제18회 한국정보처리학회 추계학술발표대회 논문집 제9권 제2호 (2002.11)

Packet Mining을 이용한 Gamer의 행위분석시스템

이미란°, 조동섭
이화여자대학교 과학기술대학원 컴퓨터학과
e-mail:{nayamira°, dscho}@ewha.ac.kr

Design of Gamer's Behavior Analysis System using Packet Mining

Mi-Ran Lee°, Dong-Sub Cho
Dept of Computer Science and Engineering, Ewha Womans University

요 약

사용자의 필요를 충족시켜 줄 수 있도록 사용자에 대한 연구가 인터넷 비즈니스에서 활발히 이루어지고 있다. 인터넷 비즈니스와 마찬가지로 컴퓨터 게임 산업 분야에서도 이러한 연구가 필요하다. 하지만 컴퓨터 게임의 특성상 기존의 인터넷 비즈니스 방식과 같은 방법으로는 게이머(Gamer)의 행동을 알기 어렵다. 이러한 문제점을 해결하고자 본 논문에서는 패킷 마이닝(Packet Mining)을 이용한 게이머의 행위 분석 시스템을 제안하고자 한다. 이 시스템은 게이머들과 시스템 사이에 전달된 패킷을 수집하여 프로토콜별 텍스트 형태로 저장하고, 일정 시간이 흐르면 텍스트 형태로 저장된 패킷을 데이터베이스로 생성한다. 게이머 행위 분석 시스템은 이렇게 생성된 데이터베이스를 분석하고, 다양한 정보를 추출해내어 게이머의 행위를 분석한다.

1. 서론

인터넷 비즈니스에 대한 관심이 대두되면서 인터넷을 통해 사용자들에게 맞춤 서비스를 제공하여 웹 사이트의 재방문 횟수를 늘리고, 판매수익을 높이고자 다양한 방법들이 연구되어지고 있다. 인터넷 비즈니스에서는 그 방법의 하나로 웹 마이닝(Web Mining)을 이용하고 있는데, 웹 마이닝은 웹 상의 문서 혹은 서비스로부터 유용한 정보를 발견하는 기법을 의미한다.

인터넷 비즈니스와 마찬가지로 컴퓨터 게임 산업 분야에서도 게이머(Gamer)의 필요를 충족시켜 줄 수 있도록 게이머에 대한 연구가 필요하다. 하지만 컴퓨터 게임의 특성상 기존의 인터넷 비즈니스와 같은 방법으로는 게이머의 행동을 알기 어렵다. 이러한 문제점을 해결하고자 본 논문에서는 패킷 마이닝(Packet Mining)을 이용한 게이머의 행위 분석 시스템에 대해 논하고자 한다.

인터넷에서 사용되는 정보들은 수많은 패킷으로 구성되어 송수신 되는데, 웹 로그만으로는 수집할 수 없는 정보들을 패킷을 이용하여 알아낼 수 있다.

이 논문은 2002년도 BK21사업에 의하여 지원되었음.

우선 각종 프로토콜과 포트들을 통해 전달되는 패킷들을 분석하여 게이머의 행위를 분석해 볼 수 있고, 또 패킷의 정보들을 통계적으로 분석해서 외부로부터의 침입과 정보의 유출을 방지할 수 있으며, 네트워크의 문제점을 파악하여 시스템을 안전하게 관리 할 수 있다. 그리고 각종 프로토콜을 분석해 네트워크의 부하와 그에 따른 요구사항 등도 알아낼 수 있다.

패킷 마이닝을 이용한 게이머의 행위 분석 시스템은 게이머들과 시스템 사이에 전달된 패킷을 수집하여 프로토콜별 텍스트 형태로 저장하고, 일정 시간이 흐르면 텍스트 형태로 저장된 패킷을 데이터베이스로 생성한다. 이렇게 생성된 데이터베이스를 분석하고, 다양한 정보를 추출해내어 게이머의 행위를 분석한다

논문의 순서는 다음과 같다. 먼저 2장에서는 데이터 마이닝의 개념 및 데이터 마이닝 도구에 대해서 기술하고, 3장에서는 패킷 정보를 이용한 기술에 대해 알아본다. 4장에서는 패킷 마이닝을 이용한 게이머 행위 분석 시스템에 대해 기술하고, 마지막으로 5장에서는 결론 및 향후 과제로 본 논문을 맺는다.

2. 데이터 마이닝의 개념 및 데이터 마이닝 도구

2.1 데이터 마이닝의 개념

데이터 마이닝은 대량의 실제 데이터로부터 묵시적이고 미리 알 수는 없지만 잠재적으로 쓸모 있는 정보를 추출하는 작업으로 정의한다. 특히 단순 의사 결정 정보와의 차별성을 위하여 데이터 마이닝은 대규모 과거 자료의 패턴 추세를 분석하기 위한 방안으로 사용자가 가설을 제공하고 통계학적으로 결과를 검증하는 기존의 통계학적 개념이 아닌 시스템 내부에서 연관규칙, 연속패턴, 분류규칙, 유사 시계열 등의 알고리즘을 이용하여 사용자의 가설 제공 없이 데이터간의 의미를 자동적으로 분석하여 유용한 정보를 추출한다.

데이터 마이닝은 기계학습, 통계학, 인공지능 등 많은 다른 연구 결과로부터 발전한 기술로서 대량 데이터 정보처리 기술, 영상 시각 기술을 추가적으로 응용하여 데이터 웨어하우스와 함께 급속히 발전하는 최신 데이터 분석 기술이다. 데이터 마이닝은 대규모 데이터 중에서 숨겨진 관계와 패턴을 발견하는 것으로 데이터의 가치 있는 지식을 표현한다. 기본적으로 데이터 마이닝은 이전에 알려지지 않은 데이터의 규칙, 패턴을 발견하는 사용자 친화적인 소프트웨어 기술이다. 사용자 친화적인 의미는 기존의 데이터 분석 방법인 전통적인 통계 기법보다 데이터 분석이 훨씬 쉽다는 뜻이다. 기존의 통계학으로의 데이터 분석은 사전에 가설, 또는 모델을 설정하고 전문적인 데이터 분석가가 반복적인 시뮬레이션을 통해 가설을 검증한 반면 데이터 마이닝은 이러한 모든 작업을 전적으로 대규모 데이터를 처리하는 데이터 마이닝 소프트웨어가 처리한다.

2.2 데이터 마이닝 도구

데이터 마이닝 기술은 분류규칙, 연관규칙, 연속패턴, 유사 시계열 등을 중심으로 연구 중이다.

분류 규칙은 클래스 레벨이 이미 알려진 데이터인 트래이닝 데이터의 집합을 분석하고 이러한 데이터의 성질을 기초로 하여 데이터베이스를 클래스로 분류한 뒤 이 클래스에 영향을 줄 규칙의 집합을 공식화한다.

연관규칙은 시험된 레코드 중에서 반영된 데이터들 사이의 유사성을 인식하는 것이다. 이러한 유사성은 규칙으로 표현된다. 이러한 발견 등에 대한 퍼센트는 연관성의 신뢰요인으로 사용된다. 대표적 응용으로 시장 바구니 분석은 연관 규칙을 제품의 유사성을 인식하기 위한 POS(Point Of Sales) 트랜잭션 데이터에 적용한 것이다.

연속패턴은 연관규칙의 변형으로 시간적인 관계를 가지고 있다. 예를 들어 배낭을 구입한 고객이 다음에 텐트를 구입하는 경향이 있는 경우, 배낭의 구입과 텐트의 구입에는 연속성이 존재한다는 의미를 부여한다.

유사 시계열의 탐사는 주식, 물가, 판매량, 과학적 실험 데이터 등이 시간 축으로 관측값 등이 나열된 시계열의 형태를 갖게 되는 경우 서로 유사한 시계열 데이터들을 발견하는 것이다. 예를 들면 특정 물품의 판매량 추이와 비슷한 판매량 추이를 갖는 다른 물품을 발견할 수 있다면 두 물품의 판매 전략에 유사성을 부여할 수 있다.

3. 패킷 정보를 이용한 기술

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 침입 탐지의 중요한 자료로서 활용될 수 있다.

3.3 Port Probing

IANA(Internet Assigned Numbers Authority)에 의해 할당된 Port 번호 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 있는지를 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.

3.5 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 잇는 패턴들을 가진 패킷을 구분하여 침입여부를 판정한다.

4. Packet Mining을 이용한 행위 분석 시스템

본 논문에서 제안하는 패킷 마이닝을 이용한 게이머의 행위 분석 시스템은 게이머 즉 사용자들과 시스템 사이에 전달된 모든 패킷을 수집하여 프로토콜별 텍스트 형태로 저장한다. 일정 시간이 지나면, 텍스트 형태로 저장한 패킷을 사용하기에 적합한 형태의 정보로 가공하여 프로토콜별 데이터베이스로 생성한다. 이렇게 생성한 데이터베이스의 프로토콜별 사용정보를 분석하여 패킷으로부터 게이머에 대한 다양한 정보를 추출해낸다. 마지막으로 추출해낸 결과를 화면에 출력하거나 파일로 저장한다.

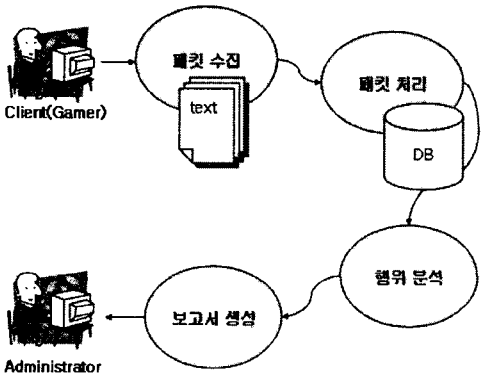

그림 1. Packet Mining System 구성도

그림 1은 패킷 마이닝을 이용한 게이머의 행위 분석 시스템의 전체적인 구성도를 보여준다.

앞에서 설명한 패킷 마이닝 시스템은 크게 패킷 수집 단계, 패킷 처리 단계, 행위 분석 단계, 보고서 생성 단계로 나눠진다.

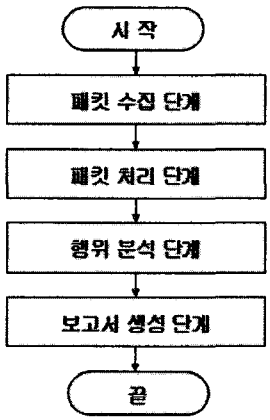

그림 2. 시스템 순서도

4.1 패킷 수집 단계

패킷 수집 단계에서는 시스템에 접속한 게이머들과 시스템 사이에서 전달된 모든 패킷들을 수집한다. 이 때 패킷을 수집하기 위하여 sniffering 기술을 사용한다. sniffering 기술은 네트워크 상의 패킷들을 실시간으로 수집하여 프로토콜을 분석하는 기술이다. 수집되는 패킷들은 프로토콜별로 텍스트 형태로 저장된다.

4.2 패킷 처리 단계

패킷 처리 단계에서는 패킷 수집 단계에서 텍스트 형태로 저장된 패킷을 일정 시간이 지나면, 행위 분석 단계에서 사용하기에 적합한 형태의 정보로 가공하여 데이터를 저장한다. 이때 패킷 수집 단계에서 텍스트 형태로 저장된 내용을 패킷 처리 단계에서 프로토콜별 시스템의 사용정보로 데이터베이스화 한다.

4.3 행위 분석 단계

패킷이 가지고 있는 정보에 대해 마이닝 알고리즘을 적용하여 프로토콜별 사용정보를 분석하여 패킷으로부터 알려지지 않은 정보를 발견한다. 패킷 헤더로부터 프로토콜별 시스템 사용기록 데이터를

읽어 서비스에 관련된 여러 정보를 종합하고 특정 패킷에 관한 축적된 정보를 분석하여 게이머의 행동을 분석한다.

4.4 보고서 생성 단계

보고서 생성 단계에서는 행위 분석 단계에서 추출한 게이머의 분석 결과를 화면에 출력하거나 파일로 저장한다. 관리자는 분석된 결과를 보고 게이머의 사용현황을 알 수 있고, 또 시스템의 관리를 편리하게 할 수 있다.

5. 결론 및 향후 연구 과제

본 논문에서 제안한 패킷 마이닝을 이용한 게이머의 행위 분석 시스템은 컴퓨터 게임 산업 분야의 특성을 고려하여 웹 로그 분석 대신, 패킷 마이닝을 이용하여 게이머의 행동을 분석하는 시스템이다.

패킷 마이닝은 컴퓨터 게임 산업 분야뿐만 아니라 기존의 인터넷 비즈니스, 포털 사이트, 보안 사이트 등에서도 이용될 수 있는 개념이다. 따라서 앞으로 효율적인 패킷 마이닝 알고리즘을 개발, 적용하여 시스템 상에서 일어나는 여러 가지 행위 패턴을 추출해 낸다면, 인터넷 산업 분야에서 효율적으로 사용될 수 있을 것이다.

향후 수집한 패킷을 효율적으로 마이닝 할 수 있는 알고리즘을 개발하고, 이 개념을 여러 분야에 적용시켜 실제로 활용해 볼 것이다.

참고문헌

[1] D.N. Doan, K.R. Narayanan, "Iterative packet combining schemes for intersymbol interference channels", IEEE Trans on Communications, Vol.50 No.04, pp.560-570, 2002.

[2] T.H. Nguyen, M.N.O. Sadiku, "Next generation networks", IEEE Potentials, Vol.21 No.02, pp.6-8, 2002.

[3] F. Braun, J. Lockwood, M. Waldvoge, "Protocol wrappers for layered network packet processing in reconfigurable hardware", IEEE Micro, Vol.22 No.01, pp.6-74, 2002.

[4] J.Koulouris, "Planning telecommunications for the Athens 2004 Olympic Games", IEEE Communications Magazine, Vol.39 No.07, pp.100-104, 2001.

[5] 이미란, 정옥란, 조동섭, "Gamer 행위 분석을 위한 Game Log Mining System의 설계", 한국게임학회 하계학술발표논문집, pp.223-225, 2002.

[6] 옥지혜, 곽미라, 조동섭, "패킷 마이닝을 위한 분산 시스템의 부하 균형", 한국정보과학회 춘계학술발표논문집, Vol.29 No.1, pp.559-561, 2002.

[7] 김형택, 민옥길, "효과적인 인터넷 마케팅을 위한 웹 로그 분석", 도서출판 비비컴, pp21-26, 2001.

[8] 김영남, "데이터 웨어하우스를 활용한 데이터 마이닝 기술", 한국정보처리학회 논문지, 제4권 제6호, pp.171-178, 1997.

[9] 임경하, 은유진, 임채호, 정태명, "네트워크 패킷 정보를 기반으로 한 보안 관리", 정보과학회 논문지, 제25권 제12호, 1998.