패킷 마이닝을 위한 분산 시스템의 부하 균형

옥지혜[0] 곽미라 조동섭
이화여자대학교 과학기술대학원 컴퓨터학과
{okwisdom[0], mirakwak, dscho}@ewha.ac.kr

Load Balancing in Distributed System for Packet Mining

Ji-hye Ok[0] Mira Kwak Dong-sub Cho
Dept. of Computer Science and Engineering, Ewha Womans University

요 약

네트워크로 연결된 컴퓨터 기술은 컴퓨터의 향상된 처리 능력과 컴퓨터 통신기술과 결합하여 자원을 보다 유용하게 이용하는 분산처리 기법이 발달하게 되었다. 이러한 분산 시스템은 단일 컴퓨터 시스템에서 시스템에 미치는 영향을 적게 함으로서 신뢰도를 높일 수 있고 저렴한 비용으로 더 큰 성능을 얻을 수 있다. 본 연구에서는 실시간으로 생성되는 패킷 데이터를 효율적으로 처리하고 분석하는데 있어서 분산 시스템을 이용하여 해결 하고자 한다. 사용자 행동으로부터 생성되는 패킷을 IP별로 분리하여 각각의 분산된 시스템에서 처리하고 이렇게 처리된 데이터를 관리자가 모니터링 할 수 있도록 부하균형을 이룬 패킷 마이닝 분산 시스템을 제안하고자 한다.

1. 서 론

네트워크에서 사용되는 정보들은 수많은 패킷으로 구성되어 송수신 되는데 이러한 패킷의 정보를 이용하여 많은 정보를 알아낼 수 있다. 패킷의 정보를 통계적으로 분석해 외부로부터의 침입과 정보의 유출을 방지할 수 있으며 네트워크의 문제점을 파악하여 시스템을 안전하게 관리 할 수 있다. 그리고 각종 프로토콜을 분석해 네트워크의 부하와 사용자의 행위 패턴과 요구사항을 알아낼 수 있다. 그러나 사용자의 시스템에서 실시간으로 나오는 패킷을 하나의 시스템에서 처리하고 분석하는 경우 많은 부하가 생긴다.

본 연구에서는 이를 효율적으로 해결하기 위해 패킷 데이터를 분산하여 처리하는 시스템을 제안하고자 한다. 제안된 패킷 마이닝 분산 시스템을 통해 사용자 행동에 의해 생성된 패킷 정보들이 IP별로 분리되고 분산된 각 시스템에서 처리되며 관리자가 이러한 데이터를 모니터링 할 수 있다.

본 논문의 구성은 다음과 같다. 2장에서는 분산처리 시스템과 최적화 기술을 살펴보고, 3장에서는 네트워크에서 정보전송의 단위인 패킷정보 기술을 설명 할 것이다. 4장에서는 본 논문에서 제안하는 패킷 마이닝을 위한 분산시스템의 부하균형에 대하여 설명하고, 5장에서는 결론과 향후 연구 과제를 논의 할 것이다.

이 논문은 2002년도 두뇌한국21사업에 의하여 지원되었음.

2. 분산처리 시스템과 최적화 기술

2.1 분산처리 시스템

분산처리시스템이란 네트워크에 의해서 서로 연결되는 여러 개의 프로세서들과 네트워크를 말하는데 이 시스템은 특별한 목적을 가지고 서로 결합된 형태로 계획, 설계된 것이라고 할 수 있다. 즉 분산처리시스템은 프로세서가 서로 다른 위치에 존재하며, 프로세서가 원거리통신에 의해서 서로 연결되어 있으며, 이러한 시스템은 결합된 형태로 계획 및 설계되어 있다.

2.2 최적화기술

분산처리 시스템에서의 최적화 기술로는 토큰패싱기법, DR(Dead Reckoning)기법, AOI(Area Of Interest)기법, Packet Aggregation기법이 있다.

2.2.1 토큰 패싱기법

중앙화된 공유상태 저장소를 갖는 경우 한번에 한 사용자만이 공유상태 저장소에 접근하여 상태를 변경시킬 수 있는 기법이다. 이 방법은 일관성을 확실하게 유지 시켜줄 수 있으나 실시간성이 요구되거나 공유 상태의 변화가 빈번한 경우 그리고 사용자가 많아지는 경우에는 사용이 어렵다.

2.2.2 DR(Dead Reckoning)기법

공유상태를 변화시킨 사용자가 다른 사용자들에게 공유

상태의 데이터들이 특정조건이 만족 될 때에만 샘플링 하여 보내준다. 이 기법은 사용자의 증가에 따라 어느 정도의 성능과 일관성을 유지시켜 준다.

2.2.3 AOI기법

서버가 사용자로부터 받은 이벤트들을 다른 모든 사용자들에게 다시 보내는 것이 아니라 그 이벤트의 관심 영역에 속해있는 사용자들에게만 이벤트를 보냄으로서 Packet Filtering 효과를 나타내어 트래픽을 줄여 준다.

2.2.4 Packet Aggregation기법

분산 환경에서 사용자들간의 트래픽을 줄이기 위한 한 방법으로서 전송 패킷이 발생할 때마다 보내는 것이 아니라 여러 개를 모아서 같이 보내는 방법이다. 여러 개를 같이 보낼 때 헤더는 하나만 보내므로 각각을 보낼 때 보다 트래픽이 줄어든다.

3. 패킷 정보를 이용한 기술

인터넷에서 사용되는 정보들은 수많은 패킷으로 구성되어 송수신 되는데 이러한 패킷의 정보를 이용한 여러 기술에 대하여 알아본다.

3.1 Protocol Probing

정보를 교환함에 있어 두 지점 사이에서 패킷 전송 수신에 따른 프로토콜의 사용은 필수라고 할 수 있다. 네트워크에서 사용되는 프로토콜은 현재 RFC 1700에 등록되어 있는 프로토콜들 이외에도 특정목적을 위해 사용되거나 새로이 만들어 사용하고 있는 프로토콜들이 있다. 현재 사용중인 프로토콜을 이용한 침입이나 새로운 프로토콜을 이용하여 시스템 내부로 침입할 가능성에 대한 탐지를 목적으로 한다.

3.2 IP Probing

인터넷에서의 주소는 정보를 정확한 목적지로 송수신 할 때에 사용되는 꼭 필요한 요소이다. 패킷의 헤더에는 전송지의 주소와 목적지의 주소가 입력되어 있는데 이러한 주소로부터 사용자의 행위 탐지와 침입탐지의 중요한 자료로서 활용될 수 있다.

3.3 Port Probing

IANA(Internet Assigned Numbers Authority)에 의해 할당된 Port번호 이외의 1024 이상의 번호를 갖는 Port들을 사용하여 들어오는 사용자를 탐지하는 기법이다. 시스템의 감시를 피해 시스템의 내부로 들어오는 사용자의 시도를 탐지한다. 보안에 취약한 서비스에 대한 접근을 탐지한다.

3.4 Source Routing

IP 헤더의 Option Field를 검사하여 라우팅 경로에 대한 옵션이 있는 지를 확인하여 옵션이 셋(Set) 되어 있으면 점검을 실시한다. 라우터에 의해 패킷이 제어되는 것을 피하기 위해 의도적으로 라우팅 경로를 조작했는지를 점검하고 라우팅 경로에 의심스러운 경로가 있는지를 점검한다. 또한 사전에 축적된 정보를 기반으로 예상 라우팅 경로 이외의 경로가 패킷 내의 라우팅 경로에 포함되어 있는지를 점검한다.

3.4 Pattern Matching

사용자가 어떤 시스템에 대하여 접근을 시도하려 할 때 사용되는 특정 유형들에 대한 패턴을 알아낸다. 관리자의 실수나 시스템의 오류로 인해 발생하는 사항들에 대하여 사전 점검을 통해 알아낸다. 불법적인 의도를 가지고 정확한 신분 인증 과정을 요구하지 않는 ID를 사용하여 시스템에 접속하거나 root로 접근시도를 탐지하여 침입의 가능성이 있는 패턴들을 가진 패킷을 구분하여 침입여부를 판정한다.

4. 패킷 마이닝을 위한 분산 시스템의 부하균형

본 연구에서는 네트워크 패킷 정보를 분석하여 이전에 알려지지 않은 네트워크 시스템 상의 정보를 발견하여 안정적이고 효율적인 네트워크 환경을 제공하고자 한다. 네트워크상의 패킷 정보를 하나의 시스템이 모두 분석하여 처리하기에는 시스템에 많은 부하가 생긴다. 따라서 제안하는 시스템은 하나의 시스템에서 모든 패킷을 처리하지 않고 분산된 시스템으로 처리하고자 한다.

4.1 효율적인 분산 시스템

네트워크로 연결된 분산된 두 대의 시스템이 실시간으로 생성되는 패킷의 정보를 처리하는 것을 그림1에서 나타내었다. 각각의 시스템은 네트워크에 연결된 시스템의 IP를 분류하여 패킷정보를 수집하고 그것을 처리하여 저장한다. 이러한 분산 시스템을 이용하여 네트워크 상에서 흐르는 패킷 정보를 처리하면 안정적이고 효율적으로 처리 가능하다.

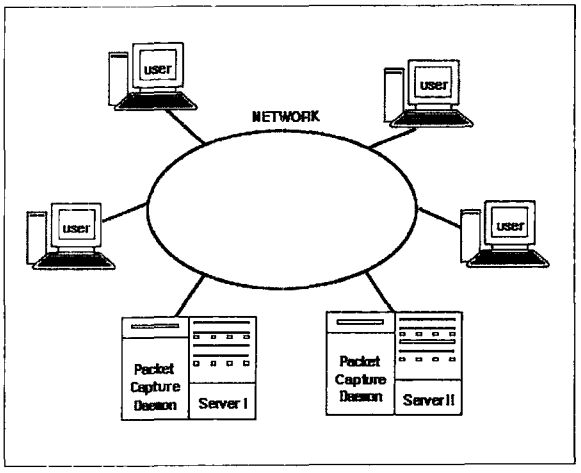

그림 1 부하 균형을 위한 분산서버 구조

4.2 패킷 마이닝을 위한 분산처리 시스템 세부구조

네트워크에서 전송되는 패킷을 잡아 분산된 시스템에서 패킷 정보를 처리하고 관리하는 것을 그림2를 통해 상세히 살펴보고자 한다. 본 시스템의 주요 구성요소를 살펴보면 네트워크의 인터페이스 장치로부터 들어오는 패킷을 IP 별로 구분하여 분산된 시스템으로 보내는 패킷 수집부, 패킷 정보를 분석하고 관리하기 쉬운 형태로 처리하는 데이터 처리부, 가공된 데이터를 가지고 시스템의 안전성과 인터넷서비스별 사용기록을 분석하는 패킷데이터 분석부와 관리자가 검사 및 분석을 수행하고 그에 관한 보고를 받을 수 있는 보고서 생성부가 있다. 제안하는 패킷 마이닝을 위한 분산 서버 시스템의 보다 세부적인 사항들은 다음과 같다.

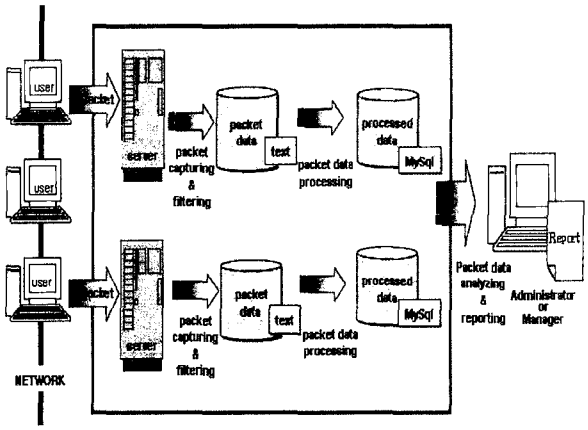

그림 2 패킷 분산처리 시스템의 세부 구조

4.2.1 패킷 수집부

로컬 네트워크 상에 있는 여러 종류의 패킷들이 패킷 마이닝을 위한 분산시스템에 접속하면 IP별로 구분하여 패킷을 수집하여 전달한다. 수집되는 패킷들은 텍스트형태로 저장된 후 일정시간이 지나면 데이터 처리부가 정보들을 처리한다. 이렇게 수집된 패킷은 이후 분석부에서 마이닝기법의 적용을 통해 시스템의 안전과 서비스별 행위분석을 위한 판단 자료로서 사용된다.

4.2.2 데이터 처리부

패킷 수집부로부터 저장되어진 패킷 데이터를 분석부로 데이터를 전송한다. 프로토콜별 시스템의 사용정보와 시스템 서비스별 사용정보를 생성하여 내장 데이터베이스 엔진에 고유한 형식으로 저장한다.

4.2.3 데이터 분석부

패킷이 가지고 있는 정보에 대해 마이닝 알고리즘을 적용하여 침입 판단 여부를 알아내고 서비스별 사용정보를 분석하여 패킷으로부터 알려지지 않은 정보를 새로 발견한다. 패킷 헤더로부터 프로토콜별 시스템 사용기록과 시스템 서비스별 사용기록데이터를 읽어 서비스에 관련된 여러 정보를 종합하고 특정 패킷에 관한 축적된 정보를 분석하여 시스템의 침입을 탐지한다. 그리고 일정 시간동안 특정 항목을 감시하여 얻은 정보를 근거로 하여 침입여부를 판정하고 서비스별 사용자의 행동도 분석한다. 분석부에서 수집된 사용자의 분석결과는 관리자에게 제공한다.

4.2.4 보고서 생성부

시스템의 안전성을 위해 침입에 대한 검사를 실행하고 서비스별 사용자의 행동을 분석하여 검사 결과를 화면에 출력한다. 관리자는 분석된 결과를 가지고 대응 조치를 실행하여 보안을 증진시키고 사용현황을 분석하여 시스템의 관리를 편리하게 이루어 질 수 있도록 해준다.

5. 결론 및 향후 연구 과제

본 논문에서 제안한 패킷 마이닝을 위한 분산 시스템은 시스템의 네트워크 패킷 정보를 분석하여 안정적이고 효율적인 네트워크 환경을 구축하고자 한다. 이를 위해 분산된 시스템으로 패킷 정보를 처리하고 분석하는 시스템을 설계하였다. 또한 관리자가 서비스별 사용자의 행위와 시스템의 안정성을 검사를 편리하게 할 수 있도록 관리 모듈을 효율적으로 설계하였다.

향후 효율적인 검색 알고리즘을 적용하고 패킷 마이닝 알고리즘을 개발 적용하여 시스템 상에서 일어나는 여러 행위의 패턴을 발견하고 이에 따른 문제를 효율적으로 해결하고자 한다. 그리고 편리한 시스템 관리를 위한 관리 인터페이스를 설계할 것이다.

[참고문헌]

[1] 심광현, 외, "분산 가상환경을 위한 네트워크 서버 기술" 정보과학회지 제19권 제5호, 2001년 5월.

[2] 성재모, "DEC를 이용한 분산 컴퓨팅" 정보과학회지 제 14권 제 1호, 1996년 1월.

[3] 이경하, 은유진, 임채호, 정태명, "네트워크 패킷 정보를 기반으로 한 보안 관리", 정보과학회 논문지(A) 제 25권 제 12호, 1998년 12월.

[4] 분산처리시스템 "http://myhome.hananet.net/~madeweb/process01.html" 2001년 2월.

[5] Busch, Costas, "A study on distributed structures", Brown University, 2000

[6] Liu, Xiaoming, "Building high-performance adaptive communication systems from components", Brown University, 2001