UNIVERSIDAD NACIONAL DE LA MATANZA

Departamento de Ingeniería e Investigaciones Tecnológicas

Seguridad y Calidad en Aplicaciones Web

Unidad N° 3: Anexo Conceptos Básicos

Fuente: "Criptografía y Seguridad en Computadoras", Manuel José Lucena López. Capitulo 2, Conceptos Básicos

Capítulo 2

Conceptos Básicos

2.1. Criptografía

Según el Diccionario de la Real Academia, la palabra *criptografía* proviene de la unión de los términos griegos $\kappa\rho\upsilon\pi\tau\acute{o}\varsigma$ (oculto) y $\gamma\rho\acute{\alpha}\varphi\epsilon\iota\nu$ (escritura), y su definición es: *"Arte de escribir con clave secreta o de un modo enigmático"*. Obviamente la Criptografía hace años que dejó de ser un arte para convertirse en una técnica, o más bien un conglomerado de técnicas, que tratan sobre la protección —ocultamiento frente a observadores no autorizados— de la información. Entre las disciplinas que engloba cabe destacar la Teoría de la Información, la Teoría de Números —o Matemática Discreta, que estudia las propiedades de los números enteros—, y la Complejidad Algorítmica.

Existen dos trabajos fundamentales sobre los que se apoya prácticamente toda la teoría criptográfica actual. Uno de ellos, desarrollado por Claude Shannon en sus artículos *"A Mathematical Theory of Communication"* (1948) y *"Communication Theory of Secrecy Systems"* (1949), sienta las bases de la Teoría de la Información y de la Criptografía moderna. El segundo, publicado por Whitfield Diffie y Martin Hellman en 1976, se titulaba *"New directions in Cryptography"*, e introducía el concepto de Criptografía Asimétrica, abriendo enormemente el abanico de aplicación de esta disciplina.

Conviene hacer notar que la palabra Criptografía sólo hace referencia al uso de códigos, por lo que no engloba a las técnicas que se usan para romper dichos códigos, conocidas en su conjunto como *Criptoanálisis*. En cualquier caso ambas disciplinas están íntimamente ligadas; no olvidemos que cuando se diseña un sistema para cifrar información, hay que tener muy presente su posible criptoanálisis, ya que en caso contrario podríamos llevarnos desagradables sorpresas.

Finalmente, el término *Criptología*, aunque no está recogido aún en el Diccionario, se emplea habitualmente para agrupar Criptografía y Criptoanálisis.

2.2. Criptosistema

Definiremos un criptosistema como una quíntupla (M, C, K, E, D), donde:

- M representa el conjunto de todos los mensajes sin cifrar (lo que se denomina texto claro, o *plaintext*) que pueden ser enviados.
- C representa el conjunto de todos los posibles mensajes cifrados, o criptogramas.
- K representa el conjunto de claves que se pueden emplear en el criptosistema.
- E es el conjunto de *transformaciones de cifrado* o familia de funciones que se aplica a cada elemento de M para obtener un elemento de C. Existe una transformación diferente E_k para cada valor posible de la clave k.
- D es el conjunto de *transformaciones de descifrado*, análogo a E.

Todo criptosistema ha de cumplir la siguiente condición:

$$D_k(E_k(m)) = m \tag{2.1}$$

es decir, que si tenemos un mensaje m, lo ciframos empleando la clave k y luego lo desciframos empleando la misma clave, obtenemos de nuevo el mensaje original m.

Existen dos tipos fundamentales de criptosistemas:

- *Criptosistemas simétricos o de clave privada*. Son aquellos que emplean la misma clave k tanto para cifrar como para descifrar. Presentan el inconveniente de que para ser empleados en comunicaciones la clave k debe estar tanto en el emisor como en el receptor, lo cual nos lleva preguntarnos cómo transmitir la clave de forma segura.
- *Criptosistemas asimétricos o de llave pública*, que emplean una doble clave (k_p, k_P). k_p se conoce como *clave privada* y k_P se conoce como *clave pública*. Una de ellas sirve para la transformación E de cifrado y la otra para la transformación D de descifrado. En muchos casos son intercambiables, esto es, si empleamos una para cifrar la otra sirve para descifrar y viceversa. Estos criptosistemas deben

cumplir además que el conocimiento de la clave pública k_P no permita calcular la clave privada k_p. Ofrecen un abanico superior de posibilidades, pudiendo emplearse para establecer comunicaciones seguras por canales inseguros —puesto que únicamente viaja por el canal la clave pública—, o para llevar a cabo autentificaciones.

En la práctica se emplea una combinación de estos dos tipos de criptosistemas, puesto que los segundos presentan el inconveniente de ser computacionalmente mucho más costosos que los primeros. En el *mundo real* se codifican los mensajes (largos) mediante algoritmos simétricos, que suelen ser muy eficientes, y luego se hace uso de la *criptografía asimétrica* para codificar las claves simétricas (cortas).

Claves *Débiles*

En la inmensa mayoría de los casos los conjuntos M y C definidos anteriormente son iguales. Esto quiere decir que tanto los textos claros como los textos cifrados se representan empleando el mismo alfabeto —por ejemplo, cuando se usa el algoritmo DES, ambos son cadenas de 64 bits—. Por esta razón puede darse la posibilidad de que exista algún $k \in K$ tal que $E_k(M) = M$, lo cual sería catastrófico para nuestros propósitos, puesto que el empleo de esas claves dejaría todos nuestros mensajes... ¡sin codificar!

También puede darse el caso de que ciertas claves concretas generen textos cifrados de *poca calidad*. Una posibilidad bastante común en ciertos algoritmos es que algunas claves tengan la siguiente propiedad: $E_k(E_k(M)) = M$, lo cual quiere decir que basta con volver a codificar el criptograma para recuperar el texto claro original. Estas circunstancias podrían llegar a simplificar enormemente un intento de violar nuestro sistema, por lo que también habrá que evitarlas a toda costa.

La existencia de claves con estas características, como es natural, depende en gran medida de las peculiaridades de cada algoritmo en concreto, y en muchos casos también de los parámetros escogidos a la hora de aplicarlo. Llamaremos en general a las claves que no codifican *correctamente* los mensajes *claves débiles (weak keys* en inglés). Normalmente en un buen criptosistema la cantidad de claves débiles es cero o muy pequeña en comparación con el número total de claves posibles. No obstante, conviene conocer esta circunstancia para poder evitar en la medida de lo posible sus consecuencias.

2.3. Esteganografía

La esteganografía —o empleo de *canales subliminales*— consiste en ocultar en el interior de una información, aparentemente inocua, otro tipo de información (cifrada o no). Este método ha cobrado bastante importancia últimamente debido a que permite burlar diferentes sistemas de control. Supongamos que un disidente político quiere enviar un mensaje fuera de su país, evitando la censura. Si lo codifica, las autoridades jamás permitirán que el mensaje atraviese las fronteras independientemente de que puedan acceder a su contenido, mientras que si ese mismo mensaje viaja camuflado en el interior de una imagen digital para una inocente felicitación navideña, tendrá muchas más posibilidades de llegar a su destino.

2.4. Criptoanálisis

El *criptoanálisis* consiste en comprometer la seguridad de un criptosistema. Esto se puede hacer descifrando un mensaje sin conocer la llave, o bien obteniendo a partir de uno o más criptogramas la clave que ha sido empleada en su codificación. No se considera criptoanálisis el descubrimiento de un algoritmo secreto de cifrado; hemos de suponer por el contrario que los algoritmos siempre son conocidos.

En general el criptoanálisis se suele llevar a cabo estudiando grandes cantidades de pares mensaje–criptograma generados con la misma clave. El mecanismo que se emplee para obtenerlos es indiferente, y puede ser resultado de *escuchar* un canal de comunicaciones, o de la posibilidad de que el objeto de nuestro ataque *responda* con un criptograma cuando le enviemos un mensaje. Obviamente, cuanto mayor sea la cantidad de pares, más probabilidades de éxito tendrá el criptoanálisis.

Uno de los tipos de análisis más interesantes es el de *texto claro escogido*, que parte de que conocemos una serie de pares de textos claros —elegidos por nosotros— y sus criptogramas correspondientes. Esta situación se suele dar cuando tenemos acceso al dispositivo de cifrado y éste nos permite efectuar operaciones, pero no nos permite leer su clave —por ejemplo, las tarjetas de los teléfonos móviles GSM—. El número de pares necesarios para obtener la clave desciende entonces significativamente. Cuando el sistema es débil, pueden ser suficientes unos cientos de mensajes para obtener información que permita deducir la clave empleada.

También podemos tratar de criptoanalizar un sistema aplicando el algoritmo de descifrado, con todas y cada una de las claves, a un mensaje codificado que poseemos y comprobar cuáles de las salidas que se obtienen *tienen sentido* como posible texto claro. En general, todas las técnicas que buscan exhaustivamente por el espacio

de claves K se denominan *de fuerza bruta*, y no suelen considerarse como auténticas técnicas de criptoanálisis, reservándose este término para aquellos mecanismos que explotan posibles debilidades intrínsecas en el algoritmo de cifrado. En general, se denomina *ataque* a cualquier técnica que permita recuperar un mensaje cifrado empleando menos esfuerzo computacional que el que se usaría por la fuerza bruta. Se da por supuesto que el espacio de claves para cualquier criptosistema digno de interés ha de ser suficientemente grande como para que los métodos basados en la fuerza bruta sean inviables. Hemos de tener en cuenta no obstante que la capacidad de cálculo de las computadoras crece a gran velocidad, por lo que algoritmos que hace unos años eran resistentes a la fuerza bruta hoy pueden resultar inseguros, como es el caso de DES. Sin embargo, existen longitudes de clave para las que resultaría imposible a todas luces, empleando computación *tradicional*, aplicar un método de este tipo. Por ejemplo, si diseñáramos una máquina capaz de recorrer todas las combinaciones que pueden tomar 256 bits, cuyo consumo fuera mínimo en cada cambio de estado[1], no habría energía suficiente en el Universo para que pudiera completar su trabajo.

Un par de métodos de criptoanálisis que han dado interesantes resultados son el *análisis diferencial* y el *análisis lineal* (ver sección 10.7, página 158). El primero de ellos, partiendo de pares de mensajes con diferencias mínimas —usualmente de un bit—, estudia las variaciones que existen entre los mensajes cifrados correspondientes, tratando de identificar patrones comunes. El segundo emplea operaciones XOR entre algunos bits del texto claro y algunos bits del texto cifrado, obteniendo finalmente un único bit. Si realizamos esto con muchos pares de texto claro–texto cifrado podemos obtener una probabilidad p en ese bit que calculamos. Si p está suficientemente sesgada (no se aproxima a $\frac{1}{2}$), tendremos la posibilidad de recuperar la clave.

Otro tipo de análisis, esta vez para los algoritmos asimétricos, consistiría en tratar de deducir la llave privada a partir de la pública. Suelen ser técnicas analíticas que básicamente intentan resolver los problemas de elevado coste computacional en los que se apoyan estos criptosistemas: factorización, logaritmos discretos, etc. Mientras estos problemas genéricos permanezcan sin solución eficiente, podremos seguir confiando en estos algoritmos.

La Criptografía no sólo se emplea para proteger información, también se utiliza para permitir su autentificación, es decir, para identificar al autor de un mensaje e impedir que nadie suplante su personalidad. En estos casos surge un nuevo tipo de criptoanálisis que está encaminado únicamente a permitir que elementos falsos pasen por buenos. Puede que ni siquiera nos interese descifrar el mensaje original, sino simplemente poder sustituirlo por otro falso y que supere las pruebas de autentificación.

[1] Según las Leyes de la Termodinámica existe una cantidad mínima de energía necesaria para poder modificar el estado de un sistema físico.

Como se puede apreciar, la gran variedad de sistemas criptográficos produce necesariamente gran variedad de técnicas de criptoanálisis, cada una de ellas adaptada a un algoritmo o familia de ellos. Con toda seguridad, cuando en el futuro aparezcan nuevos mecanismos de protección de la información, surgirán con ellos nuevos métodos de criptoanálisis. De hecho, la investigación en este campo es tan importante como el desarrollo de algoritmos criptográficos, y esto es debido a que, mientras que la presencia de fallos en un sistema es posible demostrarla, su ausencia es por definición indemostrable.

2.5. Compromiso entre Criptosistema y Criptoanálisis

En la sección 3.5 (pág. 47) veremos que pueden existir sistemas idealmente seguros, capaces de resistir cualquier ataque. También veremos que estos sistemas en la práctica carecen de interés, lo cual nos lleva a tener que adoptar un compromiso entre el coste del sistema —tanto computacional como de almacenamiento, e incluso económico— frente a su resistencia a diferentes ataques criptográficos.

La información posee un tiempo de vida, y pierde su valor transcurrido éste. Los datos sobre la estrategia de inversiones a largo plazo de una gran empresa, por ejemplo, tienen un mayor periodo de validez que la exclusiva periodística de una sentencia judicial que se va a hacer pública al día siguiente. Será suficiente, pues, tener un sistema que garantice que el tiempo que se puede tardar en comprometer su seguridad es mayor que el tiempo de vida de la propia información que éste alberga. Esto no suele ser fácil, sobre todo porque no tardará lo mismo un *oponente* que disponga de una única computadora de capacidad modesta, que otro que emplee una red de supercomputadores. Por eso también ha de tenerse en cuenta si la información que queremos proteger vale más que el esfuerzo de criptoanálisis que va a necesitar, porque entonces puede que no esté segura. La seguridad de los criptosistemas se suele medir en términos del número de computadoras y del tiempo necesarios para romperlos, y a veces simplemente en función del dinero necesario para llevar a cabo esta tarea con garantías de éxito.

En cualquier caso hoy por hoy existen sistemas que son muy poco costosos —o incluso gratuitos, como algunas versiones de PGP—, y que nos garantizan un nivel de protección tal que toda la potencia de cálculo que actualmente hay en el planeta sería insuficiente para romperlos.

Tampoco conviene depositar excesiva confianza en el algoritmo de cifrado, puesto que en el proceso de protección de la información existen otros puntos débiles que deben ser tratados con un cuidado exquisito. Por ejemplo, no tiene sentido emplear

algoritmos con niveles de seguridad extremadamente elevados si luego escogemos contraseñas (*passwords*) ridículamente fáciles de adivinar. Una práctica muy extendida por desgracia es la de escoger palabras clave que contengan fechas, nombres de familiares, nombres de personajes o lugares de ficción, etc. Son las primeras que un atacante avispado probaría. Tampoco es una práctica recomendable anotarlas o decírselas a nadie, puesto que si la clave cae en malas manos, todo nuestro sistema queda comprometido, por buenos que sean los algoritmos empleados.

2.6. Seguridad

El concepto de seguridad en la información es mucho más amplio que la simple protección de los datos a nivel *lógico*. Para proporcionar una seguridad real hemos de tener en cuenta múltiples factores, tanto internos como externos. En primer lugar habría que caracterizar el sistema que va a albergar la información para poder identificar las amenazas, y en este sentido podríamos hacer la siguiente subdivisión:

1. *Sistemas aislados*. Son los que no están conectados a ningún tipo de red. De unos años a esta parte se han convertido en minoría, debido al auge que ha experimentado Internet.

2. *Sistemas interconectados*. Hoy por hoy casi cualquier ordenador pertenece a alguna red, enviando y recogiendo información del exterior casi constantemente. Esto hace que las redes de ordenadores sean cada día más complejas y supongan un peligro potencial que no puede en ningún caso ser ignorado.

En cuanto a las cuestiones de seguridad que hemos de fijar podríamos clasificarlas de la siguiente forma:

1. *Seguridad física*. Englobaremos dentro de esta categoría a todos los asuntos relacionados con la salvaguarda de los soportes físicos de la información, más que de la información propiamente dicha. En este nivel estarían, entre otras, las medidas contra incendios y sobrecargas eléctricas, la prevención de ataques terroristas, las políticas de copias de respaldo (*backups*), etc. También se suelen tener en cuenta dentro de este punto aspectos relacionados con la restricción de acceso físico a las computadoras únicamente a personas autorizadas.

2. *Seguridad de la información*. En este apartado prestaremos atención a la preservación de la información frente a observadores no autorizados. Para ello podemos

emplear tanto criptografía simétrica como asimétrica, estando la primera únicamente indicada en sistemas aislados, ya que si la empleáramos en redes, al tener que transmitir la clave por el canal de comunicación, estaríamos asumiendo un riesgo excesivo.

3. *Seguridad del canal de comunicación*. Los canales de comunicación rara vez se consideran seguros. Debido a que en la mayoría de los casos escapan a nuestro control, ya que pertenecen a terceros, resulta imposible asegurarse totalmente de que no están siendo escuchados o intervenidos.

4. *Problemas de autentificación*. Debido a los problemas del canal de comunicación, es necesario asegurarse de que la información que recibimos en la computadora viene de quien realmente creemos que viene, y que además no ha sido alterada. Para esto se suele emplear criptografía asimétrica en conjunción con funciones resumen (*hash*).

5. *Problemas de suplantación*. En las redes tenemos el problema añadido de que cualquier usuario autorizado puede acceder al sistema desde fuera, por lo que hemos de confiar en sistemas fiables para garantizar que los usuarios no están siendo suplantados por intrusos. Para conseguir esto normalmente se emplean mecanismos basados en contraseñas.

6. *No repudio*. Cuando se recibe un mensaje no sólo es necesario poder identificar de forma unívoca al remitente, sino que éste asuma todas las responsabilidades derivadas de la información que haya podido enviar. En este sentido es fundamental impedir que el emisor pueda *repudiar* un mensaje, es decir, negar su autoría sobre él.

7. *Anonimato*. Es, en cierta manera, el concepto opuesto al del no repudio. En determinadas aplicaciones, como puede ser un proceso electoral o la simple denuncia de violaciones de los derechos humanos en entornos dictatoriales, es crucial garantizar el anonimato del ciudadano para poder preservar su intimidad y su libertad. Es una característica realmente difícil de conseguir, y desafortunadamente no goza de muy buena fama, especialmente en países donde prima la *seguridad nacional* sobre la libertad y la intimidad de los ciudadanos.

2.7. Autentificación

Como ya se ha dicho, el concepto de autentificación viene asociado a la comprobación del origen e integridad de la información. En general, y debido a los diferentes

tipos de situaciones que podemos encontrar en un sistema informático, distinguiremos tres tipos de autentificación:

- Autentificación *de mensaje*. Queremos garantizar la procedencia de un mensaje conocido, de forma que podamos asegurarnos de que no es una falsificación. Este mecanismo se conoce habitualmente como *firma digital*.
- Autentificación *de usuario mediante contraseña*. En este caso se trata de garantizar la presencia de un usuario legal en el sistema. El usuario deberá poseer una contraseña secreta que le permita identificarse.
- Autentificación *de dispositivo*. Se trata de garantizar la presencia frente al sistema de un dispositivo concreto. Este dispositivo puede estar solo o tratarse de una *llave electrónica* que sustituye a la contraseña para identificar a un usuario.

Nótese que la autentificación de usuario por medio de alguna característica biométrica, como pueden ser las huellas digitales, la retina, el iris, la voz, etc. puede reducirse a un problema de autentificación de dispositivo, solo que el *dispositivo* en este caso es el propio usuario.