Trabajo Práctico 8

Privacidad y Seguridad

- **Alumno**: Sussini Guanziroli, Patricio
- **Materia**: Bases de datos I
- **Tutora**: Constanza Uño

Pregunta 1: Cifrado Homomórfico

- ¿Qué es el Cifrado Homomórfico?

El Cifrado Homomórfico es un tipo de cifrado revolucionario que permite realizar cálculos y operaciones matemáticas directamente sobre los datos cifrados, con la particularidad de que no precisa descifrarlos primero. Al final, cuando descifras el resultado de la operación, obtienes exactamente el mismo resultado que si hubieras hecho los cálculos sobre los datos originales en texto plano.

- ¿Cuál es el principio fundamental y sus diferencias con otras técnicas de cifrado más tradicionales en términos de su aplicación a los datos en uso?

El **principio fundamental** se basa en la propiedad matemática de la “homomorfia”, que permite que una operación sobre los datos cifrados sea equivalente a la misma operación sobre los datos originales.

Diferencias Clave:

- **Cifrado Tradicional:** Protege los datos en dos estados.
    - En **reposo:** Cuando están guardados en un disco duro o una base de datos.
    - En **tránsito:** Cuando viajan por una red.
    - **Punto débil:** Para poder procesar o hacer cálculos con estos datos, obligatoriamente hay que descifrarlos en la memoria del servidor. En ese instante, los datos quedan expuestos y vulnerables.
- **Cifrado Homomórfico:** Protege los datos en tres estados.
    - **En uso:** Protege los datos mientras están siendo procesados activamente en la memoria del servidor. Elimina por completo la necesidad de descifrar los datos para trabajar con ellos, cerrando esa ventana de vulnerabilidad.

- Identificación de al menos dos ventajas y dos desafíos o limitaciones.

Ventajas Clave:

1. **Privacidad Total en la Nube:** Permite a las empresas delegar el procesamiento de datos sensibles a proveedores de servicios en la nube con la garantía de que el proveedor **nunca podrá ver información real.** Ej: Un hospital podría enviar datos de pacientes cifrados para que se realicen análisis estadísticos, y ni siquiera los ingenieros de la nube tendrían acceso a los datos médicos.
2. **Colaboración Segura:** Facilita la colaboración entre diferentes organizaciones sin que tengan que compartir sus datos confidenciales. Por ejemplo, varios bancos podrían juntar sus datos sobre transacciones fraudulentas (cifrados) para entrenar un modelo de IA que detecte fraudes, sin que ningún banco vea los datos de sus competidores.

Desafíos o Limitaciones:

1. **Sobrecarga Computacional:** Por lejos el mayor desafío actual. Las operaciones matemáticas sobre datos homomórficamente cifrados son muchísimo más lentas y consumen muchos más recursos como CPU y memoria que las mismas operaciones sobre texto plano.
2. **“Ruido” en los Datos Cifrados:** Cada operación que se realiza sobre los datos cifrados añade una pequeña cantidad de “ruido” matemático. Después de un cierto número de operaciones, este ruido se acumula tanto que corrompe el resultado, haciéndolo imposible de descifrar correctamente. Aunque existen maneras de limpiar ese ruido, son extremadamente costosas y lentas.

- Ejemplos concretos de aplicaciones potenciales

Algunos ejemplos concretos de aplicaciones potenciales para este tipo de cifrado:

1. **Análisis de Datos Médicos y Genéticos:** Permite a los investigadores realizar estudios sobre grandes conjuntos de datos de pacientes de diferentes hospitales sin comprometer su privacidad. Se podrían buscar enfermedades en materiales genéticos cifrados.
2. **Servicios Financieros y Detección de Fraude:** Los bancos podrían analizar datos de transacciones cifradas de sus clientes para detectar patrones de fraude u ofrecer servicios financieros personalizados, sin acceder a los detalles directamente.
3. **Votación Electrónica Segura:** Los votos podrían ser cifrados en el momento de la emisión. Las autoridades electorales podrían sumar todos los votos cifrados para obtener un resultado final cifrado. Solo al descifrar ese único resultado final se revelaría el conteo, garantizando que nadie pueda ver los votos individuales durante el proceso de recuento.

Pregunta 2: Técnicas de Auditoría y Monitoreo en DB Relacionales.

- ¿Cuál es el propósito principal de la auditoría y el monitoreo en la seguridad de bases de datos?

El propósito principal de la auditoría y el monitoreo en la seguridad de bases de datos cuenta con dos aristas principales: **proactiva** y **reactiva.**

- **Detección de Amenazas (proactivo):** Permite identificar comportamientos anómalos o maliciosos en tiempo real. Ej: Monitoreo constante puede detectar un ataque de fuerza bruta mientras está ocurriendo y alertar a los administradores para que tomen medidas en la inmediatez.
- **Análisis Forense (reactivo):** Si ocurre una brecha de seguridad, los registros de auditoría son fundamentales para la investigación posterior. Permiten reconstruir lo sucedido determinando la razón y la vulnerabilidad explotada para su posterior refuerzo.

- Eventos a monitorear

1. **Intentos de Inicio de Sesión Fallidos:** Un número elevado de intentos de login fallidos desde una misma IP o para un mismo usuario puede ser un claro indicador de un ataque de fuerza para adivinar la contraseña.
2. **Cambios en los Permisos de Usuario o en los Esquemas:** Monitorear sentencias como GRANT, REVOKE o ALTER TABLE es de suma importancia. Un cambio inesperado en los privilegios de un usuario podría señalar un **escalamiento de privilegios** no autorizados**,** es decir, donde un atacante procura de manera ilegítima darse a sí mismo más poder dentro del sistema.
3. **Accesos a Tablas con Datos Sensibles:** Se debe registrar cualquier acceso a tablas que contengan información crítica, como datos financieros o personales. Si un usuario es detectado consultando esas tablas fuera del normal desenlace usual, podría ser una señal de **abuso** o de una exfiltración de datos en curso.

- Mejora de detección con Herramientas de Gestión de Eventos e Información de Seguridad (SIEM)

Las Herramientas SIEM mejoran drásticamente la detección de incidentes porque actúan como un cerebro central que analiza información de múltiples fuentes.

- **Centralización y Corrección**: Un sistema SIEM recopila logs del servidor de la base de datos, del SO, del firewall y de otras aplicaciones. Su principal ventaja es su capacidad para **correlacionar eventos** de estas distintas fuentes. Ej: Conectar un inicio de sesión fallido en el firewall como evento 1, con un cambio de privilegios de la database, evento 2 y desde una dirección desconocida que sería el dato 3. Identificando un patrón de ataque complejo y de múltiples aristas casi imposible de detectar por separado.
- **Alertas Automáticas:** Al detectar estos patrones sospechosos, los SIEM generan alertas en **tiempo real**, notificando al equipo de seguridad y administradores para que puedan responder de inmediato con las medidas necesarias para garantizar la integridad.

- Importancia de la auditoría y el monitoreo para el cumplimiento legal y normativo

La auditoría y el monitoreo no son solo una buena práctica técnica, sino también una **obligación legal.** Normativas como la **Ley 25.326** en Argentina y el **GDPR** en la unión europea exigen que las organizaciones implementen “medidas de seguridad adecuadas” para proteger los datos personales.
Los registros de auditoría son la evidencia tangible que permite a una organización demostrar que ha cumplido con esas obligaciones. En caso de una auditoría legal o una investigación por una brecha de datos, estos registros sirven para probar que la empresa tenía sistemas para:

- Controlar quién accedía a los datos.
- Detectar accesos no autorizados.
- Investigar incidentes de seguridad.

Sin estos registros, es muy difícil demostrar el cumplimiento, lo que puede derivar en multas graves y pérdidas de confianza de los clientes.

Pregunta 3: Estrategias de Respaldo y Recuperación para la Resiliencia de Bases de Datos.

- Importancia fundamental de las estrategias de respaldo y recuperación

Las estrategias de respaldo y recuperación son una parte esencial de la **seguridad, disponibilidad y la continuidad del negocio.** Su importancia fundamental radica en que son la **última línea de defensa** contra la pérdida de datos. Mientras que otras son preventivas como el firewall, la copia de seguridad permite recuperarse cuando todo lo demás falla.
Los datos pueden perderse por múltiples causas, como **fallos de hardware, errores humanos, ataques maliciosos o desastres naturales.** Sin un plan de respaldo y recuperación sólido, las consecuencias pueden ser devastadoras, incluyendo **interrupciones operativas, pérdidas financieras y un daño irreparable a la reputación Defina la importancia fundamental de las estrategias de respaldo y recuperación** de la organización.

- Tipos de Respaldo

Existen diferentes tipos de respaldo, y cada uno ofrece un balance entre velocidad, espacio de almacenamiento y complejidad de restauración.

- **Respaldo Completo (Full Backup):**
    - **Descripción:** Consiste en copiar la **totalidad de los datos** de la base de datos en un momento específico.
    - **Aplicación y Consecuencias:** Es el método más simple y seguro. Su principal ventaja es que la **restauración es la más rápida y sencilla,** ya que solo se necesita un archivo. Sin embargo, es el que más tiempo tarda en realizarse y el que más espacio de almacenamiento consume. Se suele realizar con menos frecuencia.
- **Respaldo Diferencial (Differential Backup):**
    - **Descripción:** Guarda únicamente los datos que han cambiado desde el **último respaldo completo.**
    - **Aplicación y Consecuencias:** Es más rápido de crear que un respaldo completo. Para la restauración se necesita el último respaldo completo más el último respaldo diferencial. Ofrece un buen equilibrio entre velocidad de respaldo y simplicidad de restauración.
- **Respaldo Incremental (Incremental Backup):**
    - **Descripción:** Registra solo los cambios que han ocurrido desde el **último respaldo de cualquier tipo.**
    - **Aplicación y Consecuencias:** Es el tipo de respaldo más rápido y el que menos espacio consume.
- **Respaldo del Registro de Transacciones (Transaction Log Backup):**
    - **Descripción:** En las bases de datos que registran cada operación, este respaldo copia ese registro de transacciones.
    - **Aplicación y Consecuencias:** Su principal ventaja es que permite una **recuperación a un punto exacto del tiempo,** minimizando así la pérdida de datos a prácticamente cero.

- La "Regla 3-2-1" para respaldos

Esta regla es una estrategia fundamental y ampliamente reconocida para garantizar que los respaldos sean resilientes y sobrevivan a casi cualquier tipo de desastre.

- **(3) Mantener al menos TRES copias de los datos:** Esto incluye el dato original y al menos dos respaldos. La lógica es: a más copias, menos es la probabilidad de que todas fallen al mismo tiempo.
- **(2) Almacenar las copias en DOS tipos de soportes diferentes:** Los respaldos deben guardarse en medios distintos (Ej: un disco y un servicio en la nube). Esto protege contra fallos específicos del medio de almacenamiento. Si todos tus respaldos están en discos duros de la misma marca y lote, un fallo de fabricación podría comprometer a todos.
- **(1) Guardar al menos UNA copia fuera del sitio:** Almacenar una de las copias en una ubicación física diferente es vital para protegerse contra desastres locales como **incendios, inundaciones o robos.** Si la oficina se incendia, el respaldo de la nube o en otra sucursal sobrevive.

- Relación entre las estrategias de respaldo y la definición del Objetivo de Punto de Recuperación (RPO) y el Objetivo de Tiempo de Recuperación (RTO)

RPO y RTO son los dos objetivos clave que definen y guían toda la estrategia de respaldo y recuperación. Determinan qué tan “caro” es para el negocio perder datos o estar fuera de línea.

- **RPO (Recovery Point Objective - Objetivo de Punto de Recuperación):**
    - **Definición:** Es la **cantidad máxima de datos** que una empresa está dispuesta a perder, medida en el tiempo.
    - **Relación:** El RPO dicta la **frecuencia de los respaldos.** Si una empresa define el RPO de 15 minutos, significa que no puede permitirse más de 15 minutos de trabajo, por lo que realiza respaldos al menos cada 15 minutos. Un RPO más bajo requiere una estrategia de respaldo más frecuente y por lo tanto más costosa.

- **RTO (Recovery Time Objective - Objetivo de Tiempo de Recuperación):**
    - **Definición:** Es el tiempo máximo que el sistema puede estar inactivo después de un desastre.
    - **Relación:** El RTO dicta la **velocidad y eficiencia del proceso de restauración.** Si una empresa tiene un RTO de 1 hora, su equipo de TI debe ser capaz de restaurar toda la base de datos y poner el sistema en funcionamiento en menos de una hora.

Referencias:

1. Cifrado Homomórfico

- **Microsoft Research - Criptografía Homomórfica**: https://www.microsoft.com/en-us/research/project/homomorphic-encryption/
- **IBM Research - Fully Homomorphic Encryption**: https://www.ibm.com/blogs/research/2021/04/08/fully-homomorphic-encryption-is-more-real-than-you-think/
- **OpenFHE**: https://www.openfhe.org/

2. Auditoría y Monitoreo

- **Documentación de Microsoft SQL Server sobre Auditoría**: https://learn.microsoft.com/es-es/sql/relational-databases/security/auditing/sql-server-audit-database-engine?view=sql-server-ver16
- **Documentación de Oracle**: https://docs.oracle.com/en/database/oracle/oracle-database/index.html
  Documentación de PostgreSQL: https://www.postgresql.org/docs/
- **INCIBE (Instituto Nacional de Ciberseguridad de España)**: https://www.incibe.es/

3. Estrategias de Respaldo y Recuperación

- **Documentación de Microsoft SQL Server sobre Respaldo**: https://learn.microsoft.com/es-es/sql/relational-databases/backup-restore/backup-overview-sql-server?view=sql-server-ver16
- **Documentación de Oracle**: https://docs.oracle.com/en/database/oracle/oracle-database/index.html
- **Documentación de PostgreSQL**: https://www.postgresql.org/docs/
- **INCIBE (Instituto Nacional de Ciberseguridad de España)**: https://www.incibe.es/