Esquema de cifrado para la ejecución de consultas en bases de datos cifradas

Lil María Rodríguez Henríquez
Director de tesis: Dr. Debrup Chakraborty

Cinvestav

21 de junio de 2010

Contenido

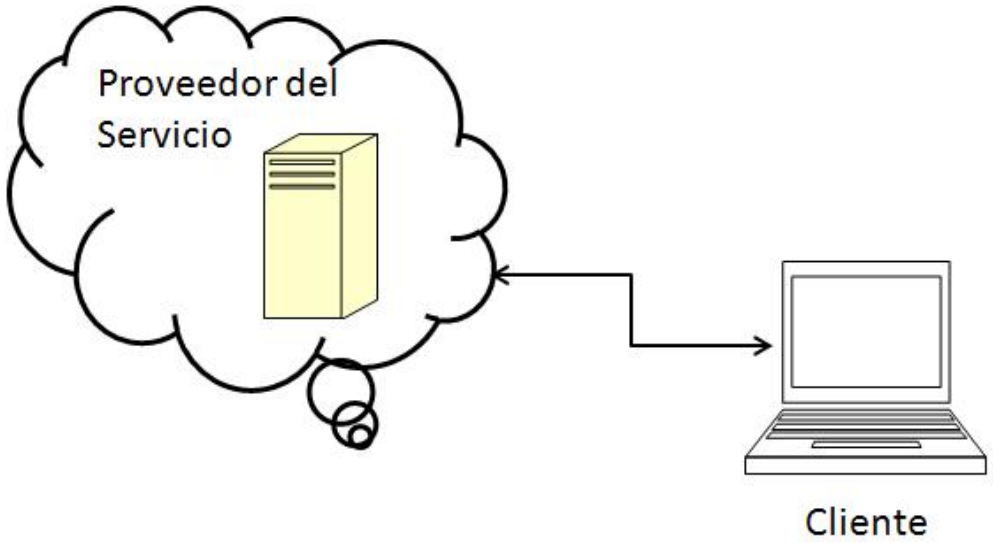

Figura 1: Cómputo Nube: Tecnología que permite ofrecer servicios de computación a través de Internet.

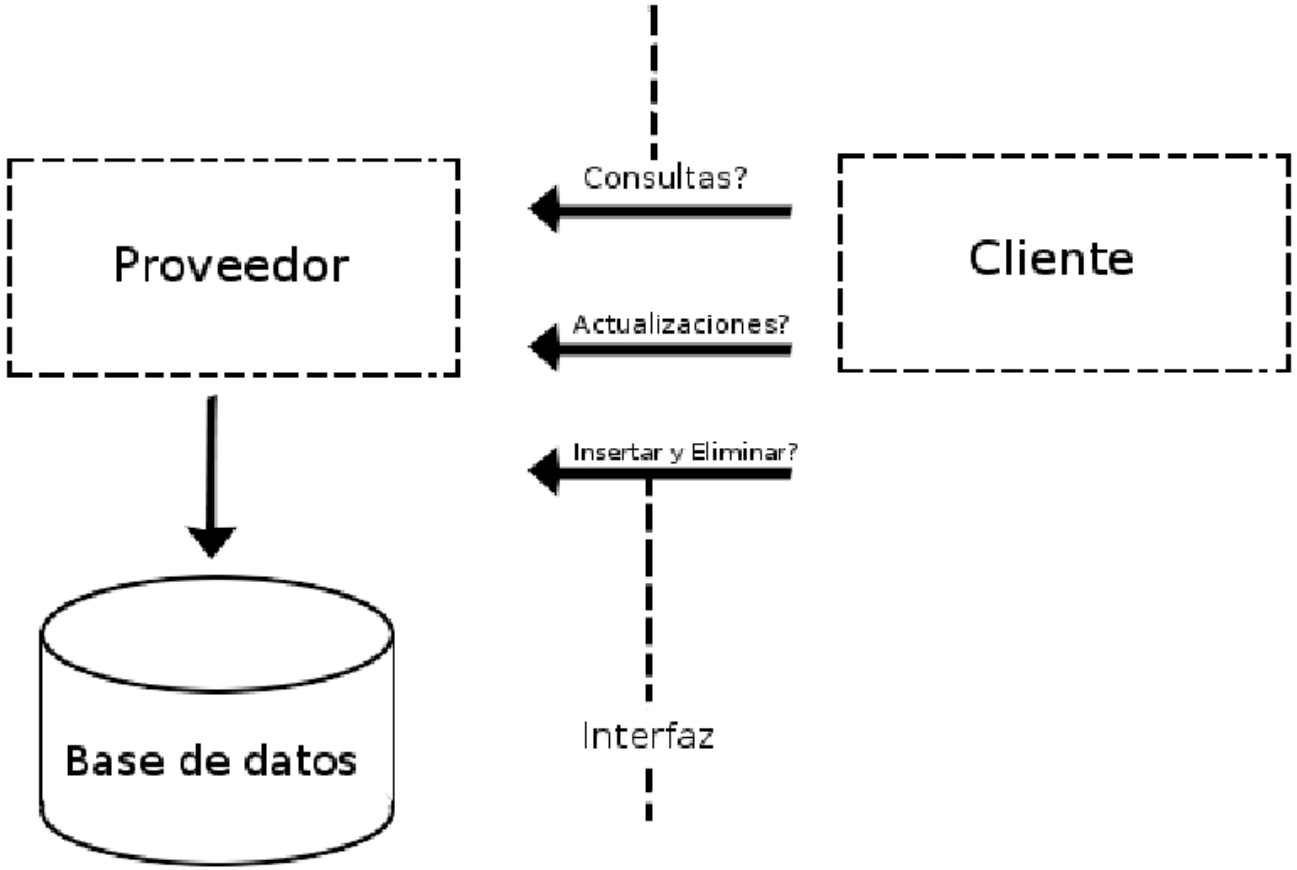

Figura 1: Escenario de bases de datos subcontratadas.

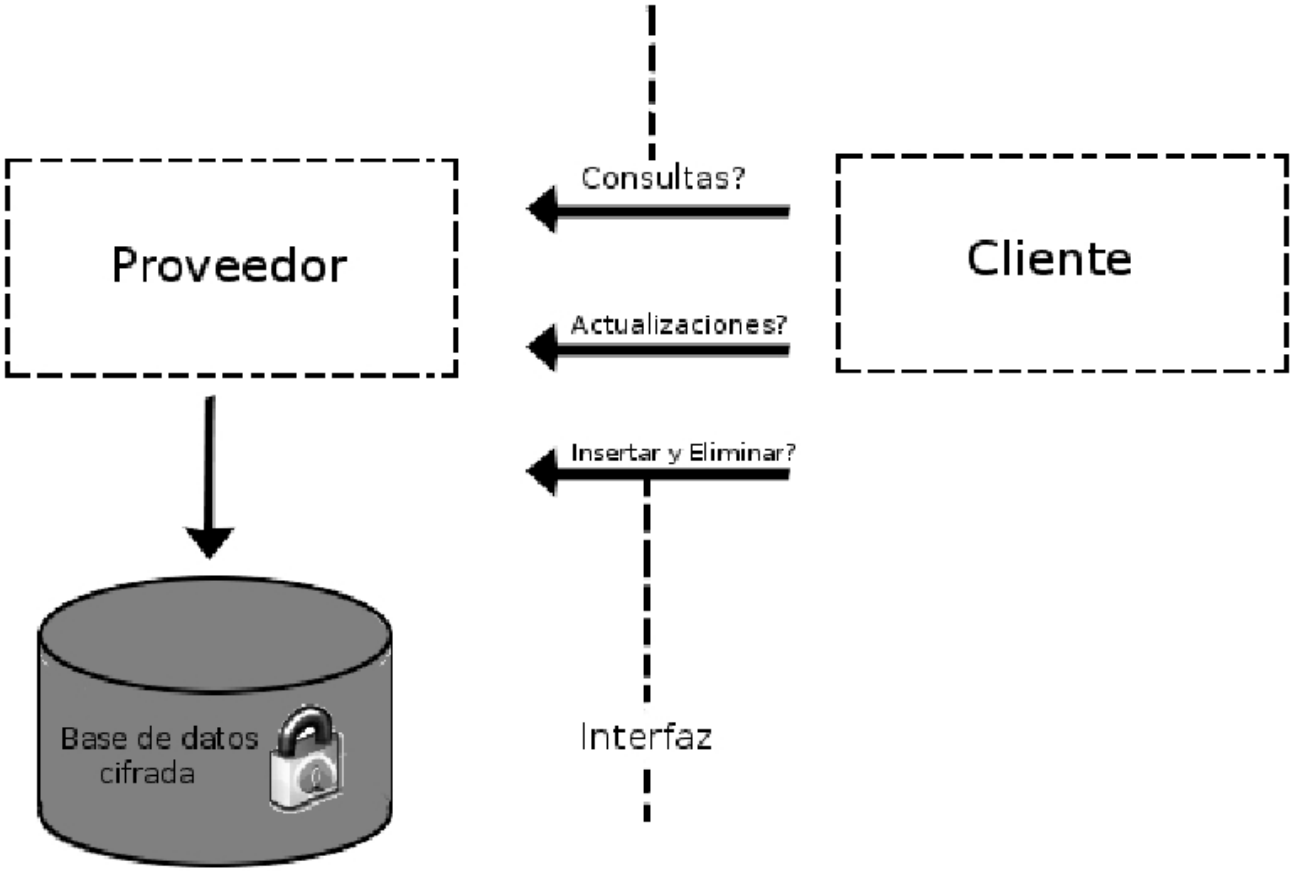

Figura 2: Escenario de bases de datos cifradas subcontratadas.

Características del Problema

- La información le pertenece al cliente, pero no puede administrarla por falta de recursos
- El servidor no es confiable
- Los datos deben ser cifrados del lado del cliente (**¿Cómo?**) y posteriormente se introducen a las tablas
- El servidor debe ser capaz de realizar consultas sin saber el contenido real de la base de datos, no conoce como descifrar

Características del Problema

- La información le pertenece al cliente, pero no puede administrarla por falta de recursos
- El servidor no es confiable
- Los datos deben ser cifrados del lado del cliente (**¿Cómo?**) y posteriormente se introducen a las tablas
- El servidor debe ser capaz de realizar consultas sin saber el contenido real de la base de datos, no conoce como descifrar

Características del Problema

- La información le pertenece al cliente, pero no puede administrarla por falta de recursos
- El servidor no es confiable
- Los datos deben ser cifrados del lado del cliente (**¿Cómo?**) y posteriormente se introducen a las tablas
- El servidor debe ser capaz de realizar consultas sin saber el contenido real de la base de datos, no conoce como descifrar

Características del Problema

- La información le pertenece al cliente, pero no puede administrarla por falta de recursos
- El servidor no es confiable
- Los datos deben ser cifrados del lado del cliente (**¿Cómo?**) y posteriormente se introducen a las tablas
- El servidor debe ser capaz de realizar consultas sin saber el contenido real de la base de datos, no conoce como descifrar

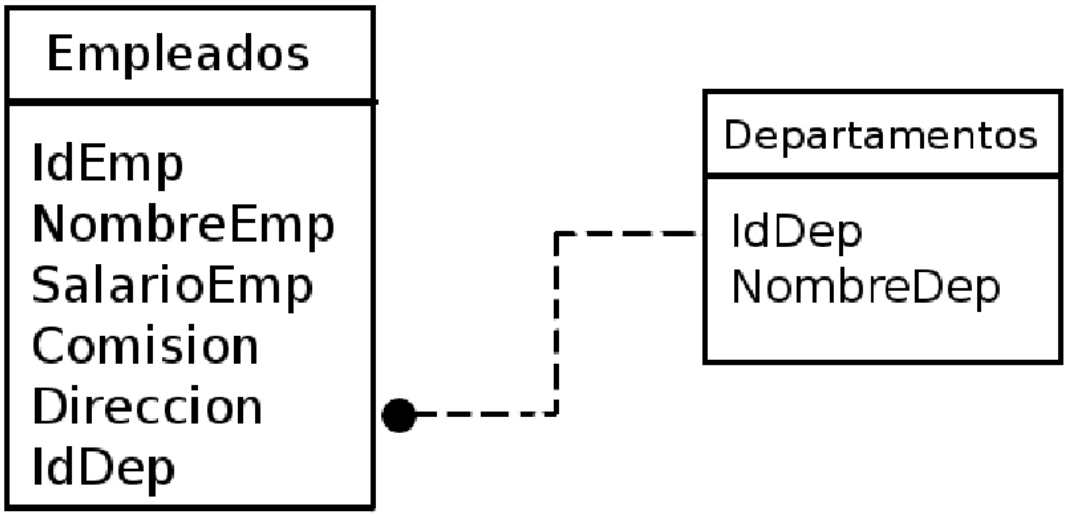

Figura 3: Bases de Datos con dos tablas.

IdEmp	NombreEmp	SalarioEmp	Comisión	Dirección	IdDep
23	Tom	10K	2k	Av.Politecnico	Fin
860	Mary	20K	1k	La Rioja	Comp
320	John	5k	1k	Bamba	Comp
44	Lucy	15k	2k	Matanzas	Comp

Cuadro 1: Tabla Empleados en claro

IdDep	NombreDep
Comp	Computación
Fin	Finanzas

Cuadro 2: Tabla Departamentos en claro

Operaciones relacionales

El cliente debe poder realizar las siguientes operaciones relacionales:

- Insertar
- Consultar
- Modificar
- Eliminar

El lenguaje bajo el cual se estructuran estas operaciones es un estándar y es conocido como *Structured Query Language* SQL, por lo que es posible identificar una sintaxis bien definida para cada una de ellas.

Operaciones relacionales

El cliente debe poder realizar las siguientes operaciones relacionales:

- Insertar
- Consultar
- Modificar
- Eliminar

El lenguaje bajo el cual se estructuran estas operaciones es un estándar y es conocido como *Structured Query Language* SQL, por lo que es posible identificar una sintaxis bien definida para cada una de ellas.

Operaciones relacionales

El cliente debe poder realizar las siguientes operaciones relacionales:

- Insertar
- Consultar
- Modificar
- Eliminar

El lenguaje bajo el cual se estructuran estas operaciones es un estándar y es conocido como *Structured Query Language* SQL, por lo que es posible identificar una sintaxis bien definida para cada una de ellas.

Operaciones relacionales

El cliente debe poder realizar las siguientes operaciones relacionales:

- Insertar
- Consultar
- Modificar
- Eliminar

El lenguaje bajo el cual se estructuran estas operaciones es un estándar y es conocido como *Structured Query Language* SQL, por lo que es posible identificar una sintaxis bien definida para cada una de ellas.

Operaciones relacionales

El cliente debe poder realizar las siguientes operaciones relacionales:

- Insertar
- Consultar
- Modificar
- Eliminar

El lenguaje bajo el cual se estructuran estas operaciones es un estándar y es conocido como *Structured Query Language* SQL, por lo que es posible identificar una sintaxis bien definida para cada una de ellas.

Trabajos Relacionados

- Z.Yang,S.Zhong and R.Wright. Privacy-Preserving Queries on Encrypted Data
  - Orientado a la búsqueda de palabras en documentos cifrados
- Hacigümüş H., Iyer B., y Mehrotra S. Efficient Execution of Aggregation Queries over Encrypted Relational Databases
  - Maneja operaciones aritméticas, permitiendo consultas de agregación, y hace una extensión para el manejo de consultas de intevalo
- Agrawal et al. Order-preserving encryption for numeric data.
  - Cifrado que preserva el orden, permitiendo consultas de intervalo

Objetivo principal

En un modelo de bases de datos subcontratadas, donde el proveedor es en sí un posible oponente, el objetivo es construir un mecanismo que permita que la información del cliente sea cifrada y almacenada en el servidor, y que las funciones del servicio de administración de base de datos (operaciones relacionales), sean realizadas remotamente de manera transparente y eficiente.

Formato de la Consulta

El cliente desea realizar la siguiente consulta:

- *La suma de todos los salarios de los empleados que trabajen en el departamento de Computación y cuyo salario sobre pase los 10k*

La estructura de las consultas responde al estándar SQL, la consulta anterior se escribe como:

- ```
  SELECT SUM(E.salario)
  FROM Empleados as E, Departamentos as D
  WHERE E.did = D.did
  AND D.IdDep =“Comp”
  AND E.salary > 10k
  ```
  ```

Esquema propuesto

Nuestro esquema de cifrado está compuesto por tres algoritmos: un cifrador por bloques (AES), un cifrador basado en homomorfismos (Paillier) y un cifrador que preserva el orden (Boldyreva et.al.); cada uno de éstos desempeña un papel esencial para hacer que el esquema sea lo suficientemente robusto, pues cada uno propicia una o algunas de las operaciones relacionales que se realizan en una base de datos en claro.

- Proceso de cifrado
  - Modelo de almacenaje
- Proceso de consulta en información cifrada
- Proceso de descifrado

Esquema propuesto

Nuestro esquema de cifrado está compuesto por tres algoritmos: un cifrador por bloques (AES), un cifrador basado en homomorfismos (Paillier) y un cifrador que preserva el orden (Boldyreva et.al.); cada uno de éstos desempeña un papel esencial para hacer que el esquema sea lo suficientemente robusto, pues cada uno propicia una o algunas de las operaciones relacionales que se realizan en una base de datos en claro.

- Proceso de cifrado
  - Modelo de almacenaje
- Proceso de consulta en información cifrada
- Proceso de descifrado

Esquema propuesto

Nuestro esquema de cifrado está compuesto por tres algoritmos: un cifrador por bloques (AES), un cifrador basado en homomorfismos (Paillier) y un cifrador que preserva el orden (Boldyreva et.al.); cada uno de éstos desempeña un papel esencial para hacer que el esquema sea lo suficientemente robusto, pues cada uno propicia una o algunas de las operaciones relacionales que se realizan en una base de datos en claro.

- Proceso de cifrado
  - Modelo de almacenaje
- Proceso de consulta en información cifrada
- Proceso de descifrado

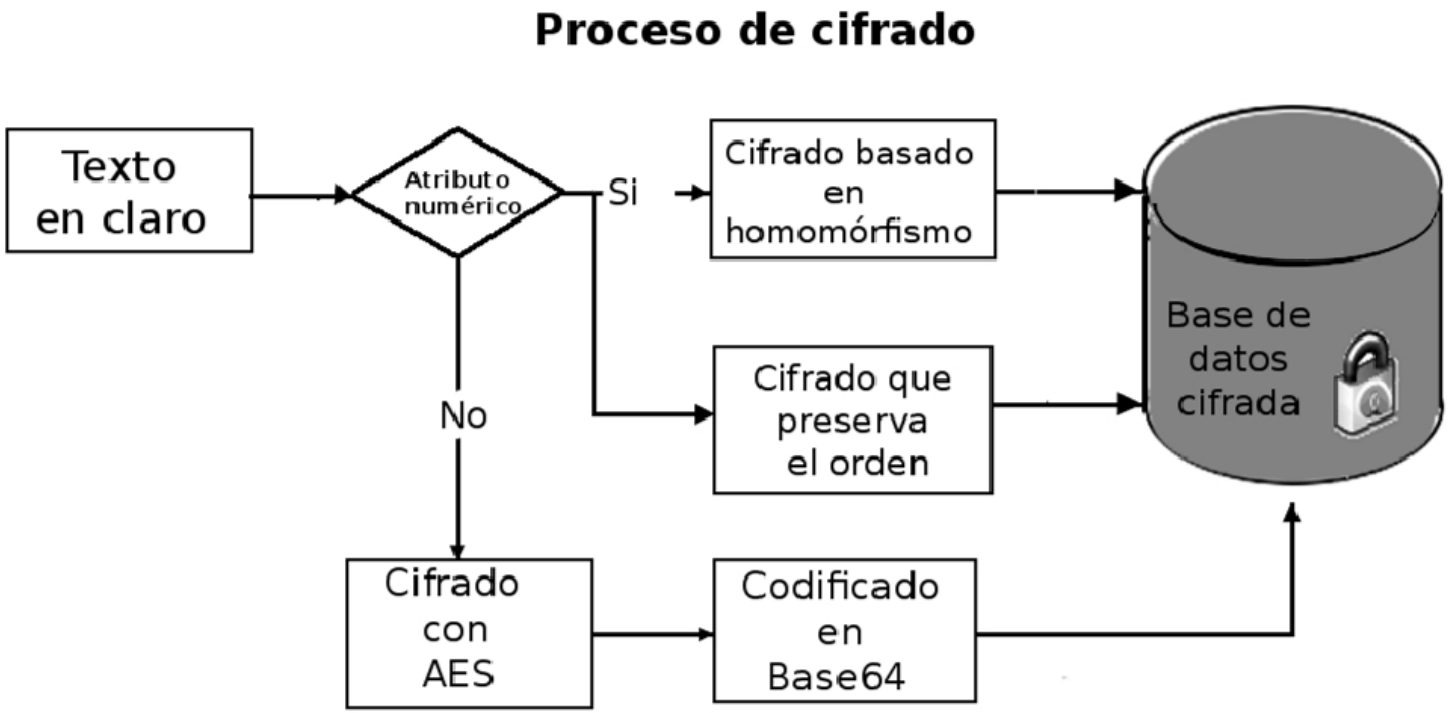

Figura 4: Esquema de cifrado propuesto.

Base de Datos en claro

IdEmp	NombreEmp	SalarioEmp	Comisión	Dirección	IdDep
23	Tom	10K	2k	Av.Politecnico	Fin
860	Mary	20K	1k	La Rioja	Comp
320	John	5k	1k	Bamba	Comp
44	Lucy	15k	2k	Matanzas	Comp

Cuadro 4: Tabla Empleados en claro

IdDep	NombreDep
Comp	Computación
Fin	Finanzas

Cuadro 5: Tabla Departamentos en claro

IdEmpC	NameEmpC	SalarioC	SalarioPO	ComisiónC	ComisiónPO	DirecciónC	IdDepC
?xEw	bcd	7	30	27	7	a34?	dre?
xr?t	ts=	18	50	17	5	dklj	kie
abtx	nmr	2	15	12	5	aiop	kie
t5wt	hlmk	31	40	41	7	tryu	kie

Cuadro 6: Tabla Empleados cifrada

IdDepC	NombreDepC
kie	atr@'?
dre?	k53l

Cuadro 7: Tabla Departamentos cifrada

Formato de la Consulta

El cliente desea realizar la siguiente consulta:

- *La suma de todos los salarios de los empleados que trabajen en el departamento de Computación y cuyo salario sobre pase los 10k*

La estructura de las consultas responde al estándar SQL, la consulta anterior se escribe como:

- ```
  SELECT SUM(E.salario)
  FROM Empleados as E, Departamentos as D
  WHERE E.did = D.did
  AND D.IdDep =“Comp”
  AND E.salary > 10k
  ```
  ```

Proceso de consulta en información cifrada

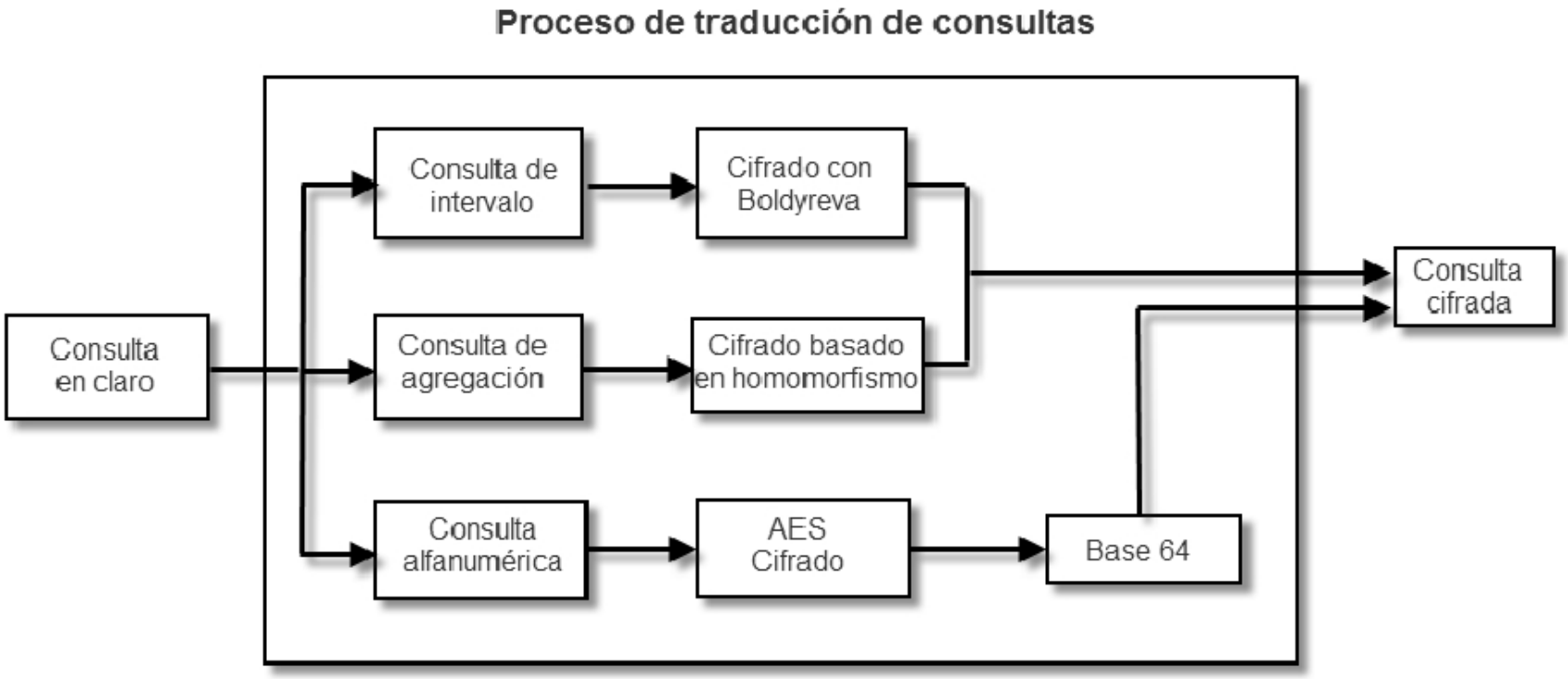

Figura 7: Proceso de traducción de consulta propuesto.

Ejecución de la Consulta

La consulta re-escrita luce asi:

- SELECT E.SalarioC as subtotal
  FROM Empleados as E, Departamentos as D
  WHERE E.IdDep = D.IdDep
  AND D.IdDep =“kie”
  AND E.SalarioPO $>$ 30

Los registros de Mary y Lucy son devueltos.
Se realiza la multiplicación de todos los subtotales, obteniendo un valor acumulado.

$$total = total * subtotal$$

Base de Datos en claro

IdEmp	NombreEmp	SalarioEmp	Comisión	Dirección	IdDep
23	Tom	10K	2k	Av.Politecnico	Fin
860	Mary	20K	1k	La Rioja	Comp
320	John	5k	1k	Bamba	Comp
44	Lucy	15k	2k	Matanzas	Comp

Cuadro 4: Tabla Empleados en claro

IdDep	NombreDep
Comp	Computación
Fin	Finanzas

Cuadro 5: Tabla Departamentos en claro

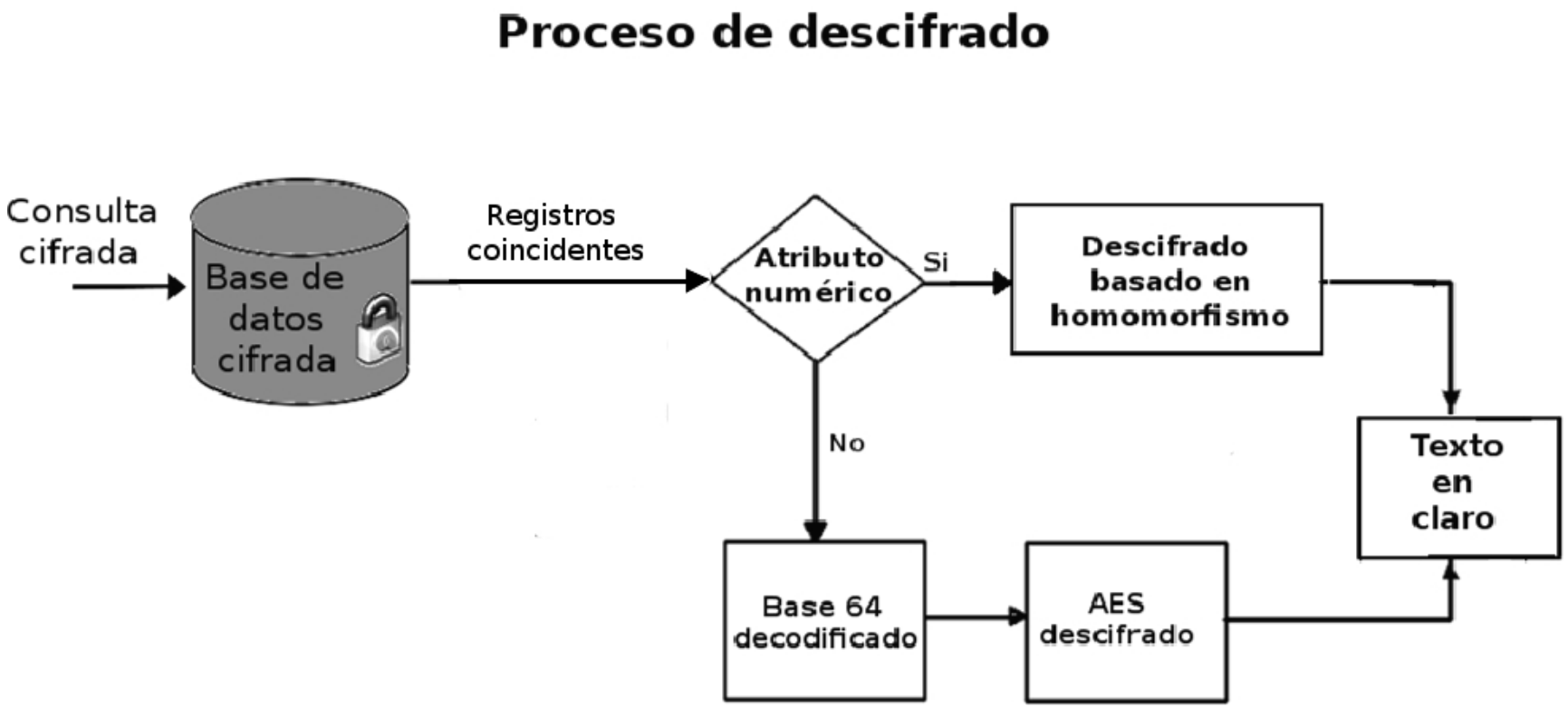

Figura 8: Esquema de descifrado propuesto.

Definición

Un homomorfismo desde un grupo $\langle \mathbb{G}, o_{\mathbb{G}} \rangle$ hacia un grupo $\langle \mathbb{H}, o_{\mathbb{H}} \rangle$, donde $o_{\mathbb{G}}, o_{\mathbb{H}}$ son respectivamente las operaciones de grupo de $\mathbb{G}$ y $\mathbb{H}$, es una función $f : \mathbb{G} \longrightarrow \mathbb{H}$ que cumple con:

$$f(g_1 \; o_{\mathbb{G}} \; g_2) = f(g_1) \; o_{\mathbb{H}} \; f(g_2) \qquad \forall \; g_1, \; g_2 \; \in \; \mathbb{G}$$

Los sistemas criptográficos basados en homomorfismo permiten realizar operaciones en información cifrada. En 1978, Rivest et.al. utilizó las propiedades que brinda el homomorfismo en un sistema criptográfico.

Paillier Scheme

Este esquema fue inventado por Pascal Paillier en 1999, y es un algoritmo asimétrico utilizado en criptografía de llave pública.

- Gen: Se debe generar p y q, donde estos son dos números primos de n-bits. Y $N = p\ q$ además debe calcularse $\phi(N)$
- Enc: la entrada es N y el mensaje $m \in \mathbb{Z}_N$ se calcula de forma aleatoria $r \leftarrow \mathbb{Z}_N^*$ y el cifrado es calculado con

$$c = [(1+N)^m \cdot r^N \ mod\ N^2]$$

- Dec: la entrada es N y $\phi(N)$ y el texto cifrado c, y se calcula:

$$m = \frac{[c^{\phi(N)}\ mod\ N^2] - 1}{N} \cdot \phi(N)^{-1}\ mod\ N$$

Paillier Scheme: Ejemplo Numérico

Sea $N = 11 \cdot 17 = 187$ y $N^2 = 34969$. Sea $m = 175$ y $r = 83 \in \mathbb{Z}^*_{187}$, el texto cifrado se calcula:

$$c = (1 + 187)^{175} \cdot 83^{187} \ mod \ 34969 = 23911$$

Para descifrar se tiene que $\phi(N) = 160$ y que su inverso multiplicativo con respecto a N es 90. Primero calculamos:

$$\tilde{c} = (23911)^{160} \ mod \ 34969 = 25620$$

Restando 1 y dividiendo entre 187 se tiene:

$$\tilde{m} = (25620 - 1)/187 = 137$$

Finalmente el mensaje es recuperado como:

$$m = 137 \cdot 90 \ mod \ 187 = 175$$

Función y cifrado que preservan el orden

Función que preserva el orden

Para $A, B \subseteq \mathbb{N}$ donde $|A| \leq |B|$ una *función que preserva el orden (no decreciente)* es aquella $f : A \longrightarrow B$ tal que para todo $i, j \in A$, $f(i) > f(j)$ si y solamente si $i > j$.

Esquema de cifrado que preserva el orden

Sea un esquema de cifrado determinístico $\mathcal{SE} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ donde $\mathcal{K}$ es el espacio de llaves, $\mathcal{E}nc, Dec$ las funciones de cifrado y descifrado respectivamente, D el espacio de textos en claro y R el espacio de textos cifrados. Entonces, se dice que $\mathcal{SE} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ es un esquema de *cifrado que preserva el orden* si $\mathcal{E}nc(k, \cdot)$ es una función que lo hace, que va D hacia R para toda $k \in \mathcal{K}$.

Función y cifrado que preservan el orden

Función que preserva el orden

Para $A, B \subseteq \mathbb{N}$ donde $|A| \leq |B|$ una *función que preserva el orden (no decreciente)* es aquella $f : A \rightarrow B$ tal que para todo $i, j \in A$, $f(i) > f(j)$ si y solamente si $i > j$.

Esquema de cifrado que preserva el orden

Sea un esquema de cifrado determinístico $\mathcal{SE} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ donde $\mathcal{K}$ es el espacio de llaves, $\mathcal{E}nc, Dec$ las funciones de cifrado y descifrado respectivamente, D el espacio de textos en claro y R el espacio de textos cifrados. Entonces, se dice que $\mathcal{SE} = (\mathcal{K}, \mathcal{E}nc, \mathcal{D}ec)$ es un esquema de *cifrado que preserva el orden* si $\mathcal{E}nc(k, \cdot)$ es una función que lo hace, que va D hacia R para toda $k \in \mathcal{K}$.

Cifrado que Preserva el Orden de Boldyreva et.al.

Se requiere cifrar los números del 1 al 4, y para ello el posible rango de los textos cifrados van del 1 al 8. Para realizar esto, se puede utilizar una distribución Hipergeométrica ya que esta función tiene la característica de preservar el orden. El experimento consiste en una caja que contiene 8 pelotas de las cuales 4 son negras y 4 son blancas, y se irán eligiendo una a una sin que exista reemplazo.

Muestreos	Pelota Resultante	Proyección de Puntos	Total de Pelotas
0	-	-	4 Blancas y 4 Negras
1	Negra	1 - 1	4 Blancas y 3 Negras
2	Blanca	2 - x	3 Blancas y 3 Negras
3	Blanca	2 - x	2 Blancas y 3 Negras
4	Blanca	2 - x	1 Blanca y 3 Negras
5	Negra	2 - 5	1 Blanca y 2 Negras
6	Blanca	3 - x	0 Blancas y 2 Negras
7	Negra	3 - 7	0 Blancas y 1 Negra
8	Negra	4 - 8	0 Blancas y 1 Negra

Cuadro 3: Cifrado que preserva el orden

Operación	Estatus
Inserción Implícita	Permitida
Inserción Explícita	Permitida
Consultas a una tabla	Permitida
Consultas con restricciones de menor o mayor que	Permitida
Consultas con restricciones de menor igual o mayor igual que	Permitica
Consultas con el operador LIKE	**No Implementable**
Consultas con modificadores	Permitida
Consultas con JOINS	Permitida
Subconsultas	Permitida
Consultas con SUM	Permitida
Consultas con MIN, MAX, COUNT	Permitida
Actualizaciones	Permitida
Eliminaciones	Permitida
Consultas que implican sumas y restas	Permitida
Consultas que implican multiplicación y divisiones	**No Implementable**
Consulta en tipos de datos enteros y cadenas	Permitida

Cuadro 10: Tabla resumen

Gracias por todo