동시 트랜잭션이 많은 데이터베이스에서 효과적인 개인정보보호 시스템 연구

강지원*

요 약

최근 개인정보보호법 제정에 따라 인터넷에서 개인정보를 취급하는 공공기관 및 기업들은 개인정보 보호를 위해 접근제어, 암호화 등 기술적 대책을 강구하고 있다. 개인정보 유출시 공공기관이나 기업은 이미지 훼손뿐만 아니라, 법적 책임을 면할 수 없는 실정이다. 그러나, 대용량 데이터베이스 시스템에서 매 접속 시마다 개인정보 항목에 대해 접근제와와 암·복호화를 하도록 하는 것은 성능 저하이 원인이 된다. 본 논문에서는 동시 트랜잭션이 많은 Oracle DBMS 환경에서 JVM(Java Virtual Machine)을 이용한 경량화 시스템을 설계·구현하여 성능에 최소한 영향을 미치면서 개인정보보호법에 정한 요구사항을 효율적이고 안전하게 수행하는 시스템을 제안하였다. 제안 시스템을 A 공공기관 포털 및 인사 시스템 내 개인정보 보호에 적용하여 성능 차이를 검증하였다.

A Study of Effective Privacy Protection System on High Concurrent Transaction Database System

Ji Won Kang*

ABSTRACT

Recently, according to the establishment of personal information protection Act, the public and private organizations are taking a step to protect personal information rights and interests by employing the technical methods such as the access control mechanism, cryptography, etc. The result of the personal information leakage causes a serious damage for the organization image and also has to face with the responsibility by law. However, applying access control and cryptographic approach on the personal information item for every connection to large database system causes significant performance degradation in a large database system. In this paper, we designed and implemented the light weight system using JVM (Java Virtual Machine) for the Oracle DBMS environment which the concurrent transaction occurs many, thereby the proposed system provides the minimum impact on the system performance and meets the need of personal information protection. The proposed system was validated on the personal information protection system which sits on a 'A' public organization's portal site and personnel information management system.

Key words : Privacy, DB Security, High Transaction, JVM model, Lightweight process

접수일(2012년 5월 8일), 수정일(1차: 2012년 5월 14일), 게재확정일(2012년 5월 21일)

* 경기대학교 정보보호학과

1. 서 론

정보화 사회로의 진입에 따라 개인정보의 개념이 인격권과 재산권이 혼재된 새로운 개념으로 이해되고 있고, 개인을 알아 볼 수 있는 정보가 유출되면 개인의 기본권인 인격권의 침해로 이어질 가능성이 높다[1].

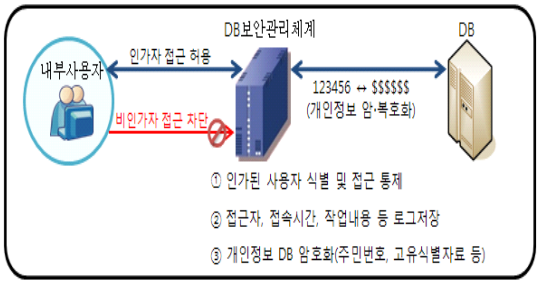
최근 국내의 A 온라인 쇼핑사이트와 S커뮤니티 사례에서 보듯이 개인정보 유출로 인한 기업이 겪는 이미지 실추와 피해 보상금은 막대하다. 공공기관 역시 2011년 9월 30일 부로 개인정보보호법이 시행되면서 각종 업무용으로 수집된 개인정보를 점검하고, 개인정보를 안전하게 관리하려는 노력을 경주하고 있다[2].

그러나 데이터베이스 시스템 내부에 저장된 개인정보들에 대해 접근제어와 암호화 등의 기술적 보호 조치는 필연적으로 시스템 성능 저하를 야기하게 되어 적용마저 쉽지 않은 실정이다. 특히, 특정 시간에 트랜잭션인 많은 데이터베이스 시스템에서 개인정보보호 메커니즘을 적용하는 것은 성능에 부담이 된다.

본 논문에서는 동시 트랜잭션이 많은 데이터베이스 시스템에 효과적이고 경량화된 개인정보보호 시스템을 제안하고자 한다.

2. 관련 연구

데이터베이스 개인정보 보호를 위한 보안 기술은 크게 접근제어(Access Control)와 암호화(Encryption) 기술로 구분할 수 있다.



(그림 1) DB 보안관리 절차

2.1 데이터베이스 접근제어

데이터베이스 접근제어의 기술은 크게 사용자나 응용 프로그램의 신분을 식별(Identification)하여 불법적인 사용자가 들어올 수 없도록 인증과, 권한에 따라 데이터베이스에 접근을 통제하는 접근제어, 그리고 시스템 기록된 로그를 바탕으로 각종 행위를 조사, 분석하는 감사 단계로 구분된다.

또한, 데이터베이스 접근제어 기술은 구성 방식에 따라 다음과 같이 구분한다.

o Agent 방식

Agent 방식이란, 서버 자체에 접근제어 및 로그 기능을 포함하는 Agent를 설치함으로써, 우회경로가 없는 강력한 보안 방법이며, Oracle DBMS의 Logical Session을 완벽히 처리하며, telnet, SSH 모두 지원한다. 그러나, 서버의 자원을 사용하기 때문에, 성능 저하 등의 단점이 있다.

o Gateway 방식

Gateway 방식은 데이터베이스 서버 앞단에, 보안서버를 설치하여, 보안 서버를 거쳐서 데이터베이스에 접근을 허용하는 방식이다. 이는 다시 프록시 방식과 인라인 방식으로 나눌 수 있는데 프록시 방식은 사용자가 접속을 데이터베이스에 직접 하는 것이 아니라, 보안서버로 접속을 하고, 보안서버에서 올바른 접근일 시 데이터베이스로 연결해 주는 방식이며, 인라인 방식은 네트워크 구성을 통해 데이터베이스 접근 시도 시, 보안서버를 거치도록 하는 방식이다.

o Sniffing 방식

Sniffing 방식이란, Tap이나, 패킷 미러링을 통해, 패킷을 분석하여 감사데이터에 남기는 방식으로써, 접근통제는 할 수 없지만, 감사 데이터를 남김으로써, 사후 감사의 역할을 하기 위한 방식이며, 다수의 데이터베이스 보안관리에 용이하고 DB서버 성능 지연에 영향이 적다[5].

2.2 데이터베이스 암호화

암호화란 의미를 알 수 없는 형식으로 정보를 변환하는 것으로 암호문의 형태로 정보를 기억 장치에 저장하거나 통신 회선을 통해 전송함으로써 정보를 보

호할 수 있다. 암호화는 암호 키를 사용하여 정보를 암호문으로 변환하는 것이고, 복호화는 복호화 키를 사용하여 원래의 정보로 복원하는 것이다[3][7]. 데이터베이스에서의 암호화란, 저장되어 있는 정보를 암호문으로 변환하여, 데이터베이스를 직접 획득하는 Stolen-verifier 공격[4]과 권한이 없는 사용자에게는 복호화 권한이 부여되지 않게 하여 원문을 해독 할 수 없게 만드는 방법이다.

암호화의 구성 방식에 따라 다음과 구분된다.

o Plug-in 방식

Plug-in 방식이란 데이터베이스 서버 내에서 암호화 수행 Agent 프로그램이 설치되어 데이터베이스 관리시스템과 연동하는 방식이다. 이는 어플리케이션의 수정이 없이 동작이 가능한 방식이며, 키 기밀성이 충족되는 강력한 방식이나, 일부 성능 저하가 발생한다.

o API 방식

API 방식이란 데이터베이스 서버와 무관하게 기존 어플리케이션에 암호화 수행 프로그램을 설치 후, 데이터베이스 관리 시스템과 연동하는 방식으로, 성능상 우수한 방식이다. 단점은 어플리케이션의 전체적인 수정이 발생한다.

o File 암호화 방식

File 단위 암호화는 운영체제 File 단위에서 암·복호화를 수행하므로 속도가 빠르고 데이터베이스 시스템에 미치는 영향이 거의 없는 방식이다. 하지만, 이 방식은 캐쉬(Cache) 영역에서 평문으로 유출될 우려 및 일방향 암호화 방식이 지원 되지 않는 단점이 있다.

3. 제안 시스템 설계 및 구현

3.1 요구사항 분석

3.1.1 접근제어 요구사항

o 법적 요구사항

<표 1> 정보통신망법 접근통제 요구사항

<정보통신망 이용촉진 및 정보보 등에 관한 법률>

제 4조(접근통제)

① 시스템 접근권한을 필요한 취급자에게만 부여.
② 취급자 변경 시 접근권한 변경 또는 말소
③ 권한 부여·변경·말소 내역 기록 및 보관(5년)
⑤ 불법적인 접근 및 침해사고 방지를 위해 다음 시스템을 설치·운영
- 시스템 접속 권한을 IP 주소 등으로 제한
- 시스템에 접속한 IP주소 등을 재분석하여 개인정보 유출 시도를 탐지

제 5조(접속기록 위 변조 방지)

① 개인정보 취급자가 접속한 기록을 월1회 이상 점검
② 접속기록 최소 2년 이상 보존·관리
③ 접속기록 위·변조되지 않도록 별도의 물리적인 저장장치에 보관 및 백업

제 9조(개인정보 표시 제한 보호조치)

개인정보 조회·출력 시 개인정보 마스킹 표시
① 성명 중 이름의 첫 번째 글자 이상
② 생년월일
③ 전화번호 또는 휴대폰 전화번호의 국번
④ 주소의 읍, 면, 동
⑤ 인터넷 주소는 v4의 경우 17~24비트 영역, v6의 경우 113~128 비트 영역

<표 2> 개인정보보호법 접근통제 요구사항

<공공기관 개인정보보호에 관한 법률>

제 29조(안전조치의무)

개인정보 분실·도난·유출·변조·훼손 방지를 위해 관리계획 수립, 접속기록 보관 등의 조치
(기술적, 관리적 및 물리적 조치 : 대통령령)

정보통신망법 제4조 ①,②,⑤항 및 정보통신망법 제9조. 개인정보보호법 제29조는 인증(login) 규칙을 이용하여, 데이터베이스에 접근 가능한 사용자를 식별하고, 로그인에 대한 권한을 부여하며, 접근통제 리스트에 의해서 세분화된 접근에 대한 권한을 부여 할 수

있다[6][1].

또한, 정보통신망법 제9조에 대해서는 접근통제 기술 중 부분컬럼 마스킹 기술을 이용하여 기능 구현을 할 수 있다. 정보통신망법 제4조 ③항의 내용은 시스템 인증과 감사 기능을 자체의 로그 저장과 보안정책 변화 등을 기록함으로써 기능 요구사항을 충족시킬 수 있다.

정보통신망법 제5조의 내용은 접근제어 기술 중 감사 기능을 이용하여 충족할 수 있다. 감사 규칙에 의하여 사용자가 데이터베이스에 접근해서 수행하는 행동을 감사기록으로 남기며, 해당 기록을 이용하여 접속 기록을 볼 수 있다.

o 시스템 요구사항

- Gateway/Sniffing/Agent 방식 선택적 조합
- DB 접근 사용자 인증 관리
- 사용자 권한 부여 및 접근제어
- DB 접속내역 실시간 모니터링
- DB 접속내역 감사추적
- 시스템 운영간 성능 저하 최소화
- 운영내역 분석 및 통계 제공

3.1.2 암호화 요구사항

o 법적 요구사항

<표 3> 정보통신망법 암호화 요구사항

<정보통신망 이용촉진 및 정보보 등에 관한 법률>
제 6조(개인정보의 암호화)
② 주민등록번호, 신용카드번호 및 계좌번호는 안전한 암호 알고리즘으로 암호화 저장

<표 4> 개인정보보호법 암호화 요구사항

<공공기관 개인정보보호에 관한 법률>
제 24조(고유식별정보의 처리 제한)
고유식별정보 처리 시 분실·도난·유출·변조·훼손되지 않도록 암호화 등 안전성 확보에 필요한 조치 (대통령령 지정)

정보통신망법 제6조와 개인정보보호법 제24조의 내용은 공공기관에 도입되는 암호화 시스템은 국가정보원 CC 인증을 받은 암호 알고리즘을 적용하여 개인정보보호를 구현한다.

여기서 중요한 것은 암호화 대상을 합리적으로 선정하는 것이다. 개인정보라는 것은 광의로 해석하면 개인을 식별하는 고유정보로써 많은 항목이 해당되므로 보호 가치와 유출시 파급효과, 시스템 성능 등을 고려하여 적절하게 선정하여야 한다. 즉 암호화 대상은 주민등록번호, 신용카드번호, 계좌번호, 여권번호, 외국인등록번호는 양방향 암호화를 하고, 비밀번호(Password)는 단방향 암호화를 한다. 또한 조합하여 개인정보가 될 수 있는 항목은 암호화가 아닌, 접근통제 규칙의 마스킹을 이용하여 처리함으로써, 데이터베이스의 성능 저하를 줄 일 수 있도록 설계하여야 한다.

o 시스템 요구사항

- 국정원 인증 암호화 알고리즘 사용
- 특정 항목 선택적 암호화 지원
- 접근제어와 통합 운영
- DB서버 암·복호화 성능 저하 최소화 (DB서버 성능 저하 10% 미만)
- 보안관리 서버 이중화
- 관리자에 의해 허용된 사용자만 복호화

3.2 시스템 설계

시스템 설계를 위해 가장 우선하는 것은 개인정보 항목을 식별하고 각 항목별 보호정책을 설정하고 적용 시스템 환경분석을 통해 현 운용 중인 데이터베이스 시스템 내에서 성능 문제를 검토해야 한다. 또한 앞서 언급한 기능 요구사항과 시스템 요구사항을 종합하여 개인정보보 호호 시스템 요구사항을 확정해야 한다.

그리고 개인정보보호 시스템 아키텍처를 설계하고 실제 운용중인 시스템에 적용이 가능한 프로토타입 시스템을 개발하는 것이다. 제안 시스템은 Java를 이용하여 쓰레드 단위로 처리하는 경량화 프로세스를 사용하고 동시 많은 트랜잭션 처리를 위해 Oracle Listener Pool을 Connection 효율성을 극대화 하도록 설계하였다.

o 적용 환경

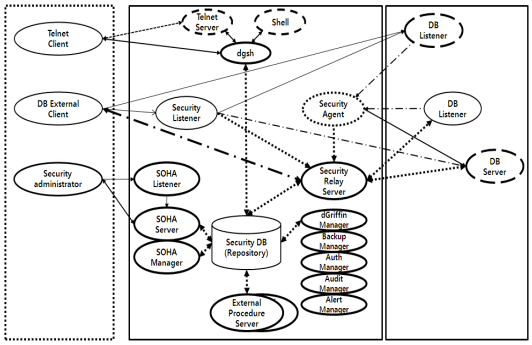
- DB 서버 : HP Super Dome
- 운영체제 : HPUX 11.31
- 개인정보 보유 시스템 : 15개 시스템
- DBMS : ORACLE 11.1 (RAC)
- 총 데이터 건수 : 100만건
- 초당 SQL 실행횟수 : 약 2,800개

o 시스템 구조 결정

- 접근제어 : 운영서버 환경에서 분석된 초당 SQL 실행횟수와 개인정보 데이터베이스 개수 등을 고려, Sniffing 방식 적용
- 암호화 : 어플리케이션 수정 없이 적용이 용이한 Plug-in 방식 적용

3.3 제안 시스템 구조

3.3.1 접근제어 시스템 구조



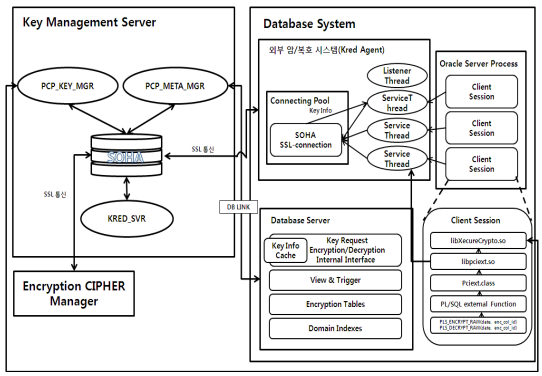
(그림 2) 접근제어 시스템 구조

(그림2)와 같이 제안 시스템은 Listener 프로세스(listener_server)를 이용해서, 클라이언트의 접속 요구를 받아 세션을 분할하고, 분할된 세션에 대한 통신 채널을 릴레이 프로세스(Relay Server)에 넘겨주는 역할을 한다. 릴레이 프로세스는 사용자와 데이터베이스 사이에 주고받는 메시지를 중계하면서 보안정책을 수행한다. 중계 하면서 생기는 감사 Log를 SOHA라는 전용 데이터베이스에 저장하게 된다. SOHA는 전용 레퍼지토리(Repository)로써 Cache로 구성된 메모리 데이터베이스이기 때문에 해당 데이터를 빠르고 최적화하여 동작한다.

3.3.2 암호화 시스템 구조

제안 시스템은 적용 대상인 데이터베이스 시스템에 (그림3)과 같이 KRED(Key Request Encryption Decryption) Agent가 설치되어 데이터베이스와 보안 시스템간 통신을 하며, 권한이 있는 사용자에게 Key를 발급해 주는 역할을 한다. 이 방식은 기존의 OEP(Oracle External Procedure) 방식보다 Connection Pool을 이용하기 때문에 대량의 트랜잭션 처리시 Key 발급 속도가 향상되어 전체적인 복호화 속도도 빨라지며 결국 시스템 성능이 향상되었다.

또한 제안 시스템은 Plug In View 형식으로써, 암호화 시 View를 제공해 권한이 있는 사용자에게는 위와 같이 View에서 자동으로 복호화 해서 보여 주기 때문에, 어플리케이션의 수정이 최소화 된다.



(그림 3) 암호화 시스템 구조

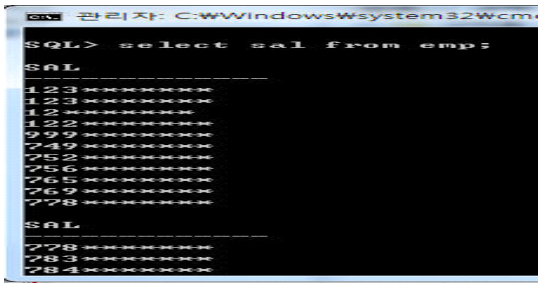
4. 성능 평가

4.1 접근제어 성능

제안 시스템은 접근제어 기능 적용시 현 운용환경에서 큰 성능 저하없이 초당 약 28,000여개의 SQL을 처리하였다.

또한, 개인정보 항목에 대해 부분 컬럼 마스킹(Masking) 기술을 통해 뭉쳐서 개인정보가 될 수 있는 항목 일부를 보이지 않게 처리함으로써 암호화 대상 컬럼을 줄일 수 있다. 특히 제안 시스템의 마스킹 기술의 장점은 다른 방식과 다르게 SQL을 분석하여 마스킹 하기 때문에, 결과 값의 패턴이 substr 등으로 변

경 되더라도, 마스킹을 처리할 수 있다.



(그림 4) 부분 컬럼 마스킹

4.2 암호화 성능

개인정보 컬럼이 데이터베이스 서버에 저장 시 국가행정망용 표준인 SEED 128비트 암호화 알고리즘에 의해 암호화되여 저장됨을 확인할 수 있다. 암호화에 따른 테이블 용량은 *암호문=(평문byte/16)*16+4* 공식에 따라 암호화 연산 시 한 블럭을 16byte 단위로 계산하기 때문에 위의 공식에 따르면 16byte 미만은 20byte로, 32byte 미만은 36byte로, 48byte는 52byte로 약간씩 컬럼 사이즈가 증가한다.

또한, 'A' 공공기관에 운용중인 2개의 대용량 시스템에서 자주 사용되는 Select 및 Update SQL문 12개(full scan 1개 포함)를 선정하여 암호화 테이블 3개, 비암호화 테이블 4개로 하고 각 1,000번씩 25회 수행하여 평균 처리시간을 실험하였다.

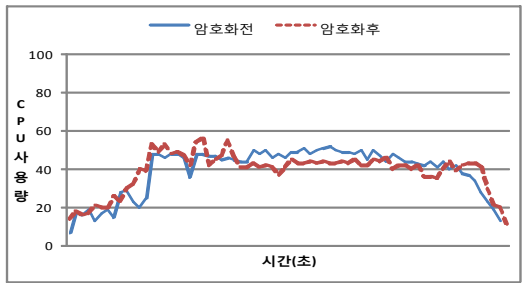
<표5>에서 보는 바와 같이 test SQL문을 1000번 수행할 경우, 적용 시스템 모두에서 암·복호화 처리로 인해 암호화 적용 전보다 처리시간이 평균 약 1.6초 정도 지연된다. 퍼센트로 환산시 성능은 암호화 전에 비해 약 4.5% 떨어지지만, 안전한 개인정보 관리를 위한 Trade-Off로서, 최초 시스템 요구사항인 10% 이내의 성능 저하 조건을 만족하였다.

<표 5> 암호화 적용 전·후 처리시간 비교

[단위 : 초]

	암호화 前	암호화 後	차 이
시스템#1	34.56	36.21	+1.65
시스템#2	35.45	37.02	+1.57

그리고 시스템 성능감시 툴로 확인한 1일 암호화 적용 전·후 CPU 사용량 그래프는 (그림5)와 같이 약 3~5% 성능 저하를 보인다.



(그림 5) DB서버 1일 CPU 성능 비교

6. 결 론

개인정보보호법 전면 시행에 따라 데이터베이스 개인정보 보호를 위한 요구사항을 식별하고 접근제어 방식과 암호화 방식을 하이브리드로 적용하였다. 앞서 살펴 본 바와 같이 메모리 데이터베이스를 저장소로 활용하고, JVM을 이용한 경량화 프로세스 기반으로 설계하였다. 또한, 국가행정망 표준인 128bit SEED 암호화 알고리즘을 적용하여 안전성을 확보하였고, Connection Pool을 이용하여 복호화 키 관리 프로세스 효율성을 극대화하여 기존 시스템 성능에 영향을 최소화 하도록 개인정보 보호 시스템을 제안하였다.

향후 대용량 트랜잭션 데이터베이스 시스템에서 개인정보 보호 시스템 적용을 위해 기존 운용 시스템의 특성과 연계하여 적용기술을 판단하고, 데이터베이스 보안관리 서버 장애에 대비하여 이중화도 필요하다. 또한 상용 DBMS 특성에 최적화된 시스템 구조를 갖도록 상용제품 제조사와 협력하여 실제 적용이 가능하고 경량화 시스템으로 발전시킬 필요가 있다.

참고문헌

[1] 공공기관의 개인정보보호에 관한 법률, 2011.9.30
[2] 데이터베이스 보안 가이드라인, 2011, KDB
[3] 네이버 백과사전
[4] C.M Chenand W.C.Ku, "Stolen-Verifier Attack on Two New Strong-Paaword Authentication Protocols," IEICE Trans. on Communications, Vol. E85-B, No 11, pp. 2519-2521, 2002
[5] 2011 DB백서, KDB, 2011.8
[6] 정보통신망 이용촉진 및 정보보호 등에 관한 법률, 2011.8
[7] 손태경, "안전한 개인정보 보호방안 연구", 2011.2, 숭실대

[저 자 소 개]

강 지 원 (Ji-Won Kang)

1988년 금오공과대학교전자공학과 (공학사)
1997년 연세대학교컴퓨터과학과 (공학석사)
2006년 경기대학교정보보호학과 (박사과정)

email: kang0158@hanmail.net