Universidad
Carlos III de Madrid

COSEC Lab · *Dpto. Informática*

TEMA 2. CRIPTOGRAFÍA
T 2.4 FUNCIONES RESUMEN Y MAC

Criptografía y seguridad informática
Seguridad en las tecnologías de la información
@ COSEC

Curso 2016-2017

Contenidos

- Funciones Resumen
  - Funciones resumen
  - Funciones resumen criptográficas
  - Ejemplos

- Message Authentication Code (MAC)
  - Generalidades MAC
  - Requisitos de seguridad MAC
  - MAC Basados en funciones resumen
  - MAC Basados en cifrado en bloque

Contenidos

- **Funciones Resumen (Hash functions)**
  - **Funciones resumen**
  - Funciones resumen criptográficas
  - Ejemplos

- Message Authentication Code (MAC)
  - Generalidades MAC
  - Requisitos de seguridad MAC
  - MAC Basados en funciones resumen
  - MAC Basados en cifrado en bloque

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

$$H(M) = hash$$

- 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' /

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

Contenidos

- Funciones Resumen
  - Funciones resumen
  - **Funciones resumen criptográficas**
  - Ejemplos

- Message Authentication Code (MAC)
  - Generalidades MAC
  - Requisitos de seguridad MAC
  - MAC Basados en funciones resumen
  - MAC Basados en cifrado en bloque

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

  Dado M, encontrar M' ≠ M / H(M) = H(M')

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

  Encontrar M y M', M ≠ 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≥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^{64} 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
  - Patrones de virus
  - Generación de números pseudo-aleatorios
  - Etc.

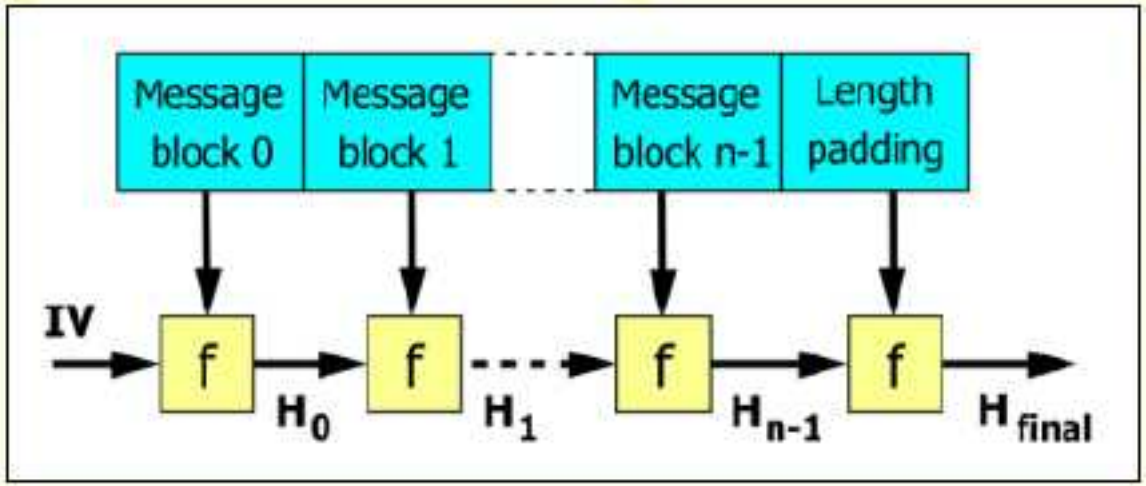
Contenidos

- Funciones Resumen
  - Funciones resumen
  - Funciones resumen criptográficas
  - **Ejemplos**

- Message Authentication Code (MAC)
  - Generalidades MAC
  - Requisitos de seguridad MAC
  - MAC Basados en funciones resumen
  - MAC Basados en cifrado en bloque

Ejemplos – Estructura Merkle-Damgard

- Usada en la mayoría de las funciones resumen actuales



$CV_0 = IV$ = 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-Damgard

- 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-Damgard

- 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
HAVAL	256/224/192/160/128	256	1024	Yes
MD2	128	384	128	Almost
MD4	128	128	512	Yes
MD5	128	128	512	Yes
RIPEMD	128	128	512	Yes
RIPEMD-128/256	128/256	128/256	512	No
RIPEMD-160/320	160/320	160/320	512	No
SHA-0	160	160	512	Yes
SHA-1	160	160	512	With flaws
SHA-256/224	256/224	256	512	No
SHA-512/384	512/384	512	1024	No
WHIRLPOOL	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)

Contenidos

- Funciones Resumen
  - Funciones resumen
  - Funciones resumen criptográficas
  - Ejemplos

- **Message Authentication Code (MAC)**
  - **Generalidades MAC**
  - Requisitos de seguridad MAC
  - MAC Basados en funciones resumen
  - MAC Basados en cifrado en bloque

Generalidades MAC

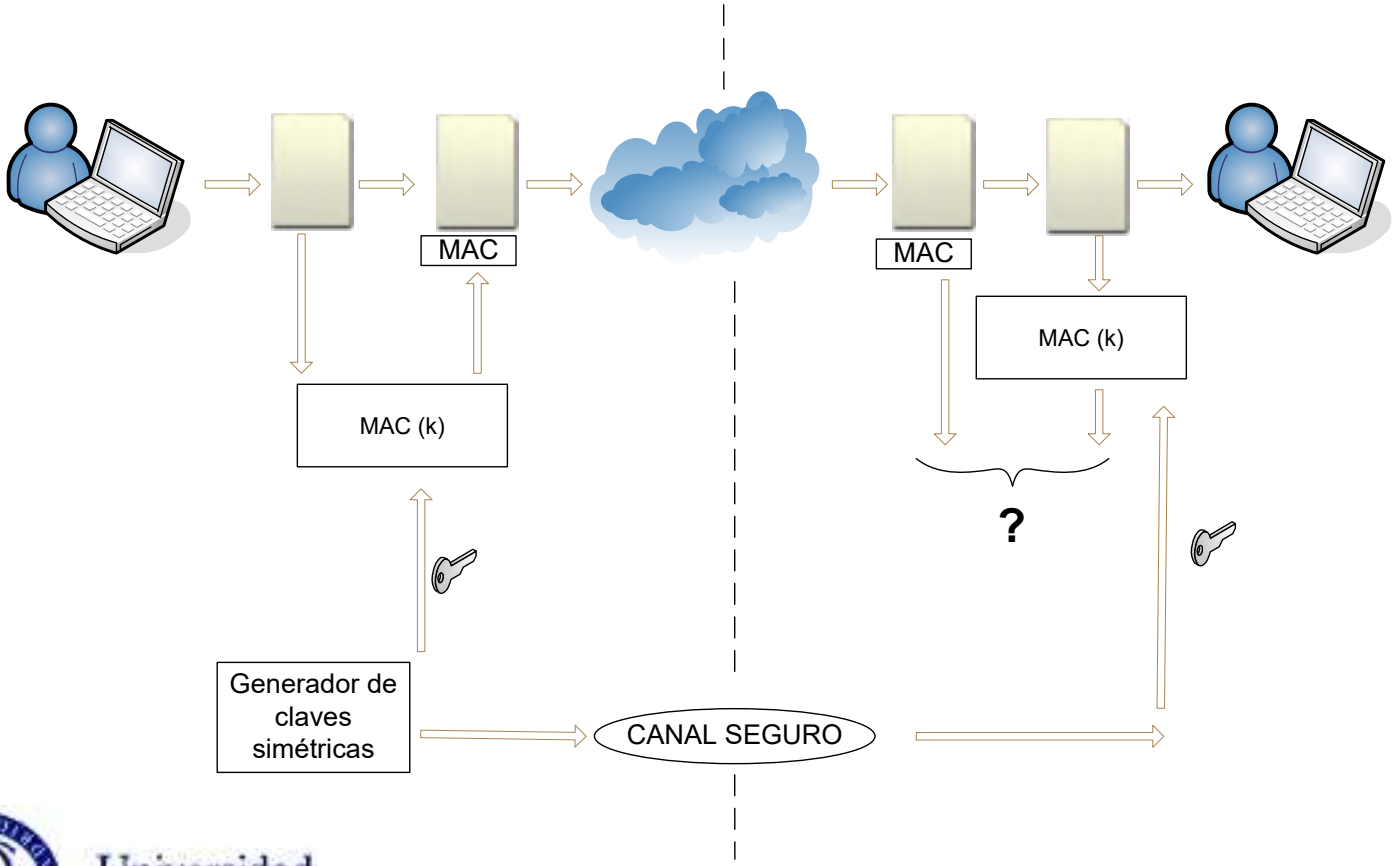
- Un código de autenticación de mensaje (MAC) es un algoritmo que emplea una clave secreta para producir un valor de longitud fija (código de autenticación) sobre un mensaje de longitud variable

- Cualquier entidad que posea la clave secreta es capaz de verificar la **integridad** del mensaje

- Un receptor que comparta la clave secreta es capaz de **autenticar** al origen del mensaje

- En caso que el mensaje incluya un número de secuencia, se evitan ataques por replicación

Generalidades MAC



Generalidades MAC

- Una función MAC no tiene porqué ser invertible

- Al igual que con las funciones resumen, se pueden producir colisiones

  $|k| = 2^k$

  $|MAC| = 2^n$

  $|M|$ = indeterminado

Contenidos

- Funciones Resumen
  - Funciones resumen
  - Funciones resumen criptográficas
  - Ejemplos

- Message Authentication Code (MAC)
  - Generalidades MAC
  - **Requisitos de seguridad MAC**
  - MAC Basados en funciones resumen
  - MAC Basados en cifrado en bloque

Requisitos de seguridad MAC

- Dado un mensaje M y el valor MAC(K, M), es *computacionalmente imposible* encontrar un mensaje M' cuyo valor MAC(K, M') coincida

  Dado M y MAC(K, M), encontrar M'≠M / MAC(K, M') = MAC(K, M)

- MAC(K, M) debe estar uniformemente distribuido, de forma que la probabilidad de encontrar dos mensajes M y M' cuyos valores MAC coincidan es $\frac{1}{2^n}$

- Sea M' un mensaje resultante de aplicar una transformación a M [M' = f(M)]. En tal caso, debe cumplirse lo siguiente:

$$Pr[MAC(K, M) = MAC(K, M')] = \frac{1}{2^n}$$

Requisitos de seguridad MAC

- Ataques a funciones MAC

  Dado un conjunto de M_i, MAC(K,M_i), el atacante desea generar M', MAC(K,M'), con $M' \neq M_i \ \forall \ i=0\ldots n$

  - Fuerza bruta

    Ataque al espacio de claves K ($\frac{1}{2^k}$) versus Ataque al valor MAC ($\frac{1}{2^n}$)

    La complejidad computacional es *Min*($\frac{1}{2^k}$, $\frac{1}{2^n}$)

  - Criptoanálisis

    Requiere la existencia de vulnerabilidades en el diseño o implementación en el algoritmo (dependerá de su estructura interna)

Contenidos

- Funciones Resumen
  - Funciones resumen
  - Funciones resumen criptográficas
  - Ejemplos

- Message Authentication Code (MAC)
  - Generalidades MAC
  - Requisitos de seguridad MAC
  - **MAC Basados en funciones resumen**
  - MAC Basados en cifrado en bloque

MAC Basados en funciones resumen

- HMAC (Hash-MAC)
- Emplean funciones resumen existentes
- Aplican la función resumen sobre una versión del mensaje al que añaden un conjunto de bits calculados a partir de la clave

$$HMAC(K, M) = H[(K' \oplus opad) \,||\, H[(K' \oplus ipad) \,||\, M]]$$

K': K *padded* con 0's a la izquierda hasta tener longitud b

b: Longitud de cada bloque procesado por la función resumen

ipad: 00110110 (0x36) repetido b/8 veces

opad: 01011100 (0x5C) repetido b/8 veces

||: operación concatenación

Contenidos

- Funciones Resumen
  - Funciones resumen
  - Funciones resumen criptográficas
  - Ejemplos

- Message Authentication Code (MAC)
  - Generalidades MAC
  - Requisitos de seguridad MAC
  - MAC Basados en funciones resumen
  - **MAC Basados en cifrado en bloque**

MAC Basados en cifrado en bloque

- Cifran el mensaje mediante un algoritmo de cifrado simétrico en bloque en modo CBC
- El valor del MAC es el resultado del cifrado del último bloque
- Consiguen que el MAC dependa de todos los bits del mensaje

Universidad
Carlos III de Madrid

***Grupo SeTI** · Dpto. Informática*

ANEXO

SEGURIDAD EN LAS TECNOLOGÍAS DE LA INFORMACIÓN

Grupo SeTI

Curso 2011-2012

Funciones Resumen Criptográficas

- Paradoja del Cumpleaños
  - Establece que si hay 23 personas reunidas hay una probabilidad del 50,7% de que al menos dos personas de ellas cumplan años el mismo día.

    P("dos cumpleaños coincidan") = 1- P("ninguno coincida")

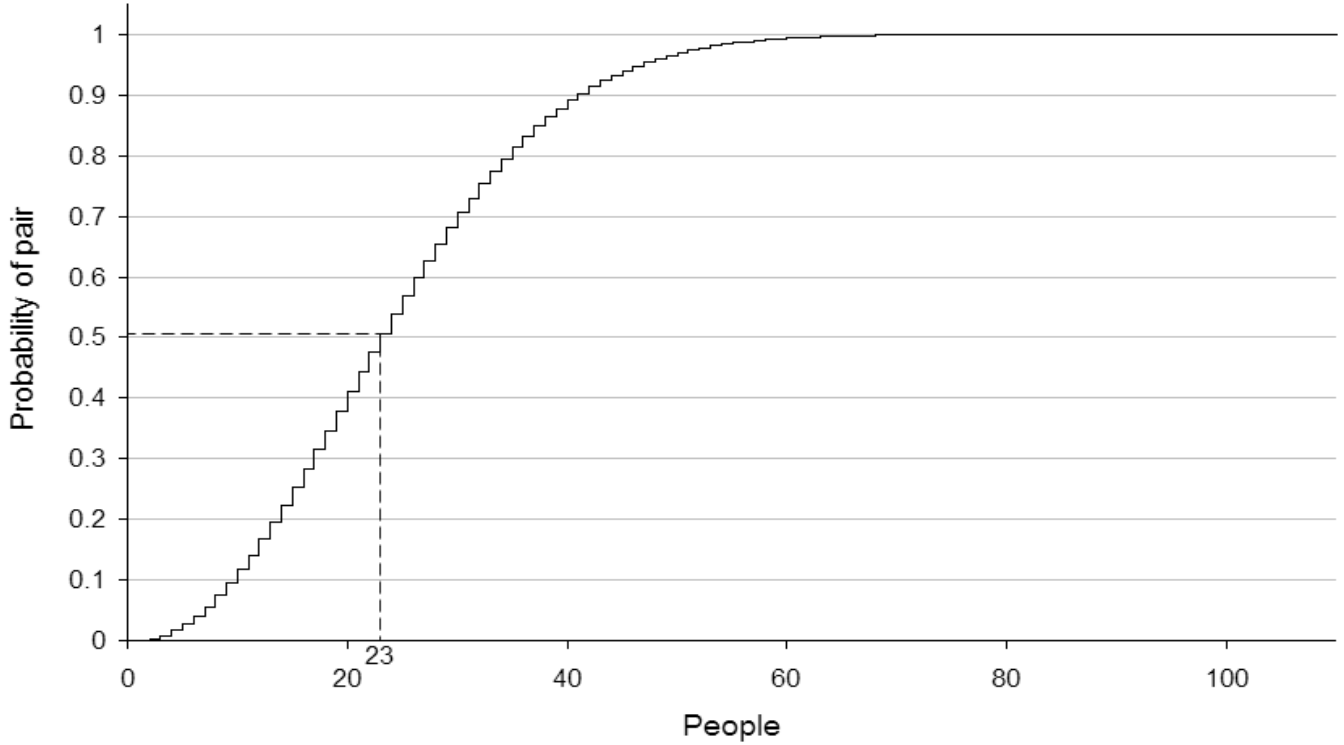
    Probabilidad de que ninguno coincida

$$p = \frac{365}{365} \cdot \frac{364}{365} \cdot \frac{363}{365} \cdots \frac{365 - n + 1}{365}$$

$$p = \begin{cases} \frac{365!}{365^n (365-n)!}, & 0 \leq n \leq 365 \\ 0, & 365 < n \end{cases}$$

    P("dos cumpleaños coincidan") > 0,5 para n= 23

Funciones Resumen Criptográficas



Funciones Resumen Criptográficas

- Por el contrario, la probabilidad de que, de un conjunto de n personas, al menos 1 cumpla años un día concreto, es mucho menor

  P("cumpla años el mismo día que yo") = 1 – P("nadie cumpla cuando yo")

$$1 - \left(\frac{364}{365}\right)^n$$

  P("cumpla años el mismo día que yo") > 0,5 para n= 253

Funciones Resumen Criptográficas

- Ataque del cumpleaños (e.g. en procedimiento de firma digital)
  - Objetivo: generar un par de mensajes M' y M'' que colisionen, y emplearlos de forma fraudulenta
  - Procedimiento
    - El atacante genera un conjunto C de m mensajes similares a un mensaje M dado (p. ej. mensaje a firmar por la víctima)
    - El atacante genera un conjunto Cf de m mensajes similares al mensaje M, pero con modificaciones que benefician al atacante
    - El tamaño de los conjuntos (m) debe ser $\geq$ n/2, siendo n la longitud del resumen generado por la función resumen dada
    - Con probabilidad 50%, y tiempo computacional n/2, se encuentra una pareja

$$M' \in C, M'' \in Cf \,/\, H(M') = H(M'')$$

  - El atacante engaña a la víctima para usar el mensaje M', y posteriormente lo sustituye por el mensaje M'' (fraudulento)