2001년 한국정보처리학회 추계 학술발표논문집 제8권 제2호

연관 규칙을 이용한 네트워크 트래픽 패턴 분석

박태진, 원용관
전남대학교 컴퓨터공학과
e-mail:{aladdin, ykwon}@grace.chonnam.ac.kr

Analysis of Network Traffic Patterns using Association Rules

Tae-Jin Park, Yong-Gwan Won
Dept of Computer Engineering, Chonnam National University

요약

네트워크에 대한 활용 범위가 방대해 지면서, 신뢰성 및 효율성을 가지는 네트워크 관리가 필요하게 되었다. 특히 네트워크 관리에 데이터 마이닝을 이용해 네트워크의 운용 상태에 대한 유용한 정보를 추출하기 위한 기법들이 연구되고 있다. 본 논문에서는 네트워크의 최적화를 위한 하나의 방법으로, 특정 노드의 트래픽 집중 현상을 줄이기 위한 방법을 제안한다. 제안된 방법은 먼저 노드별 트래픽 정보를 표현하고, 수집된 정보들간의 연관성을 가지는 규칙들을 찾으며, 이들 규칙들 중 중복되거나 유용하지 않은 규칙들을 제거하고, 마지막으로 네트워크의 구성 정보를 반영하여 트래픽의 분산에 도움이 되지 않는 정보를 담고 있는 규칙들을 제거한다. 이러한 과정으로 얻어진 규칙들은 새로운 라우팅 정책에 반영하여 병목 현상을 제거하는데 효과적으로 활용할 수 있다.

1. 서론

네트워크의 규모가 커지면서 복잡성(complexity), 상호 연동성(inter-operability), 속도(velocity) 및 유연성(flexibility)에 대한 요구 등이 증가함에 따라 네트워크에 대한 효율적인 관리의 중요성과 필요성이 부각되고 있다. 이를 위해 국제 표준화 기구(ISO)에서는 네트워크 관리를 다섯 가지 기능 영역인 장애(fault) 관리, 구성(configuration) 관리, 계정(account) 관리, 성능(performance) 관리, 보안(security) 관리 등으로 구분하여 기술하고 있다[1].

네트워크 관리의 목표인 안정적이고 효율적인 네트워크를 이용한 고품질의 서비스 제공을 위해서는 관리자에게 신속하고 정확한 운용관리 정보를 제공하고 문제점 발생 시 이를 신속히 조치 할 수 있는 체계가 필요하다. 네트워크 장애 관리(network fault management)란 네트워크에서 발생하는 장애를 검사하고, 장애가 발생한 위치를 파악, 교정, 분리시키고, 서비스를 지속적으로 제공하는 것을 의미한다[2]. 일반적으로 네트워크 장애 관리는 네트워크 감시단계, 감시 자료 분석 단계, 조치 단계로 이루어진다. 네트워크 감시 단계는 네트워크의 상태에 대한 정보를 종합하여 이해 가능한 형태로 네트워크 관리자에게 제시하는 것이다. 감시 자료 분석 단계는 전 단계에서 수집된 트래픽 혹은 알람 데이터 패턴을 분석하는 단계이다. 마지막으로 조치 단계는 분석자료를 기반으로 장애를 교정하는 단계이다.[3]

본 논문에서는 트래픽 정보에 데이터 마이닝 기법을 적용하여 전송 장치에 발생되는 트래픽 패턴을 묘사하는 연관 규칙을 추출하고, 추출된 규칙들 중에서 트래픽의 집중해소를 위한 분산에 도움이 되는 규칙들만을 얻는 기법을 제안한다. 먼저 일정 단위의 통신망을 구성하고 있는 전송 장치(예:라우터)에서 일정 기간동안 일정 주기로 트래픽 발생 정보를 수집하였다. 수집된 정보는 데이터 베이스에 저장되고 저장된 데이터 베이스에 연관 규칙을 적용하여 트래픽 패턴을 나타내는 규칙들을 추출하였다.

생성된 규칙들 중에는 동일한 의미를 갖는 중복적

인 규칙들이 존재하게 되는데 이러한 규칙들은 1차 가지치기(pruning) 과정을 거쳐 제거된다. 1차 가지치기 후 트래픽 분산에 도움이 되는 정보를 담고 있는 규칙들을 추출하기 위해 네트워크 구성 정보 및 관련 노드들의 링크 연결상태, 중간 경로 상의 지연시간, 대역폭 등의 정보를 기반으로 2차 가지치기(pruning)를 수행한다. 이러한 과정을 거쳐 최종적으로 얻어진 규칙들은 해당 네트워크의 트래픽 분산 정책을 수립하는데 유용하게 활용 될 수 있다.

본 논문의 구성은 다음과 같다. 제 2장에서는 관련 연구 분야인 연관 규칙에 대해 논의하고, 3장에서는 본 논문에서 제안한 효율적인 트래픽 분산을 위한 연관 규칙 추출 기법에 대해서 논의한다. 제 4장에서는 본 논문에서 제안하고 있는 효율적인 트래픽 분산을 위한 연관 규칙 추출 기법을 위한 가지치기(pruning) 과정에 대하여 설명을 하고, 마지막으로 5장에서는 결론과 향후 연구 방향에 대해서 논한다.

2. 연관 규칙 (Association Rule) 탐사

연관 규칙이란 동시에 발생하는 사건 그룹 내에서 사건들 사이에 존재하는 패턴을 알아내기 위한 것으로, 조사하려는 것이 무엇인지에 대한 개략적인 개념을 이미 가지고 있는 경우에 유용하다. 연관 규칙의 예로, 마켓 데이터에서 "전체 트랜잭션 중 40%가 기저귀와 맥주를 구입하였고 기저귀를 구입한 트랜잭션 중 60%가 맥주를 구입하였다"라는 규칙을 들 수 있다. 즉, 연관 규칙은 트랜잭션 안에서 항목들간의 연관 관계를 나타낸다.[6,7]

연관 규칙은 $X \Rightarrow Y$의 형식으로 표현되며 $X \subset U$, $Y \subset U$이고 $X \cap Y = \phi$을 만족한다. X를 전제부(antecedent), Y를 결과부(consequent)라고 부른다. 여기서 X를 포함하는 트랜잭션 중 C%가 Y도 포함한다면, "연관 규칙 $X \Rightarrow Y$는 C%의 신뢰도를 가진다"라고 말한다. 그리고 전체 트랜잭션 가운데 S%가 $X \cup Y$를 포함한다면, "연관 규칙 $X \Rightarrow Y$는 S%의 지지도를 가진다"라고 말한다. 즉, 연관 규칙 "IF X THEN Y" 에 대한 지지도와 신뢰도는 다음과 같이 계산된다.

지지도(Support) $\quad S(X \Rightarrow Y) = \frac{n(X \cap Y)}{n(U)}$

신뢰도(Confidence) $\quad C(X \Rightarrow Y) = \frac{n(X \cap Y)}{n(X)}$

X와 Y에 대한 연관 규칙은,

" IF X THEN Y WITH Conf(C) & Supp(S) "

로 표현된다.

연관 규칙의 탐사를 위한 기존 방법론 연구로는 빈발 항목 집합을 탐사하는 알고리즘인 Apriori 와 Apriori를 개선한 알고리즘인 DHP 등이 있다.[7]

3. 연관 규칙을 이용한 트래픽 패턴 추출

네트워크 관리의 근본 목표는 방대하고 복잡한 네트워크에서 매일 발생하는 네트워크 상태 정보를 기반으로 네트워크가 최대의 효율성을 갖도록 하는 데 있다. 이를 위해서는 네트워크 상태 정보에 대한 지능적인 분석이 선행되어야 한다.

네트워크의 트래픽 발생에 대한 패턴을 얻기 위해서 연관 규칙을 적용 할 수 있다. 즉, 네트워크로부터 일정 기간 동안 수집된 트래픽 발생 정보에 연관 규칙을 적용하여 얻은 규칙은 해당 기간 동안 트래픽이 어떠한 형태로 발생되는가를 설명 해 준다.

본 절에서는 본 논문에서 수행한 연구를 위하여 사전에 요구되는 트래픽 정보 수집 및 연관 규칙을 이용한 트래픽 패턴 추출에 대하여 설명한다.

3.1 트래픽 정보 수집

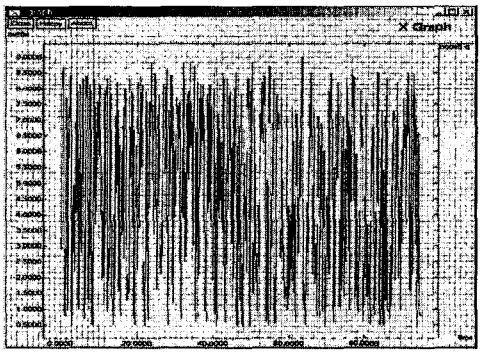
트래픽 발생 정보를 수집하기 위해서 네트워크 시뮬레이터를 이용하여 12개의 라우터를 갖는 가상의 네트워크를 구성하였다. 네트워크에 연결된 컴퓨터 단말은 5개로 가정하였고 각 단말의 트래픽 발생량 및 발생된 트래픽의 목적지는 랜덤 모델로 적용하였다. 랜덤 모델을 위하여 두 개의 간단한 random number 생성기를 이용하였고, 발생된 random number에 따라 트래픽의 양 및 목적지가 결정된다.

각 라우터에서 트래픽의 수집은 매 30초 마다 이루어 졌으며 트래픽 정보 벡터는 [노드 ID, 트래픽 상태]로 구성하였다. "노드 ID"는 네트워크 내에서 노드를 나타내는 유일한 식별자이고 "트래픽 상태"는 노드의 버퍼 상태에 기반하여 아래 <표.1>과 같은 트래픽 흐름에 관련되는 4 가지의 정보로 구분된다.

예를 들면, 버퍼의 적재 상태가 75% 이상인 노드는 트래픽 상태가 "High"로 구분된다.

State	Description
High	75 % 이상
Medium	50 % 이상 ~ 75 % 미만
Low	25 % 이상 ~ 50 % 미만
Idle	25 % 미만

<표.1> 네트워크 노드의 트래픽 상태 모델링



<그림.1> 네트워크 노드의 트래픽 적재 상태

<그림.1>은 하나의 네트워크 노드에서 버퍼의 적재 상태를 나타내는 예이다.

3.2 트래픽 패턴 추출

수집된 트래픽 데이터는 각 노드 버퍼의 트래픽 적재 상태의 관점에서의 전체 네트워크의 운용 상태를 설명하는 규칙들을 발견하기 위해 사용된다. 예를 들어, 네트워크 상에서 일정 기간동안에 발생된 트래픽 정보를 저장하고 있는 데이터 베이스에 연관 규칙을 적용하면 다음과 같은 일례의 규칙을 얻을 수 있다.

IF Node A = High
THEN Node C = Idle
WITH Conf(0.5) & Supp(0.1)

이 규칙은 해당 기간동안에 노드 A가 High 일 때 노드 C가 Idle 상태의 신뢰도 0.5와 지지도 0.1을 가지고 발생함을 의미한다.

먼저 네트워크 노드들의 트래픽 발생 패턴의 규칙들을 생성하기 위해 최소 지지도와 신뢰도를 바탕으로 데이터 내에 존재하는 모든 규칙들이 생성될 수 있도록 하였다.

<표.2>는 전체 네트워크 노드간의 트래픽을 측정한 결과로 네트워크의 서로 다른 노드들간의 상호 연관성을 나타낸 것이다.

1	[N5=High]	⇒ [N0=Low]	supp(0.25) conf(0.72)
2	[N3=Low]	⇒ [N4=Idle]	supp(0.35) conf(0.55)
3	[N6=High]	⇒ [N8=Low]	supp(0.34) conf(0.65)
4	[N7=Idle]	⇒ [N6=High]	supp(0.23) conf(0.36)
5	[N3=High]	⇒ [N5=High]	supp(0.25) conf(0.72)
6	[N9=High]	⇒ [N4=Idle]	supp(0.30) conf(0.55)
7	[N10=Idle]	⇒ [N5=Medium]	supp(0.22) conf(0.44)
8	[N5=High]	⇒ [N7=Idle]	supp(0.35) conf(0.55)
9	[N4=Idle]	⇒ [N1=Low]	supp(0.30) conf(0.47)
10	[N0=Low]	⇒ [N5=High]	supp(0.27) conf(0.51)

<표.2> 트래픽 패턴의 연관 규칙 결과

4. 규칙 분석 및 가지치기(pruning)

<표.2>의 결과에서 보듯이 연관규칙을 통해 생성된 많은 규칙들에는 네트워크 관리에 필요한 규칙 및 불필요한 규칙들이 내포되어 있다. 이중에서 목적에 맞는 유용한 규칙들만 추출하기 위해서 다음과 같은 과정을 거친다.

4.1 1차 가지치기(Pruning)

추출된 트래픽 패턴을 나타내는 연관 규칙의 결과에서 네트워크 노드가 High일 경우에 이 노드의 트래픽을 유연하게 분산시키는데 연관성을 가지는 노드들을 추출한다. 즉, 관심 대상이 되는 규칙들은 어떤 노드에 트래픽이 빈번하게 발생될 때 이 트래픽을 분산시킬 수 있는 노드를 포함하고 있는 규칙들의 집합이다. 먼저, 특정 네트워크 노드의 트래픽 상태가 High 일 때 트래픽이 빈번하지 않는 주변의 네트워크 노드를 찾아낸다. 이는 라우팅 경로 재 설정 등과 같은 트래픽의 분산을 통해 효율적인 네트워크를 운용하는데 이용될 수 있다.

<표.2>에서 보면 High 일 때 트래픽을 분산시킬 수 있는 규칙은 두 번째, 일곱 번째, 아홉 번째 규칙을 제외한 규칙 집합이다. 추출된 규칙 집합에서 첫 번째 규칙과 마지막 규칙은 전제부와 결과부가 다르지만 의미적으로 동일하다. 이러한 중복된 규칙과 다섯 번째 규칙처럼 트래픽을 분산시키는데 유용하지 않은 규칙들은 제거된다. 1차 가지치기(pruning) 과정을 <표.2>에 적용한 결과가 <표.3>에 나타나 있다.

1	[N5=High] ⇒ [N0=Low]
2	[N6=High] ⇒ [N8=Low]
3	[N7=Idle] and [N6=High] ⇒ [N1=Low]
4	[N9=High] ⇒ [N4=Idle]
5	[N5=High] ⇒ [N7=Idle]

<표.3> 1차 가지치기(pruning) 결과

4.2 2차 가지치기(pruning)

1차 가지치기(pruning)된 결과만으로도 네트워크에서 특정 노드의 라우팅 경로 재 설정 및 트래픽 분산에 도움을 줄 수 있다. 그러나 실제 네트워크의 물리적인 구성을 전혀 반영하지 못하고 있는 규칙들도 포함하고 있다. 따라서, 트래픽 분산에 도움이 되는 정보를 가지고 있는 규칙들을 추출하기 위해서 트래픽 패턴을 나타내는 연관규칙에 네트워크의 구성 정보를 반영하여 2차 적인 규칙 제거를 실행하여야 한다. 먼저, 1차 가지치기(pruning)된 규칙 집합의 규칙에서 네트워크 구성 정보를 기반으로 전제부와 결론부 사이의 전송 가능한 중간 노드들을 찾아낸다. 중간 경로 상의 대역폭과 지연 시간 등을 고려하여 전송될 중간 노드를 선택하고 중간 노드의 버퍼 적재 상태를 확인하여 전송 할 것인지를 결정한다. 전송될 중간 노드의 상태가 High일 경우 트래픽이 분산되는데 도움을 주지 못하므로 제거한다.
예를 들면, 위에서 보인 <표.3> 규칙에서 네트워크 노드 N5 와 N0 사이에 하나의 중간 노드를 갖고 이 노드의 상태가 High일 경우 이 규칙으로 인해 더 많은 양의 트래픽이 발생된다.

1	[N6=High] ⇒ [N8=Low]
2	[N7=Idle] and [N6=High] ⇒ [N1=Low]
3	[N9=High] ⇒ [N4=Idle]
4	[N5=High] ⇒ [N7=Idle]

<표.4> 2차 가지치기(pruning) 결과

<표.4>는 네트워크의 물리적인 구성 정보가 반영된 2차 가지치기 결과를 나타낸다.

5. 결론 및 향후 연구

네트워크 노드에서 발생하는 트래픽 패턴들을 수집하고, 효율적으로 트래픽 패턴을 분석하여 조치하는 것은 매우 어렵고도 중대한 문제이다. 본 논문에서는 네트워크에서 발생하는 트래픽을 분석하여 특정 노드에서의 과부하를 감소시키고 네트워크를 효율적으로 운영할 수 있는 기법을 제안하였다. 각 노드에서 발생되는 트래픽 패턴을 연관 규칙을 이용하여 추출하고 트래픽의 발생 패턴 중에서 라우팅의 효율성을 높이는데 적용이 가능한 규칙들만을 얻기 위한 가지치기(pruning)를 수행하였다.

최종적으로 얻어진 규칙들은 네트워크 운영자 및 지능적인 네트워크 관리 시스템 구현 등을 위한 의사 결정 지원에 이용 될 수 있다.

향후에는 이러한 규칙들을 토대로 실제 네트워크 관리 시스템에 적용하고, 트래픽 분산에 관련되는 두 노드의 중간 경로 상에 위치한 노드들의 상태를 반영할 수 있는 다양한 방법들에 대한 지속적인 연구가 필요하다.

6. 참고 문헌

[1] Mani Subramanian "Network Management :Principles and Practice" , Addison-Wesley.

[2] [ITU-T] Recommendation X.700: Management Framework for Open Systems Interconnection for CCITT applications, september 1992

[3] Tim Oates, "Fault identification in computer networks: A review and a new approach." Technical Report pp. 95-113, Computer Science Department 1995

[4] W. J. Frawley, G. Piatetsky-Shapiro and C. J. Matheus, "Knowledge Discovery in Database : An Overview", AAAI Press, pp.1-27, 1991

[5] R. Agrawal, T. Imielinski and A. Swami, "Database Mining: A Performance Perspective", IEEE Transformation on Knowledge and Data Engineering, pages 914-925, 1993

[6] R. Agrawal, T. Imielinski and A. Swami, "Mining Association Rules between Sets of Items in Large Database", In Proceedings of ACM SIGMOD, pp. 207-216, 1993

[7] R. Agrawal and R. Srikant, "Fast Algorithms for Mining Association Rules", In Proceedings of the 20th International Conference on Very Large Database, pp 478-499, 1994