P_m(k)$ 를 설정하는 것이

중요하다. 임계값을 너무 낮출 경우, 대부분의 시간대에서 임계값을 위반하기 때문에 적절한 탐지를 수행할 수 없으며, 임계값을 너무 높게 설정했을 경우에는 실제 임계값 위반이 발생했음에도 불구하고 제대로 탐지할 수 없는 경우가 발생할 수 있을 것이다. 실제로 선로 이용률 임계값은 시간에 따라 다른 값을 가질 수 있지만, 그 기준을 정하는 것이 쉽지 않기 때문에 최고치를 기준으로 한 고정된 임계값을 일반적으로 사용한다. 그림 5는 선로 이용률 임계값을 70%, 80%, 90%로 설정했을 경우 각 시간대별 임계값을 위반할 확률을 나타낸 그래프이다. 임계값과 마찬가지로 탐지 확률은 실제로 선로 이용률 상의 폭주 여부를 판별하는 기준값으로 어떤 값을 설정하느냐에 따라 그 적합성이 달라질 수 있다. 표 2는 선로 이용률 임계값 위반 예측 알고리즘을 이용하여 여러 임계값에 따라 실제로 임계값을 위반한 횟수와 탐지 확률별 예측 탐지 횟수와 예측 탐지 오류 횟수를 분석한 자료이다. 여기서 예측 탐지 오류 횟수는 임계값 위반이 발생했지만 탐지 못한 경우와 임계값을 위반하지 않았는데 임계값 위반으로 탐지한 값의 합을 의미한다. 임계값을 70%와 80%로 설정했을 경우 확률 0.5와 0.6에서 각각 16번의 실제 임계값 위반 중 15번, 13번 중 13번 모두를 예측한 것을 볼 수 있다. 그러나, 임계값이 90% 이상으로 설정했을 경우 확률 0.5에서는 실제 발생한 임계값 위반 횟수는 1번인데 반해 오류는 3번이나 발생한 것을 볼 수 있다. 이는 너무 낮은 탐지 확률로 인해 실제 임계값을 위반하지 않았는데도 불구하고 임계값이 위반되었다고 예측한 경우이다. 임계값 위반을 예측 탐지한 적중률과 탐지 오류 횟수로 미루어 볼 때 실험 데이터들이 수집된 네트워크의 경우, 임계값은 70% 또는 80%, 탐지 확률은 0.6일 경우 가장 높은 성능을 보이는 것으로 추론할 수 있다.

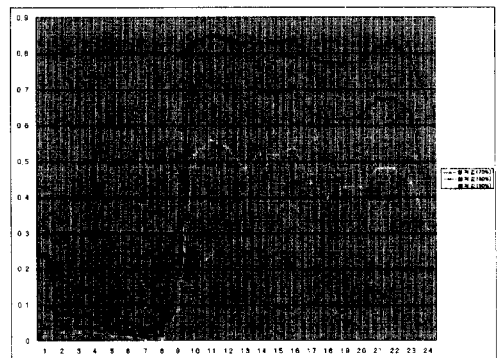


그림 5. 각 시간대별 임계값 위반 확률

표 2. 임계값 위반 탐지 분포표

확률	임계값 위반 횟수	탐지 횟수 (70%)	탐지 오류 횟수	임계값 위반 횟수	탐지 횟수 (80%)	탐지 오류 횟수	임계값 위반 횟수	탐지 횟수 (90%)	탐지 오류 횟수
0.5	16번	15번	1번	13번	13번	0번	1번	0번	3번
0.6	16번	15번	1번	13번	13번	0번	1번	0번	1번
0.7	16번	14번	2번	13번	4번	9번	1번	0번	1번
0.8	16번	9번	7번	13번	0번	13번	1번	0번	1번
0.9	16번	0번	16번	13번	0번	13번	1번	0번	1번

V. 결론

네트워크의 사용자가 급증하고 네트워크의 규모가 복잡해짐에 따라 네트워크의 장애 발생 정도는 크게 증가하고 있으며 따라서, 네트워크 관리자의 장애 대처 횟수나 범위가 늘어나고 있다. 그러나, 현재까지 대부분의 관리 행위의 형태는 결국 장애가 발생했을 경우를 임계값을 통해 탐지하는 방법으로, 네트워크 관리자가 정상적인 서비스를 제공하기 위해 미리 대응해야 하는 시간을 제공할 수 없다. 따라서, 과거의 데이터를 기반으로 예측 기법을 사용하여 앞으로 발생할 장애를 탐지하는 것이 더 효과적이라고 할 수 있다. 네트워크 관리자 입장에서는 언제 장애가 발생하는냐 하는 예측된 장애 발생 시점과 그 장애 발생 예측이 어느 정도 신뢰성을 가지느냐 하는 발생 확률을 알면 좀 더 능동적인 네트워크 관리 행위를 할 수 있을 것이다.

본 논문에서는 네트워크의 QoS에 가장 큰 영향을 미치는 네트워크 선로 이용률의 과거 데이터를 기반으로 단기간 예측과 계절성 적용이 적합한 계절자기회귀이동평균(SARIMA: seasonal ARIMA) 모형을 적용하여 앞으로의 시간대별 선로 이용률을 예측하고 그 신뢰 구간을 추정함으로써, 확률에 근거한 선로 이용률 임계값 위반 예측 기법을 제시하였다. 이 기법을 통해 선로 이용률의 임계값 위반 시점을 예측할 수 있으며, 확률에 근거한 신뢰성을 제공할 수 있었다. 또한, 제시한 알고리즘의 성능을 최대화하기 위해 적절한 임계값과 탐지 확률 추론 방법을 실험을 통하여 제시하였다. 본 실험에서는 탐지 확률이 0.6이고 임계값이 80%일 경우, 가장 좋은 성능인 100% 탐지 결과를 보였으며, 탐지 확률이 0.6이고 임계값이 70%, 80%, 90%일 경우에도 평균적으로 93.3% 탐지 결과를 나타낸 것으로 분석할 수 있다.

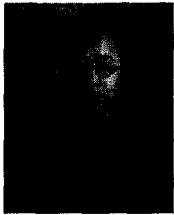
참고 문헌

- [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*, pp.212-216 vol.1, 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] Cynthia, S. Hood, Chuanyi Ji, “Intelligent Agents for Proactive Fault Detection”, *IEEE Internet Computing*, volum: 22, pp.65-72, 1998
- [5] Cynthia S. Hood, Chuanyi Ji, “Proactive Network Fault Detection”, *INFOCOM'97*, pp.1147-1155 vol.3, 1997
- [6] Marina Thottan, Chuanyi Ji, “Adaptive Thresholding for Proactive Network Problem Detection”, *System Management*, pp.108-116, 1998
- [7] Cynthia S. Hood, Chuanyi Ji, “Beyond Thresholds: An alternative Method for Extracting Information from Network Measurements”, *GLOBECOM'97*, pp.487-491 vol.1, 1997
- [8] Hellerstein, J.L, Fan Zhang, Shahabuddin, “An Approach to Predictive Detection for Service Management”, *Integrated Network Management*, pp.309-322, 1999
- [9] 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
- [10] William W. S. Wei, “Time Series Analysis: Univariate and Multivariate Methods”, *Addison-Wesley Publishing Company*, 1990
- [11] C. Chatfield, “The Analysis of Time Series: An introduction”, *Chapman&Hall*, 1989
- [12] Allan Leinwand, “Accomplishing Performance Management with SNMP”, *INET'93*, pp.CEA-1-CEA-5, 1993

[13] 안성진, 정진욱, "SNMP MIB-II를 이용한 인터넷 분석 파라미터 계산 알고리즘에 관한 연구", *정보처리학회*, 제5권 제8호, pp.2102-2116, 1998

조 강 홍(Kang Hong Cho)

정회원



1997년 2월 : 성균관대학교
정보공학과 학사
1999년 2월 : 성균관대학교
전기전자 및 컴퓨터
공학부 석사
1999년3월~현재 : 성균관대학교
전기전자 및 컴퓨터
공학부 박사과정

<주관심 분야> 네트워크 관리, 트래픽 분석

안 성 진(Seong-Jin Ahn)

정회원

1988년 2월 : 성균관대학교 정보공학과 학사
1990년 2월 : 성균관대학교 정보공학과 석사
1990년~1995년 : 한국전자통신연구원 연구전산망
개발실 연구원
1996년 : 정보통신 기술사 자격 취득
1998년 : 성균관대학교 대학원 정보공학과 박사
1999년~현재 : 성균관대학교 컴퓨터교육과 전임강사
<주관심 분야> 네트워크 관리, 트래픽분석, 유닉스
네트워킹

정 진 욱(Jin Wook Chung)

정회원



1974년 2월 : 성균관대학교
전기공학과 학사
1979년 2월 : 성균관대학교
전자공학과 석사
1991년 2월 : 서울대학교
계산통계학과 박사

1982년~1985년 : 한국과학기술 연구소 소장
1981년~1982년 : Racal Milgo Co. 객원연구원
1985년~현재 : 성균관대학교 전기전자 및 컴퓨터공
학부 교수
<주관심 분야> 컴퓨터 네트워크, 네트워크 관리, 보
안