Unidad 8 : La Capa de Presentación

Representación de datos

Los diferentes ordenadores tienen diferentes representaciones internas para los datos. Es muy difícil que los fabricantes lleguen a cambiar estas convenciones (para evitar que sus nuevos productos sean incompatibles con los anteriores) y por lo tanto, es poco probable que se logre adoptar alguna vez una norma universal para la representación interna de los datos.

El uso de diferentes formatos internos trae graves consecuencias. Para resolver este problema, se tendrá que realizar una conversión en algún lugar. Se han propuesto muchas alternativas: el extremo transmisor podría hacer la conversión; el receptor también podría hacerla; o los dos podrían hacer una conversión hacia, y desde, un formato normalizado de red.

Compresión de datos

Casi en la totalidad de los casos, el costo por utilizar una red depende de la cantidad de datos transmitidos. Por esto, cuanto más grande sea el número de octetos transmitidos, mayor será el costo de dicha transmisión, por lo que sería muy útil realizar una compresión de datos antes de enviarlos.

La compresión de datos está muy relacionada con su representación. Se ha utilizado para ahorrar espacio en la memoria, en los discos y en cintas magnéticas.

Seguridad y confidencialidad

Durante las primeras décadas de su existencia, las redes de computadoras fueron usadas principalmente por investigadores universitarios para el envío de correo electrónico, y por empleados corporativos para compartir impresoras.

Con el advenimiento de las redes, la situación cambió radicalmente. Nadie puede supervisar manualmente los millones de bits de datos que diariamente se mueven entre los ordenadores en una red. Además, las organizaciones no tienen ninguna manera de asegurar que sus datos no se puedan copiar secretamente, mediante la intercepción de líneas telefónicas o algún otro medio, en el camino que siguen hasta llegar a su destino . Por esto surge la necesidad de algún tipo de puesta en clave (cifrado), con objeto de hacer que los datos sean ininteligibles para todo el mundo, exceptuando aquéllos a los cuales se desea hacer llegar dichos datos.

Existen, por lo menos, cuatro servicios de seguridad relacionados con la seguridad en la conexión entre redes:

1. Proteger los datos para que no puedan ser leídos por personas que no tienen autorización para hacerlo.
2. Impedir que las personas sin autorización inserten o borren mensajes.
3. Verificar el emisor de cada uno de los mensajes.
4. Hacer posible que los usuarios transmitan electrónicamente documentos firmados.

La puesta en clave, puede utilizarse efectivamente para todos estos objetivos. Si bien el cifrado puede realzarse en cualquier capa del modelo OSI, el planteamiento más sofisticado consiste en colocarla en la capa de presentación, para que sólo aquellas estructuras o campos que necesiten cifrarse sufran la sobrecarga correspondiente.

Técnicas de compresión de datos

Los datos que se transmiten por un canal pueden verse como una secuencia de símbolos. Se supone que estos símbolos se extrajeron de algún conjunto (posiblemente finito) de símbolos. Algunos ejemplos de estos conjuntos son:

- Conjunto de bits: 0, 1.
- Conjunto de dígitos decimales: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.
- Conjunto de letras: A, B, C, ..., X, Y, Z.

La compresión de datos puede obtenerse de tres maneras generales, basadas en la finitud del conjunto de símbolos, las frecuencias relativas de utilización de los símbolos y el contexto en el que aparece un símbolo.

Codificación de un conjunto finito de símbolos igualmente probables

En muchas aplicaciones, los mensajes se extraen de un conjunto finito y se expresan en ASCII. En un proyecto de automatización de biblioteca, por ejemplo, los títulos de la colección de la biblioteca podrían considerarse apropiadamente como un conjunto finito de símbolos. Supongamos que cada día se envía una lista completa de los libros pedidos a cada oficina de la biblioteca. La transmisión diaria podría consistir en el número de oficina, , seguido por la lista de todos los libros pedidos ese día. Si consideramos de 20 caracteres el titulo de un libro típico, expresado en ASCII, se necesitarían 140 bits. Pero si simplemente se da a cada libro un número de secuencia, se reduce considerablemente el número de bits.

Codificación dependiente de la frecuencia

En casi todos los textos, algunos símbolos aparecen con mayor frecuencia que otros. En los textos en inglés, por ejemplo, la letra "E" ocurre 100 veces más que la letra "Q", y la palabra "THE", 10 veces más que la palabra "BE". Esta observación sugiere un esquema de codificación en la que, a los símbolos comunes se les asignen códigos cortos, mientras que a los símbolos ocasionales se les asignen códigos largos.

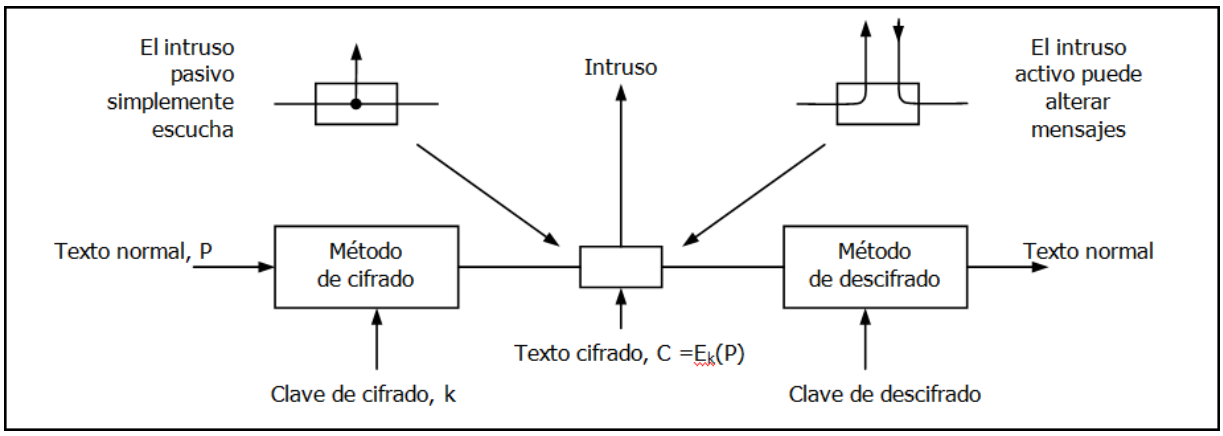
Codificación dependiente del contexto

El método anterior supone, implícitamente, que la probabilidad de ocurrencia de un símbolo es independiente de su predecesor inmediato, o sea, que la probabilidad de la aparición de una letra "T" siguiendo directamente a un a "Q" es casi cuatro veces mayor que la probabilidad de aparición de una "U", siguiendo a una "Q". Con la ayuda de un esquema un poco más sofisticado se podría determinar la probabilidad condicional de cada símbolo , para cada uno de los posibles predecesores. Si existen fuertes correlaciones entre los símbolos y sus sucesores, este método obtiene grandes ahorros, aun cuando los mismos símbolos tengan una distribución plana.

La desventaja de este método de probabilidad condicional es el gran número de tablas que se necesitan. Si hay k símbolos, por ejemplo, las tablas correspondientes deberán tener k^2 entradas.

Criptografía tradicional

Hasta la llegada de las computadoras, una de las restricciones principales de la criptografía había sido la falta de los codificadores para efectuar las transformaciones necesarias, con frecuencia en campo de batalla y contando con poco equipo. Una restricción adicional, ha sido la dificultad para conmutar rápidamente de un método criptográfico a otro, dado que esto obligaría a reentrenar un gran número de personas. Sin embargo, el peligro de que un empleado fuera capturado por el enemigo ha hecho indispensable la capacidad de cambiar el método de cifrado al instante, de ser necesario. De estos requisitos en conflicto se deriva el siguiente modelo.



Los mensajes que se tienen que poner en clave, conocidos como **texto normal**, se transforman mediante una función parametrizada por una **clave**. La salida del proceso de cifrado, conocida como **texto cifrado**, se transmite después, muchas veces mediante mensajero o radio. Suponemos que el **intruso** escucha y copia con exactitud el texto cifrado completo. Sin embargo, a diferencia del destinatario original, el intruso no conoce la clave de cifrado y no puede descifrar fácilmente el texto cifrado. A veces el intruso no sólo puede escuchar el canal de comunicación (intruso pasivo) sino que también puede registrar mensajes y reproducirlos después, inyectar sus propios mensajes y modificar los mensajes legítimos antes de que lleguen al destinatario (intruso activo). El arte de descifrar se llama **criptoanálisis**. Y el arte de diseñar cifradores (criptografía) y de descifrarlos (criptoanálisis) se conocen colectivamente como **criptología**.

Utilizaremos como notación: $C = E_k(P)$ para indicar que el cifrado del texto normal P usando la clave k da el texto cifrado C. Del mismo modo, $P = D_k(C)$ representa el descifrado de C para obtener el texto normal nuevamente. Por tanto:

$D_k (E_k (P)) = P$

Esta notación sugiere que E y D son sólo funciones matemáticas, lo cual es cierto. El único truco es que ambas son funciones de dos parámetros, y hemos escrito uno de los parámetros (la clave) como subíndice, en lugar de cómo argumento, para distinguirlo del mensaje.

La clave consiste en una cadena corta (relativamente) que selecciona uno de muchos cifrados potenciales y puede cambiarse con la frecuencia requerida.. Por tanto, el modelo básico es un método general estable y conocido públicamente pero parametrizado por una clave secreta y fácilmente cambiable.

Cuanto más grande es la clave, mayor será el **factor de trabajo** que tendrá que enfrentar el criptoanalista. Este factor crece exponencialmente con la longitud de la clave. El secreto radica en tener un algoritmo robusto (pero público) y una clave larga.

Los modelos de cifrado históricamente se dividen en dos categorías: cifrados por sustitución y cifrados por transposición.

Cifrados por sustitución

En un cifrado por sustitución, cada letra o grupo de letras se reemplaza por otra letra o grupo de letras para disfrazarla. Cada uno de los símbolos del texto normal, por ejemplo las 26 letras del abecedario (sin la ñ) inglés, tienen una correspondencia con alguna otra letra. Por ejemplo,

Texto normal:

Texto cifrado:

Este sistema general se llama **sustitución monoalfabética**, siendo la clave la cadena de 26 letras correspondientes al alfabeto completo. Para la clave anterior, el texto normal *ataque* se transformaría en el texto cifrado *QZQJXT*.

A primera vista, esto podría parecer seguro, ya que aunque el criptoanalista conoce el sistema general, no sabe cuál de las 26! = 4 x 10^{26} claves posibles se está usando. No obstante, si se cuenta con una cantidad pequeña de texto cifrado, puede descifrarse fácilmente. La forma más común es aprovechar las propiedades estadísticas de los lenguajes naturales. Otro enfoque es tratar de adivinar una palabra o frase probable, dependiente del contexto.

Cifrado por transposición

Los **cifrados por transposición** reordenan las letras del texto normal. Un cifrado de transposición común es la transposición columnar como se muestra a continuación:

M	E	G	A	B	U	C	K
7	4	5	1	2	8	3	6
p	l	e	a	s	e	t	r
a	n	s	f	e	r	o	n
e	m	i	l	l	i	o	n
d	o	l	l	a	r	s	t
o	m	y	s	w	i	s	s
b	a	n	k	a	c	c	o
u	n	t	s	i	x	t	w
o	t	w	o	a	b	c	d

Texto normal

Pleasetransferonemilliondollarst

omyswissbankaccountsixtwotwo

Texto cifrado

La clave del cifrado es una palabra o frase que no contiene letras repetidas. El propósito de la clave es numerar las columnas, estando la columna 1 bajo la letra clave más cercana al inicio del alfabeto, y así sucesivamente. El texto normal se escribe horizontalmente, en filas. El texto cifrado se lee por columnas, comenzando por la columna cuya letra clave es la más baja.

Autentificación y firmas digitales

La autenticidad de numerosos documentos legales, financieros y de otro tipo se determina por la presencia o ausencia de una firma manuscrita autorizada. Las fotocopias no cuentan. Para que los sistemas computarizados de mensajes reemplacen el transporte físico de papel y tinta, debe encontrarse una solución a estos problemas.

para proteger al cliente en el caso de que el precio del oro suba mucho

1. El receptor pueda verificar la identidad proclamada del transmisor. Ej. en los sistemas financieros.
2. El transmisor no pueda repudiar después en contenido del mensaje. Ej. para proteger al banco contra fraudes
3. El receptor no haya podido confeccionar el mensaje él mismo. Ej. para proteger al cliente en el caso de que el precio del oro suba mucho y que el banco trate de falsificar un mensaje firmado en el que el cliente solicitó un lingote de oro en lugar de una tonelada.

Autentificación

En sistemas orientados a conexión, la autentificación puede realizarse en el momento en que se establece una sesión. El planteamiento tradicional consiste en hacer que el usuario compruebe su identidad, mediante la presentación de una contraseña. Este método no solamente expone al usuario a una intercepción pasiva, sino que también puede exigirle a la computadora que autentifica (por ejemplo, el banco) mantener una lista interna de contraseñas, lo cual, en sí mismo, viene a ser un

problema potencial de seguridad. Este problema puede solucionarse mediante el empleo de las firmas digitales de clave secreta o pública.

Firma digital de clave secreta

Un enfoque de las firmas digitales sería tener una autoridad central que sepa todo y en quien todos confíen, por ejemplo el *Big Brother* (BB). Cada usuario escoge una clave secreta y la lleva personalmente a las oficinas de BB. Por tanto sólo el usuario y BB conocen la clave secreta.

Un problema potencial de este protocolo es que todos tienen que confiar en BB. Es más, el BB lee todos los mensajes firmados. Por ello es más confiable utilizar la criptografía de clave pública.

Firma digital de clave pública

Esta propuesta utiliza un algoritmo de cifrado (con clave), E, y el algoritmo de descifrado (con clave), D. E y D deben cumplir los siguientes requisitos:

1. D(E(P)) = P
2. Es excesivamente difícil deducir D de E.
3. E no puede descifrarse mediante un ataque de texto normal seleccionado.

El primer requisito dice que, si aplicamos D a un mensaje cifrado, E(P), obtenemos nuevamente el mensaje de texto original P. El tercero es necesario porque los intrusos pueden experimentar a placer con el algoritmo.

El método funciona de la siguiente manera. Una persona, diseña dos algoritmos, E y D, que cumplan los requisitos anteriores. El algoritmo de cifrado y la clave , E, se hacen públicos, de ahí el nombre de **criptografía de clave pública**., pero se mantiene secreta la clave de descifrado. Este método requiere siempre que cada usuario tenga dos claves: una pública, usada por todo el mundo para cifrar mensajes a enviar a ese usuario, y una privada, que el usuario necesita para descifrar los mensajes.