

# 시계열 모형의 트래픽 예측에 기반한 네트워크 라우팅

(Network Routing by Traffic Prediction on Time Series Models)

정상준<sup>†</sup> 정연기<sup>\*\*</sup> 김종근<sup>\*\*\*</sup>  
 (Sangjoon Jung) (Younky Chung) (Chonggun Kim)

**요약** 트래픽의 증가는 전체 네트워크 성능에 크게 영향을 미치며 네트워크 성능의 유지 및 향상을 위해서는 트래픽 관리가 필수적이다. 본 논문에서는 네트워크 트래픽을 분석하여 시계열 모형에 의해 트래픽을 예측하고 예측된 결과가 대역폭에 비해 크다면 트래픽 폭주임을 가정하고 라우팅 경로의 비용을 증가하여 트래픽을 분산되도록 한다. 즉, 라우팅 혼잡의 발생을 예측하여 라우팅 혼잡을 미리 해소하는 방안을 제안한다. 예측 모형은 실제 네트워크 망에서 트래픽을 수집하여 모형의 확률적 오차를 최소화하는 모형을 추출한다. 확률적 오차를 최소화하는 시계열 모형을 얻기 위해서는 정상성 가정에 대한 적합성을 판단하는데, 정상성 가정은 자기상관함수와 편자기상관함수를 통해 얻을 수 있다. 실험을 통하여 추출된 예측 모형이 라우팅 경로의 비용을 조정함으로써 트래픽이 분산되도록 한다. 그 결과, 트래픽 예측 라우팅이 혼잡 발생을 미연에 방지하여 네트워크 성능을 향상시킬 수 있는 방안이라는 것을 보인다.

**키워드** : 시계열 모형, 라우팅 알고리즘, 트래픽 분석

**Abstract** An increase in traffic has a large influence on the performance of a total network. Therefore, traffic management has become an important issue of network management. In this paper, we propose a new routing algorithm that attempts to analyze network conditions using time series prediction models and to propose predictive optimal routing decisions. Traffic congestion is assumed when the predicting result is bigger than the permitted bandwidth. By collecting traffic in real network, the predictable model is obtained when it minimizes statistical errors. In order to predict network traffic based on time series models, we assume that models satisfy a stationary assumption. The stationary assumption can be evaluated by using ACF(Auto Correlation Function) and PACF(Partial Auto Correlation Function). We can obtain the result of these two functions when it satisfies the stationary assumption. We modify routing paths by predicting traffic in order to avoid traffic congestion through experiments. As a result, Predicting traffic and balancing load by modifying paths allows us to avoid path congestion and increase network performance.

**Key words** : Time series models, Routing algorithms, Traffic analyses

## 1. 서론

인터넷을 선두로 하여 전세계적으로 다양한 통신망이 구축, 운용되고 있으며 이들 통신망을 이용하여 다양한 종류의 통신이 서비스되고 있다[1-4]. 사용자 수 및 통신을 필요로 하는 시스템의 증가는 네트워크 트래픽을

증가시키고 있으며 트래픽 증가는 네트워크 성능에 많은 영향을 미치고 있다[2-3]. 네트워크의 성능을 좋게 유지하기 위해서는 망의 부하량을 감시하고 이를 제어하고 성능을 향상시키는 기능이 필요하다[3].

네트워크 트래픽 모니터링은 데이터를 수집하고 분석하는 기능과 네트워크 상에 전송되는 패킷을 검사하여 네트워크 관리에 필요한 부가적인 정보를 제공하는 기능을 가진다[4]. 전체 망에서 전송되는 트래픽을 측정하고 측정된 결과를 분석하여 부하량이 많은 곳의 트래픽을 낮은 곳으로 분산시켜 망 전체의 성능 저하를 막을 수 있다[4]. 트래픽 관리는 관리자에 의해 직접 이루어지는 수동적인 조작 방법과 망 관리 시스템에 의해 이루어지는 자동적인 방법이 있다. 수동적인 조작 방법은

† 정희원 : 경일대학교 교양학부 교수  
 sjjung@kiu.ac.kr  
 \*\* 종신회원 : 경일대학교 컴퓨터공학부 교수  
 ykchung@kiu.ac.kr  
 \*\*\* 종신회원 : 영남대학교 전자정보공학부 교수  
 (corresponding author)  
 cgkim@yu.ac.kr  
 논문접수 : 2004년 11월 15일  
 심사완료 : 2005년 4월 29일

관리자가 직접 망 정보를 인지하고 조치를 취하기 때문에 관리자의 부담이 크고 명령어 입력을 통해서만 관리가 이루어진다. 이것은 수동적인 관리 방법의 비효율적인 측면이다. 자동적인 하나의 방법은 SNMP를 이용하는 망관리 시스템을 통해 트래픽을 조절하는 것으로 트래픽 폭주가 일어나는 라우터의 Metric을 조작하여 이 라우터로 진입하는 트래픽을 제어하여 트래픽 발생을 억제한다[4].

트래픽 관리를 위해서는 트래픽의 양과 종류를 파악해야 하는데 이것은 꾸준한 측정과 분석을 통해 얻을 수 있다. 시간 단위로 얻어지는 트래픽을 수학적인 확률로 정의하여 시계열 모형으로 가공하고 이 모형을 이용하여 미래의 특정 시점의 트래픽 양을 예측할 수도 있다[1,5,6]. 트래픽 예측의 일반적인 방법은 예측 알고리즘을 적용하여 과거의 측정된 값들 사이의 관계를 구하여 미래에 발생할 값을 구하는 것이다[1-3]. 과거 자료들의 관계는 자기 유사성(Self-Similarity)과 같은 반복적인 규칙 또는 과거 값들 사이의 상관관계 등으로 정의된다[7,8]. 예측을 수행하기 위해서는 과거의 자료가 가지는 반복적인 규칙의 정상성 및 과거값들간의 상관관계가 미래에도 지속될 것이라는 가정을 바탕으로 수행되나 이 가정은 확실적인 오류를 포함할 가능성이 높다[9-11]. 이 가정이 가지는 확실적인 오류를 최소화해야만 보다 정확한 예측이 이루어질 수 있다[17].

본 논문에서는 시계열 모형을 이용하여 트래픽을 예측하고 예측되는 결과가 주어진 조건의 대역폭을 벗어나면 트래픽 폭주로 간주하고 라우팅 테이블을 변경하여 트래픽의 경로를 스위칭한다. 정확한 예측을 얻기 위해서 먼저 채택되는 시계열 모형이 정상성 가정을 만족하는지 살펴본다. 정상성 가정을 만족하지 못하면 데이터의 가공이 필요하게 되는데, 데이터 간의 확실적인 오차가 최소화되도록 데이터를 가공한다. 과거값들 간의 패턴을 분석하여 적절한 시계열 모형을 구하고 이것을 이용하여 예측값을 구한다. 트래픽 예측값을 라우팅 테이블 변경의 근거로 활용함으로써 폭주될 트래픽 경로를 스위칭하여 혼잡을 회피할 수 있다. 그러므로 혼잡에 대한 예측은 안정적인 네트워크 서비스를 제공하게 된다.

## 2. 관련 연구

### 2.1 시계열 모형

시계열 자료는 시간의 흐름에 따라 관측된 일련의 결과이다[11]. 시계열 분석은 과거 시계열 자료의 패턴이 미래에도 지속적으로 유지된다는 가정 하에서, 현재까지 수집된 자료들을 분석하여 이들 자료간의 확률 과정을 모형화하여 예측을 수행하는 것이다[9]. 즉, 시계열 예측은 관측된 과거의 자료로부터 규칙적인 패턴을 찾아 이

패턴이 미래에도 계속 될 것이라는 기대 하에 그 패턴을 모형화하여 예측하는 것이다[9-11]. 이 때 과거의 값들이 가지는 규칙이 미래에도 동일하게 유지된다는 가정을 하게 되는데 이를 정상성 가정이라 한다[9-10]. 시계열 모형을 이용한 예측은 이 가정을 만족할 경우에 한해서 정확한 예측을 수행할 수 있다. 예측을 하기 위해 적용되는 시계열 모형은 다음과 같다[9-11].

#### (1) 자기회귀모형(AR, AutoRegressive Model)

현시점  $t$ 에서의 시계열  $Z_t$ 는  $p$ 개의 과거값들의 가중합과 이들로 설명되지 않는 부분인 오차항  $a_t$ 의 선형결합으로 표현된다. 자기회귀모형은 시계열 자체에 대한 회귀 형태를 취하는 모형으로 일반  $p$ 차 AR과정을 따른  $\{Z_t\}$ 는 다음과 같이 나타낸다.

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

#### (2) 이동평균모형(MA, MovingAverage Model)

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

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

#### (3) 자기회귀이동평균 모형(ARMA)

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

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

ARMA 모형은  $p$ 차의 AR( $p$ ) 모형과  $q$ 차의 MA( $q$ ) 모형의 합으로 나타나있기 때문에 두 모형의 성질을 공유한다.

#### (4) 자기회귀누적이동평균모형(ARIMA, AutoRegressive Integrated Moving Average Model)

AR, MA, ARMA는 데이터 자체의 성질을 이용하여 시계열 모형을 구하는데 반해, ARIMA 모형은 평균 또는 분산이 일정하도록 데이터를 변환하여 시계열 모형을 구한다. 일반적으로, 평균이 일정하도록 차분을 수행하고 분산이 일정하도록 로그변환을 수행한다. 시계열  $Z_t$ 의  $d$ 차 차분한 시계열  $\{W_t = (1-B)^d Z_t\}$ 이다.

AR차수가  $p$ , MA 차수가  $q$ 인 ARMA( $p, q$ ) 모형을 갖는다면 시계열  $Z_t$ 는 차수가  $(p, d, q)$ 인 자기회귀누적이동평균(ARIMA) 모형을 갖는다고 한다.

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

$$\phi(L)Z_t = (L)u_t \text{에서}$$

$$\phi(L) = 1 - \phi_1 L - \dots = \phi_p L^p$$

단,  $\theta(L) = 1 - \theta_1 L - \dots = \theta_q L^q$

즉,  $\phi(L) \nabla^d Z_t = (L) u_t$

**2.2 시계열 모형을 이용한 네트워크 관리**

네트워크 성능관리는 네트워크 자원을 효율적으로 운용하기 위해 관리자 또는 사용자에게 자원에 관한 정보를 제공한다. 이 정보를 수집하는 형태에 따라 성능 관리 방안이 여러 가지 있다. 트래픽 모니터링은 네트워크 상에서 전송되는 데이터를 수집하고 그 양을 관리자에게 제공함으로써 네트워크의 성능을 관리한다. 트래픽 관리를 위해서는 트래픽의 양과 종류를 파악해야 하는데 SNMP 기반의 성능 관리인 경우에는 트래픽의 종류를 관리자에게 제공하는 것이 일반적인 기능이다. 반면, 실시간 모니터링인 경우 트래픽의 양을 관리자에게 제공하는 경우가 많다. 실시간으로 트래픽을 감시하는 경우, 트래픽의 양을 단순히 공지하는데 그치지 않고 특정 시점에서의 트래픽 양을 제공해 준다. 실시간 모니터링 시스템에 추가적으로 트래픽 예측 기능을 제안하는 연구도 소개되고 있다[1,2,3]. 하지만, 이 연구는 인터넷 망에서가 아니라 전화망에서의 예측을 제시하였다[3]. 또한, SNMP의 MIB 정보를 분석하여 회선 이용률을 예측한 연구도 소개되었다[2]. 네트워크에서의 트래픽은 일반 전화망과는 달리 서비스의 종류가 다양할 뿐만 아니라, 전달되는 패킷의 양 또한 가늠할 수 없을 정도로 불규칙적이다. 규칙적인 패턴을 가지는 데이터인 경우는 요인 분석, 다변량 자료 분석, 회귀분석과 같은 통계적 모형을 이용하여 예측을 할 수 있으나, 불규칙적인 데이터의 경우는 예측하기 까다롭다. 불규칙성의 규칙성을 찾아내는 것이 바로 시계열 분석인 것이다.

**2.3 라우팅 알고리즘**

인터넷 사용자의 증가와 다양한 응용 서비스들의 등장으로 트래픽의 폭증은 전송 지연 및 패킷 손실 등의 망 성능 저하의 문제를 야기하고 있다. 이 문제를 해결하기 위해서는 망의 확장을 이용하여거나 트래픽 흐름을 조작하는 방법을 사용하고 있다. 손실을 최소화하는 방법으로 시간 내의 망 전체의 포괄적인 정보를 바탕으로 적절한 트래픽 분산을 수행하면 단기간에 최소의 목적을 이룰 수 있다. 이에 따라, 라우팅 경로 분산을 위해서는 현재의 라우팅 프로토콜의 문제점과 라우팅 방법을 개선할 여지가 있다.

먼저, 현재의 라우팅 알고리즘을 살펴보면 다음과 같다. 라우팅 알고리즘은 거리 벡터(Distance Vector) 알고리즘과 링크 상태(Link State) 알고리즘으로 나뉜다. 거리 벡터(Distance Vector) 알고리즘은 라우팅을 수행하기 위해 각 라우터가 이웃 라우터와 주기적으로 정보를 공유하여 라우팅 테이블을 구성하고 구성된 라우팅

테이블을 참조하여 패킷을 서비스해 준다[12]. 패킷을 전송할 때 전송 경로에 문제가 발생하면 즉시 라우팅 테이블을 변경하지 않고 관리자가 미리 설정해 둔 주기로 라우팅 테이블 전체를 전송한다. 이 결과 거리 벡터 알고리즘에서는 라우팅 정보를 공유할 때 발생하는 여러 가지 문제점을 가지고 있다[12,16]. 반면, 링크 상태(Link State) 알고리즘은 각 라우터가 전체 네트워크의 정보를 수집하여 각 라우터에서 가장 짧은 경로를 찾아 라우팅 테이블을 구성한다[13]. 각 패킷이 라우터에 도착하면 라우팅 테이블을 참조하여 비용이 최소인 경로로 전송된다. 링크 상태 알고리즘의 대표적인 라우팅 프로토콜은 OSPF(Open Shortest Path First)이다. 이 라우팅 프로토콜은 초기에 라우팅 테이블을 구성할 때 많은 라우팅 정보를 교환하여 라우팅 경로를 최적화한다. 링크의 변화가 감지되면 그 링크에 대한 정보를 이웃 라우터에게 전달하여 최적의 라우팅 테이블을 유지하여 네트워크 전체의 성능 저하를 막는다. 그 결과 수렴이 빨리 이루어져 전체 라우팅 정보를 전달할 때보다 대역폭이 많이 소모되는 문제점을 피할 수 있다. 또한, ECMP(Equal Cost MultiPath)를 지원하여 특정 목적지로 향하는 경로를 여러 개 가지고 있어 트래픽 서비스의 성능 저하를 막는다[14]. 따라서 링크 상태 알고리즘은 거리 벡터 알고리즘에 비해 라우터의 성능 저하를 막을 수 있어 많이 사용되고 있다.

이 두 가지 방법은 라우팅 구성 정보를 비용으로 이용하여 경로 결정을 수행한다. 라우팅 구성이 변경되지 않은 한 경로 변경은 이루어지지 않아 트래픽의 포화상태가 일어날 경우 병목현상이 발생하는 경우가 많다[15,16]. 실제 망에서 라우팅은 초기 네트워크의 설정 정보를 근거로 라우팅이 이루어지므로 트래픽의 혼잡으로 인한 라우팅 변경은 불가능하다. 따라서, 트래픽 관리 정보를 라우팅에 이용하는 것은 라우팅 성능을 향상하여 라우팅의 목적인 패킷 도달 시간을 줄이는데 도움을 줄 수 있다.

본 논문에서는 패킷 전송량의 예측값을 라우팅 테이블의 비용으로 결정하여 트래픽을 스위칭하도록 한다. 경로 결정 방법은 링크 상태 알고리즘인 OSPF 프로토콜과 예측 알고리즘을 결합하여 이루어진다. 초기 라우팅 테이블 구성은 OSPF 프로토콜의 구성 절차를 따르고, 갱신 과정은 예측 알고리즘에 의해 이루어진다. 트래픽의 양을 예측하여 라우팅 변경의 결정 근거로 사용함으로써 라우팅의 성능 향상과 네트워크 전체의 트래픽 관리를 할 수 있는 방안을 제안하고자 한다.

**3. 시계열 모형을 이용한 트래픽 분석**

시계열 모형을 이용하여 트래픽의 양을 예측하기 위

해서는 먼저 트래픽을 수집해야 한다. 트래픽 수집을 위해 실제 컴퓨터 망에 트래픽 수집기를 설치한다. 수집된 결과를 시계열 데이터로 분류하여 시계열 모형으로 가정하고 이 모형의 적합성을 판정한다. 수집된 데이터가 만족하는 시계열 모형을 찾으면 이 모형을 이용하여 예측을 수행한다.

**3.1 네트워크 트래픽 정보 수집 및 시계열 모형 설계**

트래픽 수집을 위해 트래픽 수집기가 설치된 컴퓨터를 네트워크에 연결한다. 트래픽 수집기는 데이터를 시간 단위로 분류하여 수집한다. 본 연구에서는 실제 네트워크의 트래픽을 수집하고 분석한다. 트래픽 관측은 2001년 7월부터 2002년 8월까지 1년 동안 이루어졌으며, 수집된 패킷의 총량은 시간대별로 분류되어 저장된다. 장기간에 걸쳐 패킷을 수집한 결과 정전이나 네트워크 장애, 바이러스가 발생한 경우에 패킷을 수집할 수 없었다. 이 경우 전체 트래픽의 정확한 모형을 얻기 위해 누락된 데이터가 발생한 기간은 제외되었다.

트래픽을 분석하기 위해 모형 식별 통계량을 이용하여 잠정적인 모형을 선택한다. 선택된 모형이 트래픽을 예측하기에 적합한지를 식별하기 위해서는 정상성 가정을 만족하는지 살펴보아야 한다. 만약, 정상성 가정을 만족하지 못하는 경우에는 선택된 모형을 예측 모형으로 사용할 수 없다. 선택된 모형의 적합성을 판정하여 모형이 부적합하다는 결과를 얻으면 다른 시계열 모형을 선택하고 정상성 가정을 검증한다. 이 단계를 정상성 가정이 만족할 때까지 계속 반복해서 수행한다. 그림 1은 시계열 모형을 구축하는 과정을 보이고 있다.

정상성 가정을 만족하는지에 대한 여부는 데이터들간의 자기상관함수와 편자기상관함수를 통해 구할 수 있다. 정확한 함수 관계를 얻기 위해 통계학에서 널리 쓰이는 SPSS 10.0 소프트웨어를 이용한다.

**3.2 정상성 가정을 만족하는 시계열 모형 구축**

시계열 모형의 적합성을 판정하기 위해 정상성 가정을 만족하는지 살펴보아야 한다. 정상성 가정은 자기상

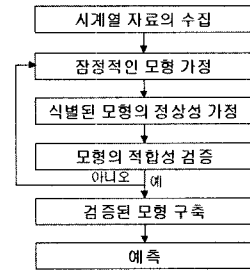


그림 1 시계열 모형의 구축

관함수와 편자기상관함수를 통해 판단할 수 있는데, 이 두 함수의 결과가 특정 시차 이내에서 주어진 신뢰구간을 만족하는 유의한 값을 가지는지에 따라 판정될 수 있다. 먼저 1년간 수집된 데이터 전체가 시계열 모형으로 적용될 수 있는가에 대한 적합성을 검증한다. 그림 2는 1년간 수집된 트래픽의 자기상관함수와 편자기상관함수의 그래프이다.

아래 그림은 자기상관함수와 편자기상관함수에 나타나는 결과가 시계열 모형을 적용하기 위해 필요한 정상성 가정을 만족하지 않는다는 것을 보이고 있다. 정상성 가정을 만족하기 위해서는 특정 시차까지 유의한 값을 나타내어야만 가능하고 이 경우에 한해서 시계열 모형을 식별할 수 있지만, 위 그래프는 특정시차가 아니라 임의적인 시차에서 유의한 값이 나타나므로 정상성 가정을 만족하지 않는다는 결과를 보여준다. 연간 수집된 트래픽 전체를 통해서도 시계열 모형을 찾을 수 없다.

위 그래프는 AR, MA, ARMA와 같은 정상 시계열 모형을 구할 수 없다는 것을 보여준다. 따라서 관측값이 정상성 가정을 만족하지 않으므로 이 관측값은 일정한 확률적 성질을 가지고 있지 않다는 것을 보여준다. 따라서 확률적인 성질 즉 평균 및 분산이 일정하도록 관측값을 변환시키는데 변환을 통해서 ARIMA와 같은 시계열 모형을 구할 수 있다. 평균이 일정하도록 차분 변환을 하고, 분산이 일정하도록 로그 변환을 수행한다. 아

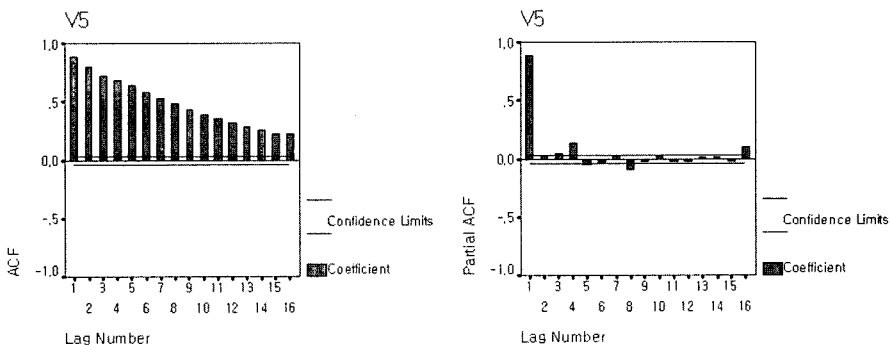


그림 2 연간 트래픽의 자기상관함수와 편자기상관함수

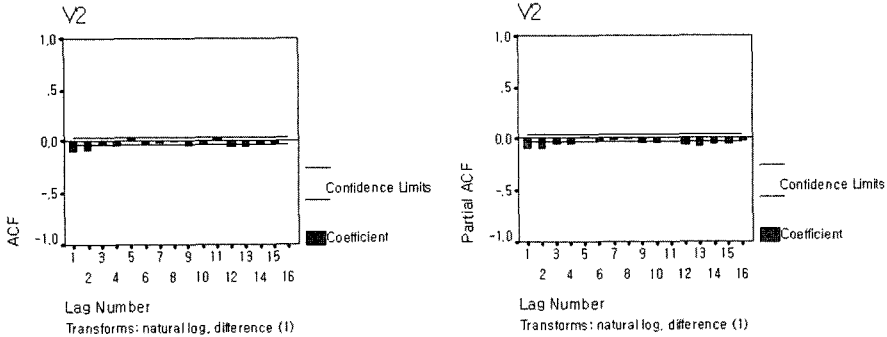


그림 3 차분, 로그변환을 수행한 후의 자기상관함수와 편자기상관함수

래 그림 3은 차분 변환과 로그 변환을 한 후 자기상관함수와 편자기상관함수를 구한 결과이다.

위의 그림에서 보는 바와 같이 차분, 로그 변환 이후의 결과도 시계열 모형을 가정할 수 있는 유의한 값을 나타내고 있지 않다. 따라서 연간 수집된 자료 전체를 대상으로 하는 시계열 모형은 구할 수 없으며, 장기간에 걸쳐 수집된 데이터를 이용하여 트래픽을 예측하는 것은 합리적이지 않다는 것을 알 수 있다.

정상성 가정을 만족하는 모형을 구하기 위해서는 적어도 중장기적인 방법으로는 어렵다는 것을 알 수 있다. 따라서 월별, 주간별, 일별로 데이터를 분류하여 분류된 데이터를 가지고 예측 모형을 식별해 보았는데, 월별 또는 주간별 데이터 또한 시계열 모형을 구할 수 없었다. 표 1은 주간별 분류에 의한 분석 결과이다.

표 1 주간별 분류에 의한 분석 결과

모형	AR	MA	ARMA	ARIMA	식별 불가능 모형
결과	6	1	1	0	14

이와 같은 결과는 시계열 모형이 적어도 중·장기적인 예측 모형으로는 적합하지 않다는 것을 보여준다. 따

라서, 시계열 모형을 이용하고자 할 경우에는 이보다는 단기적인 상황에 적합하다. 그 결과, 24시간을 주기로 하는 매일 단위의 예측에서는 정상성 가정을 만족하는 시계열 모형을 구할 수 있다. 아래 표 2는 24시간을 주기로 하여 데이터를 분류했을 때 자기상관함수와 편자기상관함수가 유의한 값을 가지는 시계열 모형의 개수를 보인다.

표 2 일별 분류에 의한 분석 결과

모형	AR	MA	ARMA	ARIMA	식별 불가능 모형
결과	105	0	17	16	12

이 표에서 150회 분석 중에 105회에 해당되는 시계열 자료가 AR 모형을 따른다는 것을 알 수 있다. 70%에 해당되는 시계열 데이터가 24시간을 주기로 가지는 AR 모형으로 식별되고 이때 차수는 1로 결정된다. 그림 4는 정상성 가정을 만족하는 AR 모형이 차수가 1일 때의 자기상관함수와 편자기상관함수 결과이다.

그림 4에서 보는바와 같이 자기상관함수는 지수적으로 감소하는 형태를 가지고 편자기상관함수는 시차 1에서만 유의한 값을 가지고 그 이후에는 절단된 형태를 가진다. 이것은 AR(1) 모형으로 식별되는 일반적인 형태이다.

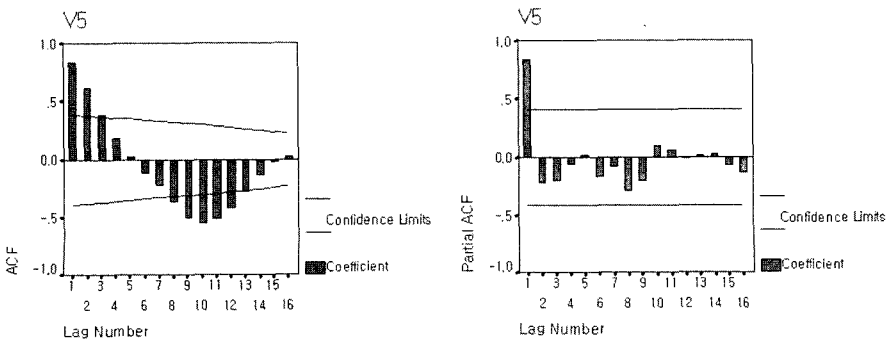


그림 4 AR(1)모형을 따르는 자기상관함수와 편자기상관함수

전체 트래픽을 분석해 본 결과 연간 총 트래픽의 변화, 월별 트래픽의 변화, 주간별 트래픽의 변화는 트래픽의 시계열 가정을 만족할 수 없었으며, 일별 분류에 의한 트래픽이 시계열 모형의 가정을 만족하였다. 따라서 정확한 예측값은 매일 24시간 단위로 시간대별로 트래픽을 분류하고 시간대별로 저장된 데이터를 이용하여 수립된 시계열 모형을 통해 유의한 값을 얻을 수 있다. 트래픽 정보의 불규칙성으로 인해 장기적인 예측을 수행할 수 없어 회귀분석, 요인분석, 다변량 자료분석과 같은 방법으로는 예측을 설명할 수 없다는 것을 알 수 있다. 따라서 시계열 모형은 단기적인 상황에 적합하다는 것을 알 수 있다.

4. 시계열 모형의 예측을 이용한 라우팅

정상성 가정을 만족하는 시계열 모형을 구하게 되면 이 모형을 이용하여 트래픽의 정확한 예측값을 구할 수 있다. 정확한 예측값은 미래의 특정 시점에서 발생할 수 있는 트래픽의 부하량을 알려주며 이것을 통해 트래픽의 혼잡을 인위적으로 피할 수 있게 된다. 따라서 예측값을 이용한 라우팅 알고리즘은 트래픽의 혼잡을 회피하여 패킷 전달시간을 최소로 가지는 경로로 패킷을 전송하여 안정적인 네트워크 서비스를 제공한다.

4.1 예측을 이용한 라우팅 수행 방법

시계열 모형을 이용하면 과거의 자료가 앞으로 일어난 트래픽 추이에 충분히 반영된 예측값을 구할 수 있다. 주기적으로 얻게 되는 예측값은 트래픽이 순간적으로 폭주하더라도 불규칙하게 증가하지는 않는다. 따라서 불필요한 라우팅 테이블의 변경을 일으키지 않아 라우팅 경로의 안정성을 제공한다. 트래픽의 증가가 일정 시간동안 지속될 경우 예측값에 영향을 주어 대역폭을 넘어서는 유의한 예측값으로 나타난다. 이 예측값이 대역폭의 한계보다 높을 때 라우팅 테이블을 조정하여 라우팅 경로를 변경시킨다.

라우팅 테이블을 구성하기 위해서는 라우팅 테이블을 작성하는 단계와 패킷 전송이 일어나는 단계로 나뉜다. 라우팅 테이블을 작성하는 단계에서 각 라우터는 링크 상태 라우팅 절차와 동일한 방법으로 전체 네트워크의 비용을 매트릭으로 구성하여 최단 경로의 라우팅을 설정한다. 모든 라우터가 라우팅 경로를 완성하면 패킷을 전송하는데 패킷이 전송되면 각 라우터는 24시간을 주기로 하는 자기회귀모형을 이용하여 예측값을 계산한다. 얻어진 예측값이 각 라우터에서 허용하는 대역폭보다 크면 라우팅 비용을 증가시켜 라우팅 테이블을 갱신한다[17]. 이 갱신 과정은 30초마다 수행된다. 예측값이 대역폭의 허용량보다 낮거나 같을 경우에는 라우터는 라

우팅 테이블을 갱신하지 않는다.

4.2 예측 알고리즘

본 논문에서 제안하는 라우팅 설정 과정에서의 예측 알고리즘은 시계열 모형의 하나인 자기회귀모형을 이용하여 예측값을 구한다. 시차를 24로 하는 자기회귀모형 [17]이 실제 트래픽 추이를 가장 잘 설명하고 있으므로 예측 알고리즘의 주기는 24로 결정된다( $t = 24$ ). 이 때 자기 회귀 모형에서의 차수( $p$ )는 1이다. 이 모형에서의 모수  $\hat{\phi}_1$ 값은 조건부 최소제곱추정법을 이용하여 추정된다. 시간대별로 저장된 패킷양이 가지는 허용 오차는 정규분포를 이용하여 구하게 되는데, 이 방법은 일반적으로 널리 사용되는 오차항 계산법이다. 현재 시점에서의 패킷 양에  $\hat{\phi}_1$ 추정값을 연산하고 오차항을 더하여 미래에 발생할 패킷의 예측값을 구한다. 이 과정은 그림 5와 같다.

<p>Step 1. 측정되는 트래픽 정보를 시간대별로 저장하여 시계열 자료로 추출</p> $X_1, X_2, \dots, X_t \quad \text{where } t = 1, \dots, n$ <p>Step 2. 시차 <math>p</math> 결정(<math>p = 1</math>, ACF와 PACF에 의해 결정됨)</p> <p>Step 3. 모수 추정 (조건부 최소제곱 추정법)</p> $\hat{\phi}_1 = \frac{\sum_{t=1}^n (X_t - \bar{X})(X_{t-1} - \bar{X})}{\sum_{t=1}^n (X_{t-1} - \bar{X})^2}$ <p>Step 4. 정규분포를 이용하여 오차항(<math>a_t</math>) 계산</p> $a_t = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{a_t^2}{2\sigma^2}}, \quad X_t \sim N(0, 1)$ <p>Step 5. 과거의 트래픽을 이용하여 트래픽의 예측값 생성</p> $X_{t+1} = \hat{\phi}_1 X_t + a_t$
---

그림 5 예측 알고리즘

4.3 예측 결과에 따른 비용 결정

링크 상태 라우팅 방법의 하나인 OSPF 프로토콜은 대역폭에 근거하여 라우팅 비용을 결정한다. 비용 결정은 (1) 식과 같이 이루어진다.

$$COST_{ospf} = 10^8 / \text{Bandwidth} \quad (1)$$

예측값을 이용하는 라우팅 비용을 결정하기 위해서는 먼저 패킷의 양이 허용할 수 있는 대역폭을 구한다. 패킷이 허용할 수 있는 링크의 대역폭 사용량은 다음과 같이 구할 수 있다.

$$\text{PredictionBandwidth}(PB) = X_{t+1} \times 8 / 1,000,000 \quad (2)$$

패킷의 예측값이 링크가 허용할 수 있는 대역폭의 90% 이상일 경우, 라우팅 경로의 비용(Costpr)은 기존의 링크가 가지는 비용의 10%가 증가되도록 한다.

if  $\text{PredictionBandwidth}(PB) > \text{Bandwidth}_{medium} \times 0.9$   
 then  $COST_{pr} = COST_{ospf} + COST_{ospf} \quad (3)$

이 과정을 단계별로 정의하면 그림 6과 같다.

4.4 예측 수행 라우팅 알고리즘

먼저 초기 라우팅 결정은 Dijkstra 알고리즘을 이용하

```

Step1. Costospf = 108 / Bandwidthmedium
Step2. Prediction Bandwidth(PB) = Yt+1 × 8 / 1,000,000
Step3. if Prediction Bandwidth(PB) > Bandwidthmedium × 0.9
      then Costpr = Costospf + Costospf × 0.1
    
```

그림 6 비용 결정 알고리즘

여 라우팅 테이블을 구성한다. 모든 노드가 라우팅 테이블을 가지게 되면, 패킷의 전송이 이루어진다. 패킷이 발생하면 매 30초마다 패킷의 양을 예측하게 되는데, 예측된 패킷의 양이 전송 경로 대역폭의 90%가 넘게 되는 경우 라우팅 비용을 10% 증가시킨다. 증가된 라우팅 비용은 모든 라우터에게 전송되며 각 라우터는 라우팅 테이블을 갱신한다. 그림 7은 예측을 통한 라우팅 알고리즘을 순서화한 것이다.

```

Step 1. Dijkstra 알고리즘을 이용한 라우팅 경로의 비용 계산
Step 2. 모든 노드가 라우팅 테이블 작성
Step 3. while (트래픽 발생 && 타이머 == 30초) {
  1. 예측 알고리즘 수행
  2. 예측 결과 생성(Xt+1)
  3. 비용 결정 알고리즘 수행
  4. Costpr을 모든 노드에게 전송
  5. if (링크 상태 변경 == true)
      break;
}
Step 4. 변경된 링크 비용 전송
    
```

그림 7 예측을 이용한 라우팅 알고리즘

### 5. 실험 및 분석

본 논문에서 제안하고 있는 라우팅 알고리즘의 성능을 확인하기 위해 NS-2 시뮬레이터를 이용한다. 시뮬레이션을 위해 각 노드를 라우터로 가정한다. 각 노드는 주어진 대역폭을 비용으로 하여 라우팅 테이블을 설정한다.

#### 5.1 실험 환경 및 평가 방법

본 논문에서 제안하는 방안을 실험하고자 그림 8과 같은 토폴로지를 사용한다.

라우팅의 성능 평가는 2X4 Grid와 3 X 6 Grid로 구

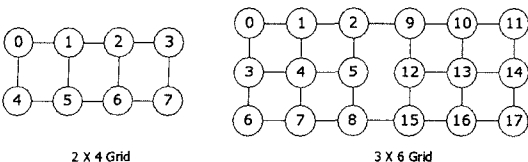


그림 8 시뮬레이션 토폴로지

분하여 평가된다. 이 때 고려해야할 시뮬레이션 환경은 표 3과 같다.

표 3 시뮬레이션 환경

	2 X 4 Grid	3 X 6 Grid
노드 수	8 개	18 개
대역폭/전달 지연 시간	1.5 Mbps 10ms	
시뮬레이션 실행 시간	400 초	
트래픽 유형	FTP, TELNET, CBR, Exponential, Pareto	
패킷 생성 최대 노드수	4 개	
링크당 최대 패킷 허용량	12,284 Bytes	

2X4 Grid에서는 좌측 4개의 노드에서 하나를 출발지 노드로 설정하고 오른쪽 노드 중의 하나를 목적지 노드로 설정한다. 이 때, 1번 노드와 5번 노드는 경유노드로 하여 라우팅 비용은 동일하도록 설정한다. 3X6 Grid도 같은 방법으로 2번 노드와 8번 노드를 경유노드로 하여 왼쪽 노드들의 집합에서 오른쪽 노드들의 집합으로 전송되도록 한다. 왼쪽 집합에서 최대 4개의 노드가 출발지 노드가 되며 오른쪽 노드 집합에서 최대 4개가 목적지 노드가 된다. 이 때 2번 노드를 경유할 때의 비용과 8번 노드를 경유할 때의 비용은 동일하다. 적용되는 라우팅 알고리즘은 링크 상태 알고리즘인 OSPF 라우팅 알고리즘(LS, Link State)과 본 논문에서 제안하는 예측을 이용한 라우팅 알고리즘(PR, Prediction Routing)이며 이 두 라우팅 알고리즘을 비교 분석하게 된다.

#### 5.2 실험 결과

라우팅 프로토콜의 성능을 평가하기 위해 다음과 같은 요소를 비교한다.

##### 5.2.1 패킷 전달시간의 비교

패킷을 전달하기 위한 라우팅의 목표는 송신지에서 출발한 패킷이 얼마나 빨리 목적지로 도착하는 것이다. 따라서 OSPF 라우팅 프로토콜과 본 논문에서 제안하는 예측을 이용한 라우팅 알고리즘의 패킷 전달시간을 비교해 본다. 그림 9는 각 토폴로지에서의 패킷 전달시간을 그래프로 보이고 있다.

그림 10에서 보는 바와 같이 라우팅의 궁극적인 목적인 패킷 전달시간이 미세하게 예측을 이용할 경우 우수한 형태로 나타나고 있다. 두 라우팅 알고리즘에서의 패킷 전달시간의 평균을 비교하면 표 4와 같다.

표 4 패킷 전달시간의 평균

평균	LS	PR
2 X 4 Grid	0.091912	0.087601
3 X 6 Grid	0.137759	0.135098

표 3에서 보는 바와 같이 제안하는 라우팅 알고리즘

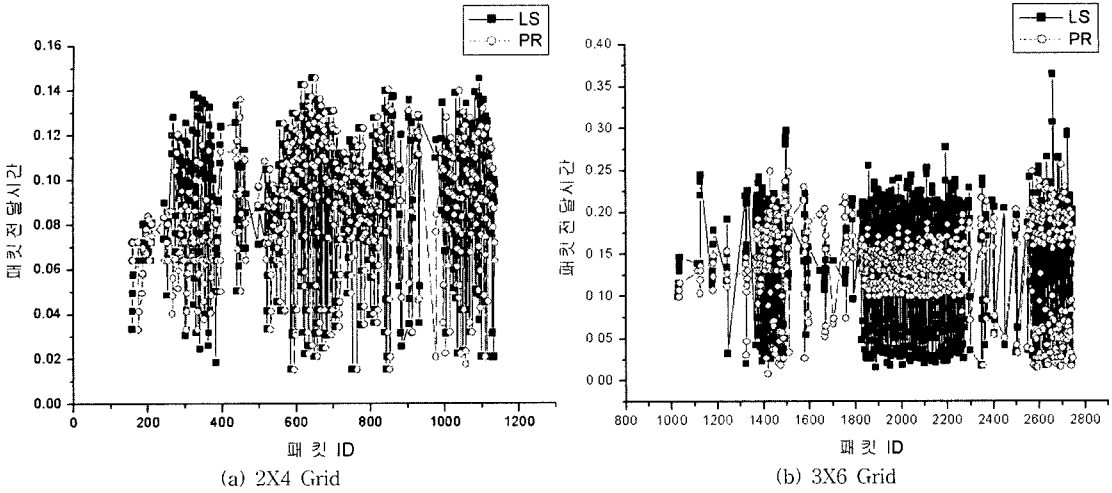


그림 9 각 토폴로지에서의 패킷 전달 시간

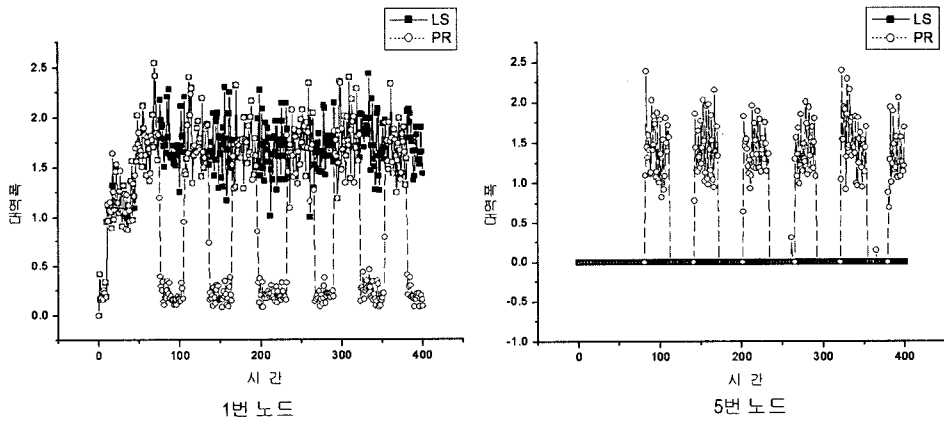


그림 10 2X4 Grid에서의 경유노드별 대역폭 추이

은 혼잡으로 인해 발생하는 패킷의 전달 지연 시간을 감소할 수 있어 패킷을 전달할 때 걸리는 시간을 줄인다. 위의 표에서 두 토폴로지에서도 제안하고 있는 예측 라우팅 알고리즘이 패킷 전달시간을 감소시킨다.

5.2.2 경유 노드의 시간대별 대역폭

2X4 Grid와 3X6 Grid 토폴로지에서도 경유하는 노드의 시간대별 대역폭을 비교해 본다. 경유노드의 대역폭은 트래픽이 어디로 전송되는지 보여준다. 먼저 2X4 Grid의 1번 노드와 5번 노드는 최종 목적지까지 도착하기 위해 거쳐야 하는 노드이다. 3X6 Grid 토폴로지에서는 2번 노드와 8번 노드를 비교한다.

같은 라우팅 비용을 가질 경우 OSPF에서는 반드시 1번 노드로만 라우팅이 이루어지지만 예측을 이용한 라우팅은 트래픽의 혼잡이 발생할 경우 5번 노드로 라우팅이 설정된다. 이 결과 트래픽의 혼잡을 회피한다.

그림 10과 11에서 보는 바와 같이 LS형 라우팅 방법은 혼잡이 발생하더라도 링크 상태의 변화가 없는 한 라우팅 경로가 변경되지 않는다. 따라서 혼잡으로 인한 라우팅 스위칭은 일어나지 않아 패킷의 전달을 지연시키게 된다. 제안하는 예측을 이용한 라우팅 알고리즘은 트래픽 전송이 이루어질 때마다 예측이 발생하며 이 결과를 30초 단위로 이웃 라우터에게 알려주어 라우팅 테이블을 변경시킨다. 트래픽의 예측값이 대역폭 보다 높을 경우 각 라우터는 라우팅 비용을 증가시켜 트래픽을 스위칭한다. 따라서 각 패킷은 혼잡을 회피하여 패킷 전달 지연 시간이 감소되어 목적지에 빨리 도달할 수 있게 된다.

5.2.3 목적지노드의 시간대별 대역폭

각 토폴로지에서도 목적지노드에서의 대역폭을 비교해 보면 제안하는 라우팅 알고리즘이 라우팅 스위칭을 가



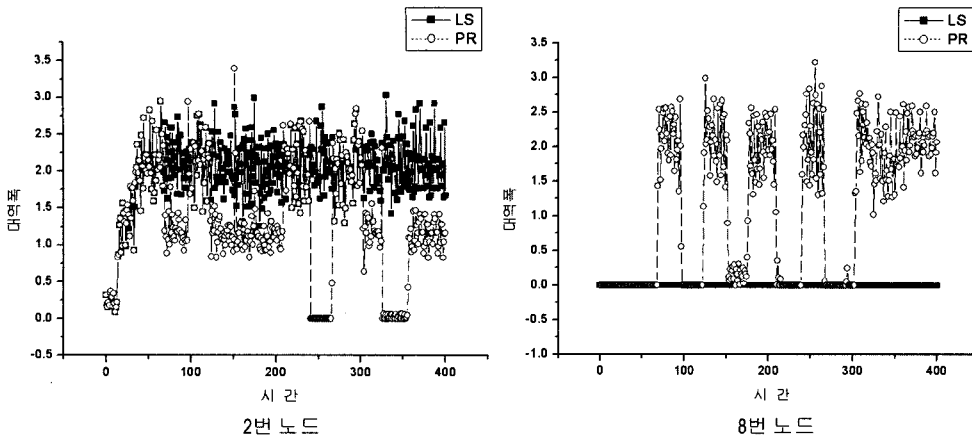


그림 11 3X6 Grid에서의 경유노드별 대역폭 추이

지더라도 일정한 대역폭을 가진다는 것을 알 수 있다. 예측을 이용한 라우팅 알고리즘이 패킷 전달 지연 시간을 줄여 링크 상태 알고리즘보다 일정한 수준의 대역폭을 가진다.

표 5 목적지노드에서의 대역폭의 표준 편차

표준 편차	LS	PR
2 X 4 Grid	0.040493	0.040264
3 X 6 Grid	0.054203	0.053989

표 5에서 보는 바와 같이 라우팅 비용의 변경으로 경유 노드가 스위칭되더라도 목적지노드는 일정한 대역폭을 가진다. 혼잡을 회피하는 예측을 이용한 알고리즘은 안정적인 대역폭을 확보하여 사용자의 중단 없는 서비스를 제공한다.

#### 5.2.4 손실되는 패킷의 양

각 토폴로지에서 라우팅을 수행한 후, 비교하는 두 알고리즘이 가지는 패킷의 손실 양을 살펴본다. 표 6은 각 토폴로지에서 두 라우팅 알고리즘에 의해 얻어진 패킷 전송 결과를 보이고 있다.

패킷의 양을 예측하여 트래픽의 혼잡을 회피함으로써 발생하는 패킷의 손실을 줄일 수 있다. 두 토폴로지에서 손실 비율을 비교해보면 링크 상태 알고리즘보다 제안하는 알고리즘의 손실 비율이 낮다는 것을 알 수 있다. 제안하는 알고리즘이 혼잡으로 인해 손실되는 패킷

표 6 손실되는 패킷 양

	2 X 4 Grid		3 X 6 Grid	
	LS	PR	LS	PR
패킷 총량	629,123	646,136	635,366	654,160
손실 패킷량	10,019	8,682	6,319	4,852
손실 비율(%)	1.593%	1.343%	0.995%	0.741%

을 줄여 링크 상태 알고리즘보다 많은 패킷들을 목적지 노드에 도착하도록 한다.

#### 5.3 결과

네트워크의 트래픽을 분석해 본 결과 시계열 모형으로 예측할 수 있다는 것을 확인하였다. 시계열 모형은 중장기적인 예측보다는 단기간의 예측에 적합하다는 것을 확인하였고, 이 사실을 이용하여 네트워크 성능에 영향을 미치는 트래픽 정보를 예측하여, 라우팅을 결정하는 근거로 사용하고자 하였다. 실험을 통해 네트워크의 예측값을 이용한 라우팅은 일정시간 동안에는 예측값에 따른 트래픽 증가를 반영하지 않는다. 즉, 일정 시간 동안에는 라우팅 경로를 안정적으로 확보할 수 있었고, 트래픽의 증가가 지속되는 상황과 같이 충분히 예측값에 영향을 주는 경우에 라우팅 테이블을 변경하였다. 시계열 모형을 이용할 경우, 바이러스나 네트워크의 침입 등과 같은 네트워크의 이상 징후가 짧은 시간동안 일어나는 경우에는 예측할 수 없겠지만, 증가 추세와 같은 경향성을 가질 경우에는 예측이 가능하다. 그 결과 목적지 노드에서는 안정적인 대역폭을 확보하게 되어 향상된 성능의 서비스를 제공받을 수 있었다.

#### 6. 결론

OSPF 라우팅 프로토콜은 ECMP를 제공하지만 라우팅 경로가 끊어지지 않는한 다른 경로로 패킷을 전송하지 않는다. 따라서 순간적으로 과도한 트래픽이 유입될 경우 라우팅 경로를 스위칭하지 않아 혼잡이 일어나게 되어 트래픽의 손실이 발생한다. 본 논문에서는 서비스되는 패킷의 양을 측정하고 특정 시점에서의 트래픽 양을 예측하여 혼잡 발생의 징후를 발견하면 라우팅 테이블을 미리 갱신하여 라우팅 경로를 변경하도록 하였다. 그 결과 패킷의 혼잡을 회피함으로써 출발지 노드에서

목적지 노드까지의 패킷 전달 시간을 줄일 수 있게 되었다. 본 연구의 예측값을 이용한 라우팅은 일정시간 동안에는 예측값에 따른 트래픽 증가를 반영하지 않는다. 즉, 일정 시간 동안에는 라우팅 경로를 안정적으로 확보할 수 있었고, 트래픽의 증가가 지속되는 상황과 같이 충분히 예측값에 영향을 주는 경우에 라우팅 테이블을 변경하였다. 그 결과 목적지 노드에서는 안정적인 대역폭을 확보하게 되어 향상된 성능의 서비스를 제공받을 수 있었다. 또한, 혼잡을 회피함으로써 손실되는 패킷의 양을 줄일 수 있었다. 따라서 시계열 모형을 이용한 예측은 실제 망에 적용될 경우 트래픽 관리를 가능하게 함으로써 서비스 향상에 많은 도움을 줄 것으로 보인다. 향후 연구로는 예측의 정확성을 높일 수 있는 방안과 네트워크의 혼잡을 예측함으로써 예견되는 상황을 복구할 수 있는 성능 관리 방안에 대해 연구하고자 한다.

### 참 고 문 헌

- [1] Yantai Shu, Zhigang Jin, Lianfang Zhang, Lei Wang, "Traffic Prediction Using FARIMA Models," IEEE International Conference on Communications, pp. 891-895. 1999, 6.
- [2] 홍원택, 안성진, 정진욱, "시계열 분석을 이용한 SNMP MIB-II 기반의 회선 이용률 예측 기법", 한국정보처리학회 논문지 제 6 권 제 9 호, 1999, 9.
- [3] 이강원, 김태윤, "효율적인 통신망 설계를 위한 예측 시스템 설계", 한국정보과학회 논문지, 제25권 제1호, pp. 76-82. 1998, 1.
- [4] 김희진, 박상근, 최덕재, "네트워크 트래픽 감시를 통한 동적인 트래픽 부하 분산에 관한 연구", 한국통신학회 춘계학술발표논문집. 2002, 5.
- [5] Ryan Kastner, Elaheh Bozorgzadeh and Majid Sarrafzadeh, "Predictable routing," Computer Aided Design, 2000. ICCAD-2000. IEEE/ACM International Conference on , 5-9 Nov. 2000.
- [6] Xun Su and Gustavo de Veciana, "Predictive routing to enhance QoS for stream-based flows sharing excess bandwidth," Computer Networks, Volume 42, Issue 1, 15 May 2003, pp. 65-80.
- [7] 김동일, 김창호, "Ethernet 트래픽의 장기간 의존성 및 Self-Similar 트래픽 소스 모델링에 관한 연구", Telecommunication Review, 제 11 권 6호, 2001, 12.
- [8] Wilinger, W., Wilson, D., Taqqu, M. "Self-similar Traffic Modeling for Highspeed Networks," ConneXions, Nov. 1994.
- [9] 이덕기, "예측방법의 이해", SPSS 아카데미, 1999.
- [10] 정동빈, 원태연, "SPSS를 활용한 시계열 자료와 단순화 분석", SPSS 아카데미, 2001.
- [11] 최기현, 이종협, "SAS/ETS를 이용한 시계열 분석과 그 응용", 자유아카데미, 1994.
- [12] F. C. M. Lau, Guihai Chen, Hao Huang and Li Xie, "A distance-vector routing protocol for networks with unidirectional links," Computer Communications,

Volume 23, Issue 4, 15 February 2000, pp. 418-424

- [13] G. Michael Schneider, Tamas Nemeth, "A simulation study of the OSPF-OMP routing algorithm," Computer Networks, Volume 39, Issue 4, 15 July 2002, pp. 457-468.
- [14] M. Littman and J. Boyan, "A distributed reinforcement learning scheme for network routing," Technical Report CMU-CS-93-165, School of Computer Science, Carnegie Mellon University, 1993.
- [15] J. A. Boyan and M. L. Littman, "Packet routing in dynamically changing network: A reinforcement learning approach," In J. Cowan, G. Tesauro, and J. Alspector, editors, Advances in Neural Information Processing Systems 6. MIT Press, Cambridge, MA, 1994.
- [16] Dong Ho Kim, Keun Ho Ryu and Young So Cho, "A new routing control technique using active temporal data management," Journal of Systems and Software, Volume 51, Issue 1, 1 April 2000, pp. 37-48.
- [17] 정상준, 김동주, 권영현, 김종근, "네트워크 트래픽 예측을 위한 시계열 모형의 적합성 검증", 한국통신학회 논문지 제 29 권 제 2B 호. 2004, 2.

#### 정 상 준

1999년 2월 영남대학교 통계학과 이학사  
2001년 2월 영남대학교 컴퓨터공학과 석사.  
2003년 2월 영남대학교 컴퓨터공학과 박사수료.  
2003년 3월~현재 경일대학교 교양학부 전임강사.  
관심분야는 네트워크 관리, ubiquitous 네트워크, 원격 교육



#### 정 연 기



1982년 2월 영남대학교 전자공학과 졸업 (공학사).  
1984년 2월 영남대학교대학원 정보통신 전공(공학석사).  
1996년 2월 영남대학교대학원 정보통신 전공(공학박사).  
1985년 3월~1990년 2월 가톨릭상지대학 전산정보처리과 조교수.  
1990년 3월~현재 경일대학교 IT대학 컴퓨터공학부 교수.  
1998년 1월~1998년 12월 호주 뉴캐슬대학교 컴퓨터공학과 방문교수.  
관심분야는 멀티미디어 통신, LAN/WAN 기술, TMN/TINA 체계의 통신망 운용관리, 차세대 인터넷

#### 김 종 근



1981년 영남대학교 전자공학과 학사  
1987년 영남대학교 전자공학과 석사  
1991년 (일본) 전기통신대학 박사.  
1997년 (미국) Virginia Tech. 연구교수  
2003년 (미국) UCSC 연구교수.  
현재 영남대학교 컴퓨터공학전공 교수.  
관심분야는 컴퓨터 네트워크, ubiquitous 네트워크, 분산처리, 인터넷 응용, 멀티미디어 기반 가상강의 시스템