Encripción de Bases de Datos

Capetillo R. María del Rocío , Gómez C. Roberto
ITESM-CEM
rociocapetillo@hotmail.com, rogomez@campus.cem.itesm.mx

Actualmente existen diferentes métodos para garantizar la seguridad de una base de datos, pero ninguno por si solo nos puede garantizar completamente tanto la confidencialidad como la integridad de la base de datos. Es necesario combinar varios mecanismos para lograr la seguridad deseada. Algunos de estos mecanismos son la seguridad física, los sistemas de autentificación y la encripción. El presente trabajo se sitúa dentro del contexto de este último mecanismo: la encripción.

La seguridad de las bases de datos debe considerar que ninguna persona o grupo de personas deben ser capaces de leer, escribir, destruir o modificar datos directamente de una manera no autorizada. Así mismo, debe ser imposible para cualquier persona o grupo de personas inferir el valor de cualquier objeto dato por manipulación directa o computacional. Los mecanismos de seguridad no deben degradar el desempeño de las operaciones básicas de la base de datos y no debe producirse un aumento grande en el almacenamiento de datos. Sin olvidar que el costo computacional debe ser razonable.

El algoritmo de encripción a usar en una base de datos, debe ser teóricamente seguro o requerir un trabajo extremadamente alto para romperlo. La encripción y decripción deben ser suficientemente rápidos para no degradar el desempeño del sistema, por ejemplo la decripción debe ser rápida ya que será utilizada con mas frecuencia al hacer las consultas a las bases de datos. Además los datos encriptados no deben ocupar más espacio que los datos decriptados.

La encripción de datos debe ser orientada a registros. Debe ser posible la encripción y decripción sobre un solo registro sin considerar una posición física o lógica en la base de datos. Esta propiedad es requerida para que, en el caso de que un registro cambie de posición, no se tenga que decriptar n-1 registros para decriptar el registro n.

La encripción debe soportar la lógica del enfoque del subesquema para bases de datos. En un sistema de bases de datos, el DBMS (sistema manejador de bases de datos) es la primera línea de defensa de nuestra información. El DBMS presenta a un usuario solo los campos que ha solicitado y, sobre todo, solo a los que tiene acceso. El sistema de encripción debe impedir que si un atacante obtiene partes del registro ilegítimamente, este pueda usarlos. Es necesario que un registro encriptado no sea un conjunto de campos encriptados individualmente. Un registro encriptado debe ser un valor único encriptado en función de todos los campos. Esto previene dos cosas: la comparación de patrones y la substitución de valores encriptados. El esquema de encripción debe ser flexible, estimando cualquier combinación de privilegios. Idealmente debe ser posible otorgar a cada usuario privilegios por cada campo, así como poder detectar y rechazar registros que hayan sido creados o modificados bajo una falsa llave de encripción.

El DBMS no debe ser forzado por el sistema de encripción a guardar copias duplicadas de objetos de datos. Por último, los registros probablemente serán construidos sobre un periodo de tiempo, por lo que debe ser posible recuperar información de registros parcialmente completos.

Un primer enfoque propone un esquema de encripción orientado a registro, en el cual un registro encriptado es una función individual de los valores de todos los campos, y en los cuales cada campo separado es encriptado/decriptado por llaves separadas.

En este esquema las llaves se consideran subllaves en el sentido que, mientras la encripción de registros es una función de todas las llaves de encripción y los valores de los campos, cada llave de decripción solo decripta un subconjunto de campos del registro original. Esto puede ser descrito matemáticamente como

$$Ci = E((e1,f1i),(e2,f2i),...,(en,fni))$$
$$Fij=D(di,Ci),$$

donde fij es el valor del campo j y el registro i; ej, la llave de encripción para el campo j; dj, la llave de decripción por cada campo j; Ci, la versión encriptada del registro i; E representa el algoritmo de encripción; y D es el algoritmo de decripción. Se asume que E,D y Ci son conocidos por todos, y que tanto ej como dj solo son conocidos por usuarios que tienen los privilegios correspondientes (el DBMS no cuenta con estos). Fji son los datos a proteger. Además, ej y dj solo son conocidos por los usuarios que tienen privilegios sobre esos campos.

Nuestra propuesta es una variante del esquema anterior. En primer lugar nos enfocamos al modelo relacional, ya que es él mas utilizado en el modelado de bases de datos. El esquema de encripción de subllaves se cambió por uno de encripción simétrica, utilizando el algoritmo AES. Por otro lado, el esquema de asignación de privilegios no se basa en la posesión de un conjunto de llaves sino en un modelo de autorización de privilegios propio de las bases de datos relacionales.

Nosotros proponemos encriptar el registro como un todo, sin separaciones entre campo y campo. Con una sola llave de encripción/decripción que será elegida por el propietario o administrador de la base de datos. A diferencia del esquema anterior este algoritmo es más sencillo en cuanto a su administración y diseño por que solo existe una llave, además de ser más seguro y rápido.

En este momento nos encontramos en la implementación y prueba del esquema propuesto para la encripción de bases de datos. No se implementó un manejador de bases de datos completamente, más bien se modificó uno existente, MySQL. Elegimos este manejador por que su sistema de autentificación se acopla al que proponemos y sobre todo por que es código abierto. Las pruebas se han estado realizando con archivos de datos de varios tamaños, pero sobre todo con tamaños muy grandes para probar la eficiencia en base a la velocidad y tamaño de archivos ya encriptados contra los archivos sin encriptar.