가상환경에서 멀티미디어 데이터를 사용하는 사용자 행동 분석 시스템 개발

이윤경*, 이민수*, 손유승**, Gunnar Wallmann**, Miguel Fernandes**
*이화여자대학교 컴퓨터정보통신학과
**Manufacturing Enterprises Security Team, Institute for Graphic Interfaces
e-mail:narouge@ewhain.net

A user behavior monitoring system on multimedia data for virtual engineering environments

Yoon-Kyung Lee*, Minsoo Lee*,
Yuseung Sohn **, Gunnar Wallmann **, Miguel Fernandes **
* Dept. of Computer Science and Engineering, Ewha Womans University
** Manufacturing Enterprises Security Team, Institute for Graphic Interfaces

요 약

사용자의 행동을 모니터링 하는 것에 대한 이전의 기술적인 연구는 네트워크 트래픽과 데이터베이스 접근 패턴에 집중되어 있으나 이러한 접근은 사용자간의 데이터를 교환하고 공유하는 등의 상호 작용을 관찰하기에는 부족하다. 따라서 'BHave' 라는 가상 환경에서 사용자의 행동을 추적할 수 있는 시스템을 개발하여 문서에 접근하는 사용자의 행동을 모니터링한다. 서버쪽의 데이터베이스에서 데이터를 가져와서 클라이언트의 API 를 통하여 사용자가 선택한 데이터를 분석한 뒤 사용자에게 그래프를 통해서 시각적으로 분석 결과를 보여준다.

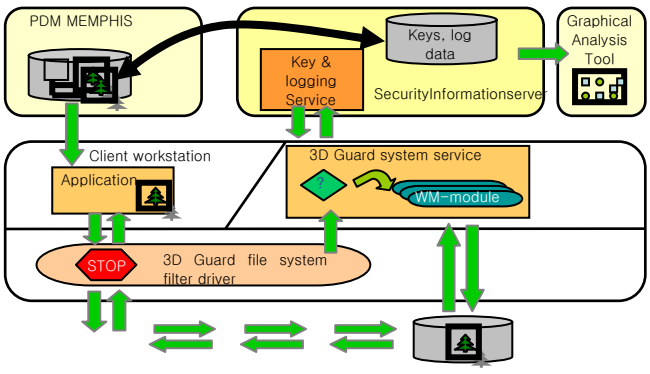
1. 서론

사용자의 행동을 모니터링 하는 것에 대한 이전의 기술적인 연구는 네트워크 트래픽과 데이터베이스 접근 패턴에 집중되어 있으나 이러한 접근은 사용자간의 데이터를 교환하고 공유하는 등의 상호 작용을 관찰하기에는 부족하다.

BHave 라는 가상 환경에서의 사용자 행동을 추적하는 시스템은 문서상의 사용자의 행동을 모니터할 수 있는 기능을 제공한다. BHave 가 수집한 정보는 Security Information Server 라는 중앙 서버에 저장되고, 접근된 문서에 관한 정보와 문서 접근이 일어난 환경에 대한 정보로 분류될 수 있다. BHave 는 흥미롭거나 어딘가 의심스러운 사용자의 행동을 검출할 수 있을 것이다. 게다가 DeviceGuard, FileGuard 같이 IGI 보안 모듈의 가동에 영향을 미칠 수 있을 것이다.

이 프로젝트의 목적은 사용자의 행동 정보를 수집하여 분석한 뒤 그래프를 사용하여 시각적으로 보여주는 소프트웨어 도구를 개발하는 것이다. 이 도구는 사용자가 원하는 차트를 선택할 수 있고, 차후 사용자 행동 패턴을 정의하고 관리하며, 사용자 행동 정보를 시각화하는 도구중의 하나로 발전을 하게 될 것이다.

본 논문의 구성은 다음과 같다. 2 장에서는 기존의 분석 및 레포팅 도구에 관한 내용과 차트 컴포넌트에 대해 기술하며 3 장에서는 시스템의 구조 및 디자인에 대한 내용을 다룬다. 4 장에서는 시스템 구현 및 결과를 설명하며 5 장에서는 결론을 내린다.



(그림 1. BHave 의 개요)

2. 관련연구

Oracle Discoverer [10]와 Hyperion [11]과 같은 몇몇 일반적으로 잘 알려진 자료 분석 툴이 있지만 이러한 도구들은 대부분 OLAP 을 위해 사용되는 도구이다. Oracle Discoverer 는 ad hoc 쿼리, 보고서, 분석 및 Oracle Corporation 의 웹 간행물을 위한 비즈니스 인텔

리전스 도구이고, Hyperion 또한 자료를 수집하고, 분류하고 분석하여 사업 성과 관리를 더 쉽고 편리하게 할 수 있다. Bhave 시스템을 위한 우리의 그래픽 분석 시스템에서는 사용자 행동 자료는 더 효과적인 분석을 지원할 수 있는 방법으로 그래픽을 통해 시각적으로 표시될 수 있다.

2D 와 3D, scatter, bubble 차트 등등 다양한 형태의 그래프를 제공해주는 컴포넌트들로 Nevron Chart for .NET[1], Dundas [6] Graphics Server [7], Sharp Graph [8], TeeChart [9], Chart FX [4], XCEED [5] 등등 여러 차트 컴포넌트가 있다. 이들 컴포넌트의 기능들을 비교, 분석하여 프로젝트에서 사용할 기능과 알맞은 기능을 가진 Nevron Chart for .NET 을 사용했다. Nevron Chart for .NET 은 .NET 을 지원하고 짧은 시간에 차트를 구성하여 보여준다. 유연성, 속도, 표현 능력에 있어서도 뛰어난 성능을 보인다.

3. BHAVE 시스템을 위한 분석 툴의 설계

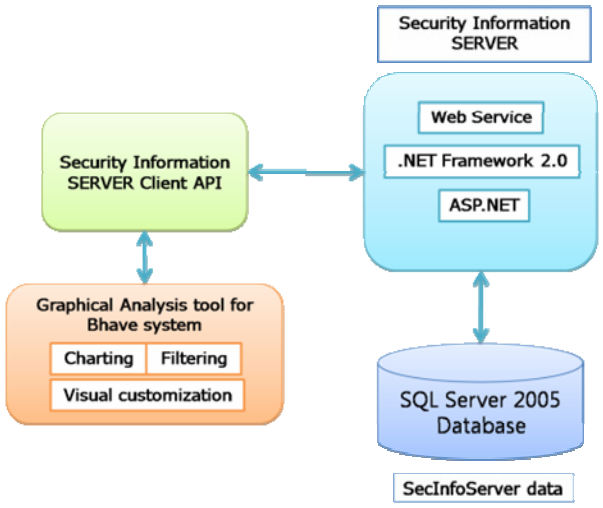
3.1 시스템 구조

BHave 시스템을 위한 그래픽 분석 도구는 효과적인 분석을 지원해주는 방법으로써 사용자의 행위를 그래픽한 화면에 매핑하여 보여주는 것을 가능하게 한다. 이 도구는 사용자의 행동에 관련한 데이터를 필터링 해주는 기능을 지원하며 X 축과 Y 축의 데이터를 바꾸거나 데이터 포인트의 레이블을 필요에 따라 보이거나 감추는 기능도 지원을 한다. 데이터가 많을 경우나 자세히 보고 싶은 데이터를 위해 확대하거나 축소기능을 추가하고, 스크롤이 가능하게 하였다.

그림 2 는 BHave 시스템의 전체적인 구조를 나타내는 것으로써 Security Information Server 라고 불리는 BHave 시스템의 중앙 서버는 사용자의 행동에 관한 데이터를 저장하는 데이터베이스를 가지고 있다. 이 데이터베이스는 SecInfoServer 로 총 14 개의 테이블을 가지고 있다. 이 데이터베이스의 스키마에 대해서는 뒤에서 다루도록 하겠다.

Security Information Server 는 서버 쪽의 웹 서비스 컴포넌트와 클라이언트 API 로 구성되어 있으며 클라이언트 API 는 BHave 시스템을 분석하기 위한 도구로 사용된다. 이 때 사용하는 데이터는 서버쪽의 SecInfoServer 에 있는 데이터를 가져와 사용한다. 클라이언트쪽의 API 는 IGI 측에서 제공하는 API 를 사용한다. 웹 서비스는 응용 프로그램의 작성 시 HTTP, XML, SOAP 와 같은 표준화된 웹 프로토콜과 데이터 형식을 사용함으로써 운영 체계(OS) 등 특정 플랫폼과 상관없이 모든 컴퓨터 간 원활한 데이터의 흐름을 보장해 준다.

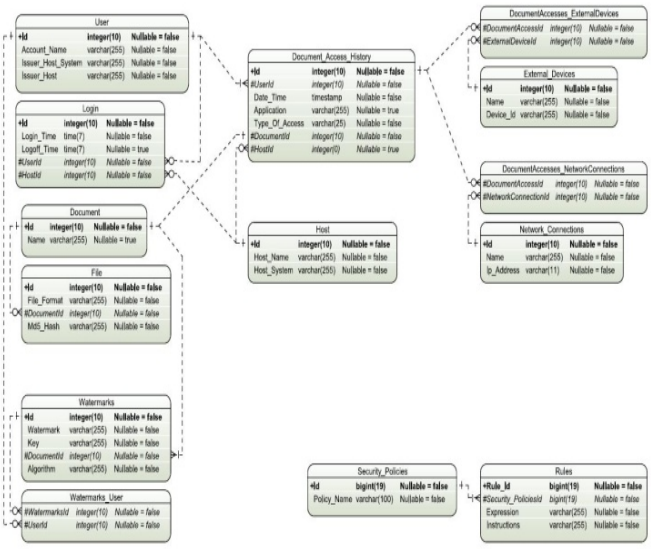
BHave 시스템을 위한 그래픽 분석 도구는 세가지 유형의 컴포넌트들로 모듈화 되어 있다. 첫 번째는 데이터를 제공해주는 것과 두 번째는 제공받은 데이터를 필터링하는 것, 세 번째는 필터링된 사용자 행동 데이터를 분석하여 화면에 그래프로 표시해 주는 것이다.



(그림 2. BHave 시스템 구조)

3.2 데이터베이스 스키마

BHave 시스템과 관련된 DB 를 SecInfoServer 라고 할 때, SecInfoServer 에는 User, Login, Document, File, Document_Access_History, DocumentAccess_ExternalDivices, DocumentAccess_NetworkConnection, Watermarks, Waternarks_User, Security_Policies, Rules, Host, External_Device 등 총 14 개의 테이블이 존재하며 각 테이블의 필드 및 관계들은 아래의 그림과 같다.



(그림 3. SecInfoServer Schema)

3.3 요구사항

● 구조

- 데이터베이스에 자동적으로 연결이 되어야 한다.
- 분석 도구는 stand alone 컴포넌트여야 하고 필요하다면 .NET 플랫폼에 기초한 미래의 플랫폼에 통합되어 질 수 있어야 한다
- 분석 도구는 상업적인 그래픽 컴포넌트를 사용해도 되지만 기능성을 더 확장한 API 를 제공하는 그래픽 컴포넌트를 사용해야 한다.

● 기본적인 그래프 기능

- scatter 그래프와 time interval 그래프를 제공해야 한다.
- 확대 및 축소 기능과 스크롤 기능을 제공해야 한다.
- 그래프에서 각 축은 데이터베이스에 저장된 데이터에 대하여 적절한 데이터 값을 보여주어야 한다. 축에 시간 데이터를 나타낼 때는 시간은 유저가 정의한 time interval 을 보여주어야 한다. 축에 시간 값은 일, 시간, 또는 분과 같은 유저가 정의한 time interval 의 정밀함을 나타낼 수 있어야 한다.
- 데이터 포인트 시리즈 또는 intervals 의 다른 종류는 다른 모양, 색깔, 크기 등등으로 표현되어야 한다.

● 데이터 필터링

- 사용자가 필터링을 할 수 있는 컴포넌트를 제공하고 사용자가 그래픽적인 분석 도구에서 분석을 위해 사용할 데이터만을 명시한다.
- 데이터를 필터링 할 때, 사용자, 호스트, 문서, 로그인 시간, 로그 오프 시간에 대한 값들은 모두 리스트 박스에 표시되어야 하고 이들은 모두 사용자가 선택할 수 있다.
- 날짜와 시간 필터링은 시작 시간과 종료시간의 두 가지 부분으로 나뉜다. 타임 픽커를 사용하여 사용자는 특정 시간을 선택할 수 있다.

● 시각화

- 사용자가 그래픽적인 포맷에 데이터들이 나타나는 방법을 제어할 수 있는 컴포넌트들을 제공한다.
- 그래프의 종류를 사용자가 선택할 수 있도록 제공한다.
- 각 축에 경계인 데이터를 유저가 선택하기 쉽게 한다. 사용자는 X, Y 축 각각을 표현하는 두 가지 데이터 필드를 선택하고, 데이터 포인트 레이블에 표현될 필드 또한 선택한다.
- Time interval 그래프에서는 X 축 값은 사용자가 선택한 TIME_OF_LOGIN_OF_THE_USER 와 TIME_OF_LOGOFF_OF_THE_USER 의 간격에 의해서 표현되고, Y 축 값은 ACCOUNT_NAME_OF_THE_USER 와 같은 값들에 의해서 표현된다. 데이터 포인트들은 사용자에 의해서 선택될 수 있고, HOST_TYPE, NAME_OF_THE_ HOST, NAME_OF_THE_DOCUMENT 과 같은 값들을 갖는다.
- 그래프의 눈금은 데이터의 간격과 값들을 알아보기 쉽게 하기 위해서 제공되어야 한다.
- 데이터 포인트 레이블을 표시하고 해제하는 체크박스가 제공되어야 한다.

4. BHAVE 시스템을 위한 분석 툴의 구현

4.1 구현환경

BHave 시스템을 위한 그래픽 분석 도구는 다음과 같은 환경에서 개발하였다.

운영체제는 마이크로소프트의 Windows XP Media Center 를 사용하였고, 사용한 프로그래밍 언어는 C# 으로 Microsoft Visual Studio 2005 에서 개발되었다. 그래프 컴포넌트는 앞에서 언급하였던 Nevro- n .NET Vision Q4 2006 for Visual Studio 2005 를 사용하였고, 데이터베이스는 MS SQL Server 2005 를 사용하였다.

.

4.2 구현결과

사용자 인터페이스는 차트를 선택할 수 있는 영역, 데이터를 필터링하는 영역, 차트를 그리는 영역, 그래프에서 포인트의 레이블을 보여주는지 결정하는 체크박스, 그리고 다시 그리기 버튼으로 구성된다. 차트를 선택하는 영역은 사용할 차트의 종류를 선택하는 부분이다. BHave 시스템은 그래픽적인 분석 도구를 위해 지원하는 두 가지 종류의 차트를 갖는다. 각각은 scatter 차트와 time interval 차트이다. 두 가지 차트를 선택하기 위해서는 각 차트 이름 앞의 라디오 버튼을 클릭하면 된다. 차트의 종류가 결정되면, X 축, Y 축의 데이터 타입과 데이터 포인트 레이블이 선택되어야 한다. 이 과정은 각 리스트 박스에 데이터 필드를 선택함으로써 이루어진다.

BHave 시스템을 위한 그래픽적인 분석 도구의 데이터 필터링 함수는 데이터를 필터링하는 영역에서 수행된다. 데이터 필드의 체크 박스에 체크함으로써 선택된 필드로 필터링할 수 있다. 각 리스트는 현재 데이터베이스에 저장된 데이터를 보여준다.

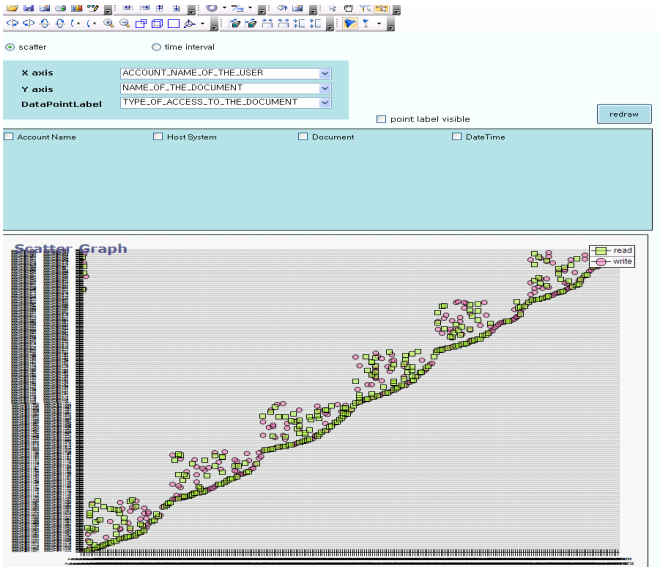
결과 차트는 차트를 그리는 영역에 그려진다. 차트의 결과를 나타내는 점은 앞의 과정에서 선택한 필터링 조건에 따라 데이터베이스로부터 읽은 결과에 대응한다. 차트는 확대/축소 아이콘을 사용하여 확대/축소가 가능하다. 또한, 차트는 차트를 움직이기 위한 아이콘을 사용하여 위치를 이동시킬 수 있다.

만약 포인트 레이블을 보여주는지 결정하는 체크박스가 체크되어 있다면, 포인트의 레이블들이 차트에 나타난다. 다시 그리기 버튼을 다른 차트 컴포넌트 영역에 대한 선택을 한 후 누르면, 차트를 다시 그리고 갱신하는 과정이 수행된다. 그리고 결과차트는 차트를 그리는 영역에 그려진다.

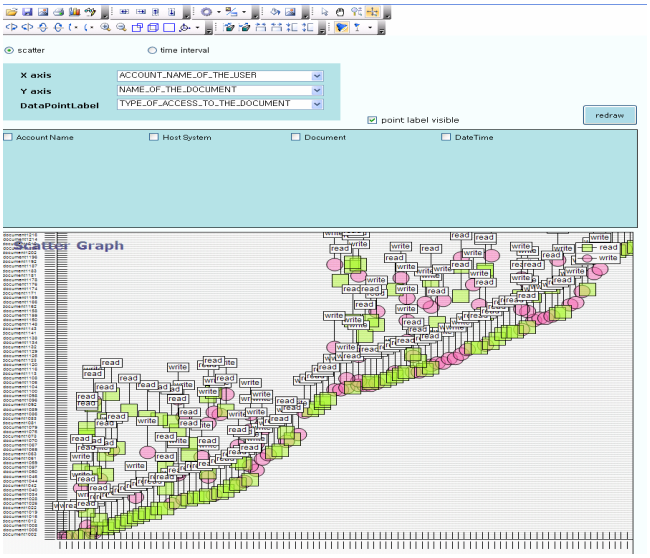
아래는 SecInfoServer 의 데이터를 이용하여 분석한 그래프 화면이다. 그림 4 에서 X 축에 표현할 데이터들은 User 테이블의 Account Name 필드에 해당하는 값인 ACCOUN-T_NAME_OF_THE_USER 이며, Y 축에 표현할 데이터는 Document 테이블의 Name 필드에 해당하는 값인 NAME_OF_THE_DOCUMENT 이고, 이때 각 포인트에 표현될 값은 사용자가 문서를 어떤 식으로 이용하였는지에 대한 내용을 나타내는 TYPE_OF_ACCESS_TO_THE_DOCUMENT 를 표현한다. 연두색은 READ, 분홍색은 WRITE 를 나타낸다.

그림 5 는 그림 4 의 화면을 확대하고 데이터 레이블을 표시하여 사용자에게 가독성을 높였다. 레이블의 표시를 해제하면 전체적인 데이터 분포를 파악하는데 용이하다.

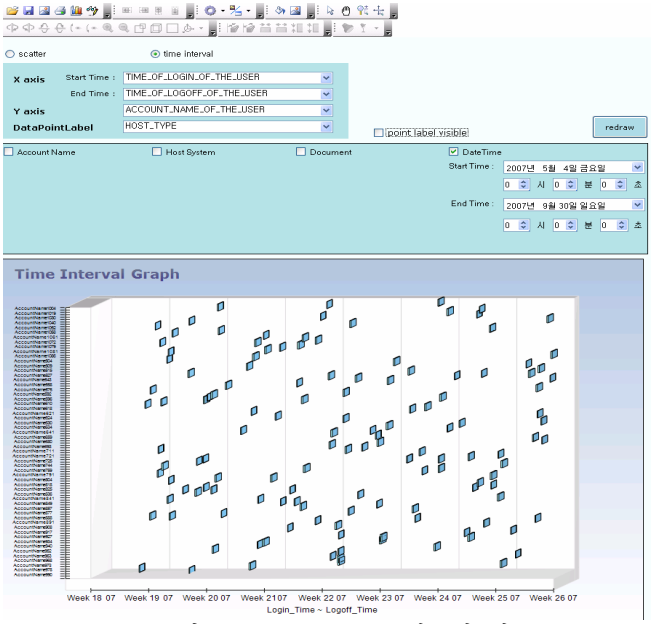
그림 6 은 Time interval 그래프로써 사용자의 로그인 시간부터 로그아웃 시간을 통하여 얼마나 오래 접속해 있었는지 알 수 있으며 특정시간에 접속한 사용자가 누구인지도 파악이 가능하다. X 축은 TIME_OF_LOGIN_OF_THE_USER 와 TIME_OF_LOGOFF_OF_THE_USER 필드를 이용해서 두 시간의 차를 이용하여 3D 막대그래프를 생성하였다.



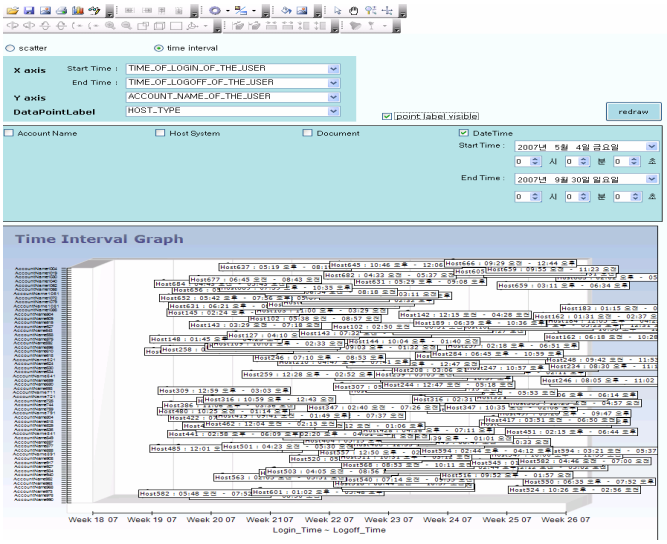
(그림 4. Scatter 그래프의 예)



(그림 5. Scatter 그래프 확대 및 데이터 레이블을 표시)



(그림 6. Time interval 그래프의 예)



(그림 7. Time interval 그래프의 데이터 레이블을 표시)

5. 결론

이 프로젝트의 목적은 사용자의 행동 정보를 수집하여 분석한 뒤 그래프를 사용하여 시각적으로 보여주는 소프트웨어 도구를 개발하는 것이다. Nevron 사의 컴포넌트를 사용하였고, BHave 데이터베이스와 차트 컴포넌트를 연동하여 추가적인 기능들을 구현하였다. 구현된 기능들은 데이터 분석시 사용자들에게 편의를 제공한다. 이 도구는 scatter 그래프와 time interval 그래프 두 종류의 그래프를 제공하고 있으며 scatter 그래프는 분산되어 있는 데이터를 분석하는데 유용하며 time interval 그래프는 사용자의 로그인 시간부터 로그아웃 시간을 파악하거나 그 시간 동안 어떤 일을 했는지 파악을 하는데 유용하다.

앞으로는 데이터를 집계하여 평균값이나 합, 최대값 최소값 등 현재는 지원하지 않는 기능들을 추가하는데 초점을 둘 것이며, 이 들 기능이 지원이 된다면 복잡한 시나리오에도 효과적인 분석을 할 수 있는 도구가 될 것이다.

※ 본 연구는 정보통신부 및 정보통신 연구진흥원의 IT 신성장동력 핵심기술 개발사업의 일환으로 수행했습니다. [2005-S004-02, 실감형 Virtual Engineering 기술]

참고문헌

[1] http://www.nevron.com/
[2] David S. Platt , "Introducing Microsoft .Net", MICROSOFT , 2003
[3] Lee, Wei Meng, Jepson, Brian , "Programming the .Net Compact Framework", Oreilly & Associates Inc, 2006
[4] http://www.softwarefx.co.kr/
[5] http://xceed.com/
[6] http://www.dundas.com/
[7] http://www.graphicsserver.com/
[8] http://www.datadynamics.com/
[9] http://www.steema.com/
[10] www.oracle.com/technology/documentation/disc-overer.html
[11] http:// www.hyperion.com
[12] Nick Symmonds , "GDI+ Programming in C# and VB.NET" , APRESS , 2002