Secure Distributed Databases Using Cryptography

Prof.dr. Ion IVAN, asist. Cristian Valeriu TOMA
Catedra de Informatică Economică, A.S.E. Bucureşti

The computational encryption is used intensively by different databases management systems for ensuring privacy and integrity of information that are physically stored in files. Also, the information is sent over network and is replicated on different distributed systems. It is proved that a satisfying level of security is achieved if the rows and columns of tables are encrypted independently of table or computer that sustains the data. Also, it is very important that the SQL - Structured Query Language query requests and responses to be encrypted over the network connection between the client and databases server. All this techniques and methods must be implemented by the databases administrators, designer and developers in a consistent security policy.
Keywords*: databases, encryption, cryptography, digital signature.*

1 Introducere

În securitatea sistemelor distribuite în general şi în securitatea bazelor de date în particular se urmăreşte securizarea mesajelor şi a tranzacţiilor ce se efectuează în sistem. Principalele obiective urmărite în securizarea unui sistem distribuit sunt:

1.*Confidenţialitatea* – protejează conţinutul mesajelor (tranzacţiilor) împotriva citirii neautorizate, de către alte persoane decât receptorii specificaţi de emiţător.

2.*Autentificarea* – permite receptorului unui mesaj să determine în mod sigur identitatea expeditorului.

3.*Integritatea datelor* – în reţea furnizează receptorului unei tranzacţii siguranţa că mesajul primit este identic cu mesajul emis de expeditor.

4.Prevenirea nerecunoaşterii mesajului (tranzacţiei) de către expeditor (*Non-Repudierea*) – garantează integritatea şi originea mesajelor (tranzacţiilor) din punctul de vedere al expeditorului şi nu al destinatarului. Se împiedică astfel ca expeditorul unei tranzacţii electronice să nege trimiterea ei.

5.*Aplicarea selectivă a unor servicii* – de multe ori este necesară acoperirea unor părţi ale mesajelor (tranzacţiilor), de exemplu cele conţinând numărul cărţii de credit al unui client. Aceasta nu trebuie să fie în clar vânzătorului, care poate abuza de utilizarea ei.

Criptografia este o "unealtă" folositoare în realizarea securităţii sistemului (nu este singura, se adoptă pachete de măsuri grupate în politici de securitate pentru a asigura securitatea sistemului).

Există două tipuri de sisteme criptografice: simetrice şi asimetrice. *Sistemele criptografice simetrice* (cu cheie secretă) folosesc aceeaşi cheie, atât la criptarea cât şi la decriptarea mesajelor. *Sistemele criptografice asimetrice* (cu cheie publică) folosesc chei distincte la criptare şi decriptare (dar legate una de alta). Una din chei este ţinută secretă şi este cunoscută doar de proprietarul ei. A doua cheie (perechea ei) este făcută publică. Algoritmi criptografici (cifrurile) folosiţi în sisteme criptografice simetrice se împart în cifruri flux (stream ciphers) şi cifruri bloc (block ciphers). Cifrurile flux pot cripta un singur bit de text clar la un moment dat, pe când cifrurile bloc criptează mai mulţi biţi (64 sau 128 de biţi) la un moment dat.

2. Sisteme criptografice şi modalităţi de criptare

În acest articol se operează o serie de concepte din criptografia computaţională cum ar fi: funcţiile hash, algoritmi criptografici simetrici, asimetrici şi modalităţi de criptare.

2.1. Funcţii hash

Funcţiile hash (de dispersie) joacă un rol important în autentificarea conţinutului unui mesaj transmis în comunicaţii. Obiectivul lor NU este de a asigura secretul transmisiilor, ci de a

crea o valoare h=H(m), numita rezumat (digest), cu rol în procedura de semnătură digitală, valoare h foarte greu de falsificat.
În procedura de semnare digitală sunt implicate 3 entități:

- M = mesajul de "digerat";
- h = H(M) rezumatul calculat prin hash;
- S = Sign(H(M)) semnătura digitală.

Funcțiile hash au câteva caracteristici comune:

- fiind dat M, este simplu să se calculeze h;
- fiind dat h, este IMPOSIBIL să se calculeze M, astfel încât H(M) = h;
- fiind dat M, este greu să se găsească alt mesaj M' (imposibil chiar), astfel încât H(M) = H(M');
- este imposibil să se găsească 2 mesaje aleatoare, astfel încât H(M) = H(M'), proprietate numită rezistență la coliziune.

Una din cerințele fundamentale pentru o astfel de funcție este ca, modificând un singur bit la intrare să producă o avalanșă de modificări în biții de la ieşire. Unele din cele mai utilizate funcții de dispersie sunt: **SHA-1, MD5**, MD2, RIMPED-160.

2.2. Algoritmi criptografici cu cheie simetrică

În cazul sistemelor criptografice simetrice (cu cheie secretă) se foloseşte aceeaşi cheie, atât la cifrarea cât şi descifrarea mesajelor. Cheia este ținută secretă şi folosită în comun de către emițător şi receptor (figura 1). Sistemele simetrice sunt bine cunoscute, conduc la performanțe bune şi sunt folosite pentru protecția datelor utilizatorilor. La baza cifrurilor simetrice stau cifrurile elementare bazate pe: transpoziție şi substituție.

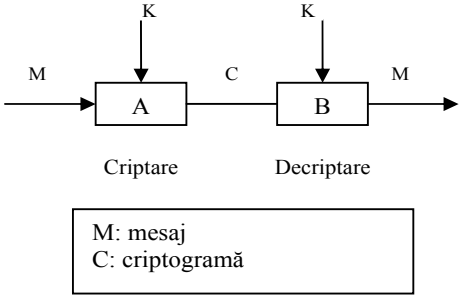

Fig.1. Cifrare simetrică.

Următorii algoritmi din „era 128 de biți" (şi cheia şi mesajul au minim 128 de biți) sunt cei mai cunoscuți: Rijndael (finalist şi câştigător AES), Twofish (finalist AES), Serpent (finalist), RC6 – Rivest Cipher 6 (finalist), MARS (finalist), Blowfish, DEAL, SAFER+, FROG, CAST-256, Magenta, SkipJack, DFC – Decorrelated Fast Cipher.

2.3. Algoritmi criptografici cu cheie asimetrică

Cifrurile asimetrice (cu chei publice) folosesc chei distincte de cifrare şi descifrare (dar dependente una de alta). Deoarece este imposibil deducerea unei chei din cealaltă, una din chei este făcută publică, fiind pusă la dispoziția oricui doreşte să trimită un mesaj cifrat. Doar destinatarul, care deține cea de-a doua cheie, poate descifra şi utiliza mesajul. Tehnica cheilor publice este folosită şi în semnătura digitală (electronică), fapt ce a sporit popularitatea sistemelor criptografice cu cheii publice. Modul de folosire a unui cripto-sistem asimetric este redat în figura 2.
În figura 2 sunt prezentate trei utilizări ale acestor tipuri de criptosisteme asimetrice. Exemplul 2 şi 3 din figură detaliază procedura de semnătură digitală deoarece aceste exemple arată modul cum se execută o semnătură digitală asupra unui mesaj M. Exemplul 3 ilustrează procesul de semnătură electronică plus procesul confidențialitate.

2.4. Moduri de criptare

La nivelul modului de utilizare ai algoritmilor simetrici (indiferent de algoritmul folosit, iar uneori chiar şi asimetrici) se întâlnesc în practică două tipuri de cifrări: cifrarea bloc (block ciphering) şi cifrarea flux (secvențială) (stream ciphering).
Cifrarea bloc operează cu blocuri de date în clar şi cifrate – de regulă 64 şi 128 biți dar, uneori, şi mai mari. Cele mai cunoscute moduri de acest tip sunt: ECB, CBC, PCBC, OFBNLF.
Cifrarea secvențială operează cu secvențe de date în clar şi cifrate de mărime un bit sau octet dar, uneori şi cu date de 32 de biți. Cele mai cunoscute moduri de acest tip sunt: cifrarea secvențială, cifrarea secvențială cu autosincronizare, cifrarea cu reacție, cifrarea sec-

venţială sincronă, cifrarea secvenţială cu reacţie la ieşire, cifrarea cu contor.

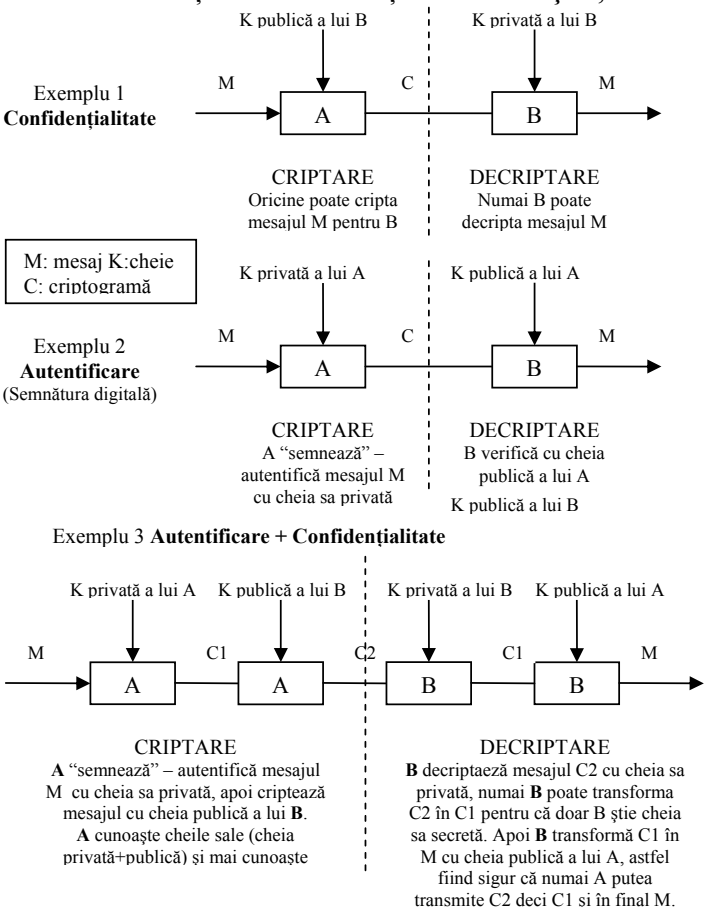

Fig.2. Sistemul de cifrare cu cheie publică folosit pentru confidenţialitatea sau/şi autentificare unui mesaj M între expeditorul A şi destinatarul B.

3. Dimensiunile potenţial securizabile ale bazelor de date

Există trei mari dimensiuni ce adeseori sunt analizate în ce priveşte criptarea în baze de date:

• Dimensiunea granularităţii datelor ce sunt criptate şi decriptate. Câmpul, atributul sau tuplul sunt câteva alternative în ce priveşte criptarea datelor. In practică s-a dovedit că ce mai bună alegere este să fie criptat câmpul pentru că minimizează numărul de octeţi ce trebuie criptaţi şi decriptaţi.

• Dimensiunea implementării algoritmilor criptografici la nivel hardware versus software. Cele mai bune rezultate dar şi cu cele mai mari investiţii iniţiale sunt cele furnizate de sisteme hibrid, apoi se consideră criptarea/decriptarea hardware şi apoi cea software. Fiecare model, hibrid, hardware sau software are avantajele/dezavantajele lui, precum şi costuri şi performanţe operaţionale diferite.

• Dimensiunea locaţiei serviciului de criptare. Serviciul de criptare este fie local, fie apelabil prin proceduri la distanţă (RPC – Remote Procedure Call), fie serviciu ataşat reţelei.

În mod practic se observă că dacă criptarea respectiv decriptarea
are loc cât mai aproape de mediul unde sunt stocate datele, atunci gradul de securizare prin criptare creşte. Criptarea executată de sistemele de gestiune a bazelor de date protejează datele ce aparţin tabelelor din bazele de date existente şi este o alegere a administratorilor, utilizatorilor şi a dezvoltatorilor de aplicaţii de baze de date dacă trebuie criptată/decriptată şi comunicaţia între baze de date şi aplicaţii ce utilizează acele date.

4. Alegerea topologiei pentru implementarea mecanismelor criptografice

Se iau în considerare câteva combinaţii de criptare precum cea la nivel software şi har-

dware şi cele legate de granularitatea datelor. In practică analizele au fost întâi efectuate la nivel de câmp/atribut. S-au dezvoltat acceleratori de căutare peste câmpuri criptate, incluzând câmpurile cheie primară. La nivel hardware se preferă criptarea doar a cheilor de administrare şi a stocării.

*Criptare/decriptare la nivel elementar software***.** In testele efectuate pe bazele de date relaţionale distribuite sau nu, s-au luat în considerare algoritmi criptografici cu cheie simetrică precum AES – Rijndael sau Blowfish şi algoritmi cu cheie asimetrică precum RSA sau DSA. Testele au arătat ceea ce era de aşteptat că algoritmi cu cheie simetrică sunt mult mai rapizi decât cei cu cheie asimetrică. Procedura este destul de simplă, se stochează algoritmul AES sau RSA ca fiind funcţie definită de utilizator (stored procedure). Odată înregistrată această funcţie poate fi apelată de un declanşator (trigger) ori de câte ori apare o comandă SQL de inserare date (INSERT) sau interogare date (SELECT).

Criptare/decriptare la nivel hardware. Unul din cele mai utilizate sisteme de criptare hardware este HSM FIPS-140-1. Cheia principală este creată şi criptată de sistemul HSM. Ca şi algoritm principal de criptare se utilizează sistemul criptografic cu cheie simetrică DES. Se preferă criptarea/decriptarea software pentru ca are cel mai bun raport calitate preţ. Cel mai utilizat mod de criptare a datelor şi structurilor de date în SGBD-uri este CBC.

Cifrarea bloc cu înlănţuire CBC (Cipher Block Chaining) - Acest mod de cifrare bloc, adaugă mecanismului de criptare un bloc cu reacţie. Rezultatul criptării unui bloc anterior revine prin buclă şi intervine în criptarea blocului curent.

În acest mod, datele cifrate nu mai depind doar de datele în clar, ci şi de modul de cifrare a blocului anterior. În CBC, datele în clar, înainte de a intra în blocul decriptare propriuzis, sunt însumate modulo 2 (XOR) cu blocul de date cifrat anterior. Figura 3 reprezintă modul de criptare CBC.

Paşii pe scurt sunt următorii: se iniţializează registrul de reacţie cu o funcţie hash de dispersie MD5, ce produce rezumatul unei parole. Apoi pentru i (un contor) de la 0 la numărul de blocuri a fişierului sau structurii de date se execută XOR (sau exclusiv) între blocul citit din fişier şi blocul de date din registrul de reacţie. Se scrie blocul criptat în fişier şi apoi se atribuie registrului de reacţie blocul de biţi criptat. Se incrementează i şi apoi se reia procesul. Aceasta e o structură de date compusă ce implică două structuri de date de tip fişier şi o structură de tip masiv sau după necesităţi poate fi una dinamică (listă de liste de octeţi).

CBC face ca acelaşi bloc de date să se transforme în blocuri de date diferite, deoarece la diferite rulări valoarea de iniţializare a registrului de reacţie poate fi diferită. Dacă valoarea iniţială a registrului de reacţie rămâne neschimbată între rulări, atunci două mesaje identice folosind aceeaşi cheie se vor transforma în acelaşi mesaj criptat.

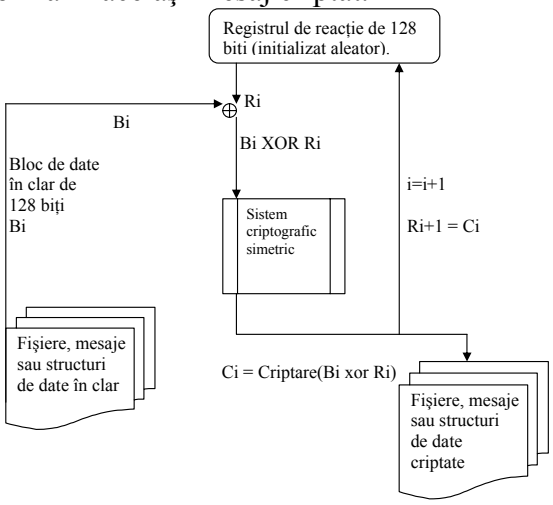

Fig.3. Cifrare CBC.

Vectorul de iniţializare (valoare iniţială a registrului de reacţie) nu trebuie neapărat să fie secret (deşi se poate genera printr-o funcţie hash în urma unei parole, astfel încât să nu fie nevoie de transmiterea – valorii iniţiale – prin reţea la receptor).

Chiar dacă acest lucru pare greşit (de a nu ţine secret valoarea iniţială), nu este deoarece, oricum prin canal (reţea) circulă blocurile criptate dar nu şi cheia deci, cineva ce ar dori să spargă cifrul va trebui să cunoască ce structură de date s-a folosit, ce algoritmi şi protocolul de transmisie a datelor, şi apoi să spargă algoritmul.

5. Concluzii

În acest articol au fost elaborate şi analizate o serie de concepte prezente în securizarea bazelor de date prin utilizarea criptografiei computaționale. Acest domeniu este o mică parte din ce înseamnă securitatea la nivelul bazelor de date unde se aplică politici de securitate consistente la nivel aplicație, utilizator şi administrator de baze de date.

S-a observat că în practică datorită comodității şi lipsei de interes a unor dezvoltatori şi administratori de baze de date şi de aplicații ce utilizează baze de date, apar breşe de securitate ce sunt uneori vulnerabilități greu de acceptat în contextul unei cerințe cât mai pronunțate de produse program şi servicii cu grad ridicat de valoare adăugată şi de securitate. Modul prin care poate fi combătut acest lucru este conştientizarea pericolelor ce se pot produce la nivel informații precum şi comunicarea şi dezbaterea la nivelul comunității dezvoltatorilor şi utilizatorilor a acestor probleme.

Bibliografie

[1] J. He and M. Wang. Encryption in relational database management systems. In Proc. Fourteenth Annual IFIP WG 11.3 Working Conference on Database Security (DBSec'00), Schoorl, The Netherlands, 2000.

[2] R.Agrawal,J.Kiernan,R.Srikant,and Y.Xu. An XPath-based preference language for P3P. In Proc. of the 12th Int'l World Wide Web Conference, Budapest, Hungary, May 2003.

[3] S. Muchnick. Advanced Compiler Design and Implementation. Morgan Kaufmann Publishers, 1997.

[4] N. R. Adam and J. C. Wortman. Security-control methods for statistical databases. ACMComputing Surveys, 21(4):515– 556, Dec. 1989.

[5] R. Agrawal and J. Kiernan. Watermarking relational databases. In 28th Int'l Conference on Very Large Databases, Hong Kong, China, August 2002.

[6] R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic databases. In Proc. of the 28th Int'l Conference on Very Large Databases, Hong Kong, China, August 2002.

[7] Agrawal, J.Kiernan,R.Srikant,andY.Xu.Implementing P3P using database technology. In Proc. of the 19th Int'l Conference on Data Engineering, Bangalore, India, March 2003.