

Funciones resumen

CURSO CRIPTOGRAFÍA Y SEGURIDAD INFORMÁTICA

Ana I. González-Tablas Ferreres

José María de Fuentes García-Romero de Tejada

Lorena González Manzano

Pablo Martín González

uc3m | Universidad **Carlos III** de Madrid

COSEC



ÍNDICE

- 9. Funciones Resumen (*hash*)
 - Funciones resumen
 - Funciones resumen criptográficas
 - Ejemplos

ÍNDICE

- 9. Funciones Resumen (*hash*)
 - **Funciones resumen**
 - Funciones resumen criptográficas
 - Ejemplos

FUNCIONES RESUMEN

- Una función resumen (función *hash*) es una función que acepta un bloque de datos (M) de longitud variable y genera un resumen (hash) de longitud fija

$$H(M) = \text{hash}$$

- El resumen, de tamaño limitado, identifica de “manera única” el bloque de datos, de longitud no limitada

FUNCIONES RESUMEN

- Colisión
 - Espacio de resúmenes $|h|$ para una función resumen dada

$$|h| = 2^n$$

Siendo n el número de bits de salida de la función resumen

- Dado que una función resumen genera resúmenes de longitud fija, pero el número de mensajes de entrada a la función es teóricamente infinito, es posible encontrar 2 mensajes M y M' tal que

$$H(M) = H(M') \rightarrow \text{Colisión}$$

ÍNDICE

- 9. Funciones Resumen (*hash*)
 - Funciones resumen
 - **Funciones resumen criptográficas**
 - Ejemplos

FUNCIONES RESUMEN CRIPTOGRÁFICAS

- Función resumen que debe cumplir los siguientes requisitos:

- Ser aplicable a mensajes de entrada de cualquier longitud

- Producir resúmenes de salida de una longitud fija

Compresión

- La salida generada por la función resumen debe satisfacer los requisitos para pseudo-aleatoriedad.

FUNCIONES RESUMEN CRIPTOGRÁFICAS

- **Difusión:** si se modifica un solo bit del mensaje M , el resumen debería cambiar la mitad de sus bits aproximadamente
- **Determinista:** la aplicación de la misma función resumen sobre los mismos datos debe producir el mismo resumen
- **Eficiente:** El cálculo del resumen de un mensaje dado debe ser rápido tanto en implementaciones software como hardware

FUNCIONES RESUMEN CRIPTOGRÁFICAS

- **Resistente a preimágenes:** Dado un resumen h , es *computacionalmente imposible* encontrar un mensaje M' cuyo resumen coincida con el primero (propiedad de una sola vía):

Dado h , encontrar $M' / H(M') = h$

- **Resistente a segunda preimagen (resistencia débil a colisiones):** Dado un mensaje M , es *computacionalmente imposible* encontrar un M' tal que el resumen de ambos coincidan:

Dado M , encontrar $M' \neq M / H(M) = H(M')$

- **Resistente a colisiones (resistencia fuerte a colisiones):** Es *computacionalmente imposible* encontrar dos mensajes M y M' tales que sus resúmenes coincidan:

Encontrar M y M' , $M \neq M' / H(M) = H(M')$

FUNCIONES RESUMEN CRIPTOGRÁFICAS

- “Computacionalmente imposible”
 - No existe algoritmo o técnica para la búsqueda de colisiones que sea más eficiente que la fuerza bruta.
 - Si el espacio de resúmenes generados es suficientemente grande, se puede estimar que, con los recursos HW/SW existentes, la probabilidad de encontrar una colisión es nula en un tiempo razonable
- La fortaleza de una función resumen radica en:
 - Que su diseño sólo permita ataques por fuerza bruta (no criptoanalizable)
 - n (longitud del resumen) sea suficientemente grande

FUNCIONES RESUMEN CRIPTOGRÁFICAS

- Probabilidades de encontrar una colisión (fuerza bruta)
 - Ataque de preimagen: $\frac{1}{2^n}$
 - Ataque de segunda preimagen: $\frac{1}{2^n}$
 - Ataque de colisión: $\frac{1}{2^{n/2}}$ ($p \geq 50\%$) (ataque del cumpleaños)
- En definitiva, la **complejidad algorítmica** (fortaleza) de una función resumen viene determinada por la probabilidad de encontrar una colisión mediante un ataque de colisión.

FUNCIONES RESUMEN CRIPTOGRÁFICAS

- Un algoritmo se considera roto cuando existe un algoritmo de complejidad menor al de fuerza bruta, aunque en la práctica resulte inviable el ataque
- La barrera de 2^{112} establece el mínimo aceptable para una complejidad algorítmica
- Cualquier ataque que requiera un menor número de operaciones convierte al algoritmo en no seguro

FUNCIONES RESUMEN CRIPTOGRÁFICAS

- POSIBLES ATAQUES

- Ataque de preimagen

- Suplantación de identidad en sistemas que almacenen los resúmenes de las contraseñas
 - Forzar falsos positivos en tablas de hashing

- Ataque de segunda preimagen

- Falsificación de certificados digitales, documentos firmados digitalmente, código fuente, etc.

- Ataque de colisión

- Ataque del cumpleaños para la falsificación de documentos firmados digitalmente

FUNCIONES RESUMEN CRIPTOGRÁFICAS

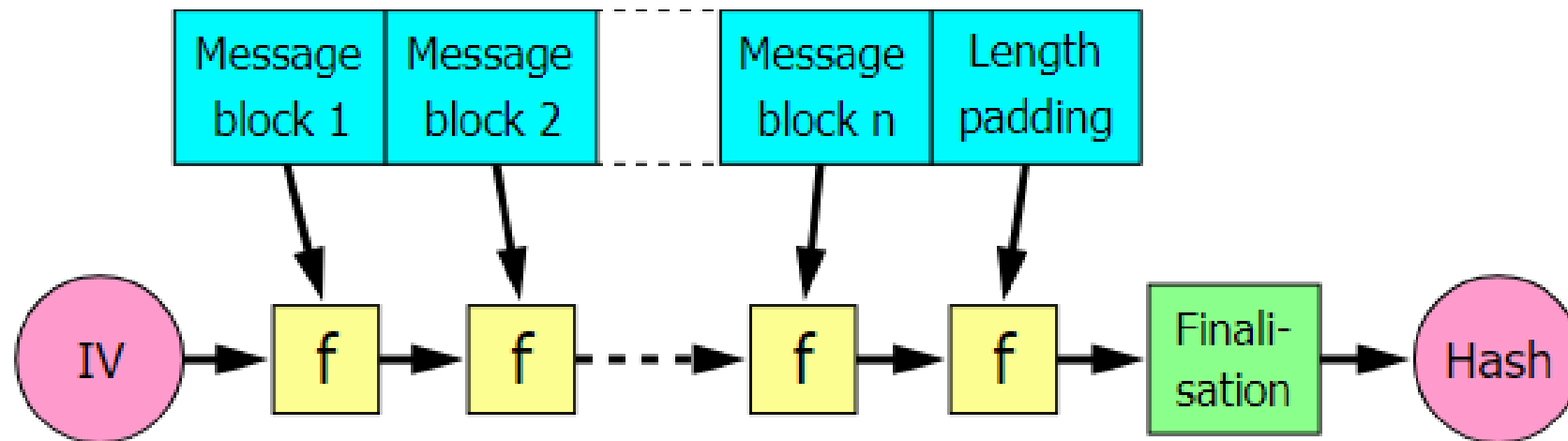
- APLICACIONES PRÁCTICAS
 - Verificación de integridad de datos
 - Firmas digitales
 - Uso en funciones MAC (Message Authentication Code)
 - Indexación en bases de datos, estructuras de datos, etc.
 - Almacenamiento de contraseñas
 - Detección de intrusiones
 - Reconocimiento de patrones de virus
 - Generación de números pseudo-aleatorios
 - Etc.

ÍNDICE

- 9. Funciones Resumen (*hash*)
 - Funciones resumen
 - Funciones resumen criptográficas
 - **Ejemplos**

EJEMPLOS: ESTRUCTURA MERKLE-DAMGÄRD

- USADA EN LA MAYORÍA DE FUNCIONES RESUMEN ACTUALES



$CV_0 = IV = \text{valor inicial del resumen}$

$CV_i = f(CV_{i-1}, B_{i-1}) \quad 1 \leq i \leq L$

$H(M) = CV_L$

EJEMPLOS: ESTRUCTURA MERKLE-DAMGÄRD

- Algoritmo con iteraciones (etapas) encadenadas
- Fase de adecuación del mensaje de entrada
 - El mensaje se divide en L bloques B de longitud b
 - Al último bloque se le añade la longitud total del mensaje
 - Se añade un relleno (*padding*) en caso necesario. Dificulta la búsqueda de colisiones:
 - 2 mensajes de igual longitud que colisionen
 - 2 mensajes de diferente longitud que, con sus longitudes añadidas, colisionen

EJEMPLOS: ESTRUCTURA MERKLE-DAMGÄRD

- FUNCIÓN DE COMPRESIÓN
 - 2 entradas: salida de la etapa anterior (vector de inicialización si primera etapa) + bloque correspondiente
 - Cada etapa produce un resumen de n bits
 - El resumen final generado es de longitud n bits
- Si la función de compresión es resistente a colisiones, también lo es la función resumen (lo contrario no tiene porqué ser cierto)
- Diseño de la función de compresión → núcleo de la seguridad
 - El criptoanálisis a una función resumen se centra en la función de compresión

EJEMPLOS: MD5

- Diseñada por Ronald L. Rivest en 1991
- Modo de operación
 - Genera un resumen de 128 bits
 - El mensaje de entrada se divide en bloques de 512 bits
 - Se produce una operación de relleno sobre el último bloque
 - Cada bloque se descompone a su vez en 16 sub-bloques de 32 bits cada uno
 - Se realizan 4 rondas de 16 operaciones cada una basadas en:
 - Funciones no lineales
 - Suma módulo 2^{32}
 - Rotación de bits

EJEMPLOS: MD5

- ATAQUES

- Primeras señales de vulnerabilidad (1996)
- Primeros algoritmos que darían lugar a las primeras colisiones (2004)

<http://eprint.iacr.org/2004/199>

- Lenstra, Wang y Weger, logran construir dos certificados de claves públicas distintas con la misma firma digital (MD5-RSA) (2005)

<http://eprint.iacr.org/2005/067>

- Algoritmo que encuentra colisiones en un minuto (*Preimage attack by Tunneling*) (2006)

<http://eprint.iacr.org/2006/105>

EJEMPLOS: SHA-0, SHA-1

- SHA-0
 - Genera un resumen de 160 bits
 - Roto en 2005 al publicarse un algoritmo que encontraba colisiones con tan sólo 2^{39} operaciones
- SHA-1
 - Diseñado por la NSA
 - Genera un resumen de 160 bits
 - Estructura similar a la de MD5
 - En 2005, Wang, Yin y Yun publican un algoritmo que encuentra colisiones con 2^{69} operaciones (2^{80} sería con fuerza bruta)
 - En 2005, Wang, Yao y Yao reducen la complejidad del algoritmo a 2^{63} operaciones

EJEMPLOS: FAMILIA SHA-2

- SHA-224, SHA-256, SHA-384 y SHA-512
- Diseñados por NSA
- Nueva estructura común a todas ellas
- SHA-224 y SHA-384 son versiones truncadas de SHA-256 y SHA-512 (64 rondas en vez de 80 y con valores iniciales diferentes)
- No se han encontrado vulnerabilidades
- Son las opciones que ofrecen más garantías

Ejemplos

Algorithm	Output size	Internal state size	Block size	Collision
<u>HAVAL</u>	256/224/192/160/128	256	1024	Yes
<u>MD2</u>	128	384	128	Almost
<u>MD4</u>	128	128	512	Yes
<u>MD5</u>	128	128	512	Yes
<u>RIPEMD</u>	128	128	512	Yes
<u>RIPEMD-128/256</u>	128/256	128/256	512	No
<u>RIPEMD-160/320</u>	160/320	160/320	512	No
<u>SHA-0</u>	160	160	512	Yes
<u>SHA-1</u>	160	160	512	With flaws
<u>SHA-256/224</u>	256/224	256	512	No
<u>SHA-512/384</u>	512/384	512	1024	No
<u>WHIRLPOOL</u>	512	512	512	No

EJEMPLOS: FAMILIA SHA-3

- Competición para la selección de la nueva familia de funciones resumen SHA-3

(<http://csrc.nist.gov/groups/ST/hash/sha-3/>)

- 2007: Establecimiento de los requisitos a cumplir.
- 2008: Envío de propuestas.
- 2009 (Febrero): Primer Congreso sobre las funciones resumen candidatas. Revisión pública de las candidatas.
- 2010 (2Q): Segundo Congreso sobre las funciones resumen candidatas. Análisis de resultados y propuesta de mejoras.
- 2010 (3Q): Selección de las funciones resumen finalistas.
- 2010 (4Q): Últimos “retoques” por parte de los autores.
- 2011: Análisis de la comunidad científica mundial.
- 2012 (4Q): **Keccak** seleccionado como algoritmo ganador (<http://csrc.nist.gov/groups/ST/hash/sha-3/documents/Keccak-slides-at-NIST.pdf>)

CURSO CRIPTOGRAFÍA Y SEGURIDAD INFORMÁTICA

COSEC

uc3m | Universidad **Carlos III** de Madrid

