Adivinando passwords. Una propuesta para su búsqueda eficiente

Alejandro Mor Michael

9 de Julio de 2019

Índice

Funciones resumen

Toman como entrada mensajes de longitud arbitraria, produciendo como salida cadenas de longitud fija. Dicha salida es conocida como valor resumen o valor *hash*. Tomando una cadena de entrada x y una función resumen H:

$$H(x) = \text{cadena de longitud fija } y$$

Colisiones para funciones resumen

Una función resumen h cumple que, para una cadena binaria de salida x de longitud igual a n bits, la probabilidad de que un mensaje de entrada m aleatorio resulte en un valor *hash* igual que x es de 2^{-n}.

$$x = \text{'112233'},\ H = \text{CRC-32}$$

$$H(x) = 3570655599 = y$$

$$\text{Probabilidad}[H(m) = y] = 2^{-32}$$

Proceso

Obtener un mensaje de entrada p.

Generar su valor resumen $h = H(p)$.

Reconstruir el valor resumen h en un mensaje de entrada válido, según el contexto.

Función de reconstrucción

Toma como entrada un valor resumen y lo reconstruye en un mensaje válido según el contexto de la implementación.

$$R(h) = p'$$

Tablas del arco iris

$$
\begin{array}{ccccccccccc}
p_{1_1} & \xrightarrow{h} & h_{1_1} & \xrightarrow{r} & p_{1_2} & \xrightarrow{h} & \dots & \xrightarrow{r} & p_{1_t} & \xrightarrow{h} & h(p_{1_t}) \\
p_{2_1} & \xrightarrow{h} & h_{2_1} & \xrightarrow{r} & p_{2_2} & \xrightarrow{h} & \dots & \xrightarrow{r} & p_{2_t} & \xrightarrow{h} & h(p_{2_t}) \\
p_{3_1} & \xrightarrow{h} & h_{3_1} & \xrightarrow{r} & p_{3_2} & \xrightarrow{h} & \dots & \xrightarrow{r} & p_{3_t} & \xrightarrow{h} & h(p_{3_t}) \\
 & & & & \vdots & & & & \vdots & & \\
p_{m_1} & \xrightarrow{h} & h_{m_1} & \xrightarrow{r} & p_{m_2} & \xrightarrow{h} & \dots & \xrightarrow{r} & p_{m_t} & \xrightarrow{h} & h(p_{m_t})
\end{array}
$$

Almacenar el mensaje inicial y el resumen final de cada fila

Búsqueda de colisiones con una tabla del arco iris

Teniendo $\{p_{1_1}, h(p_{1_t})\}, \{p_{2_1}, h(p_{2_t})\}, \{p_{3_1}, h(p_{3_t})\}, \ldots, \{p_{m_1}, h(p_{m_t})\}$

Búsqueda de colisiones con una tabla del arco iris

Teniendo $\{p_{1_1}, h(p_{1_t})\}, \{p_{2_1}, h(p_{2_t})\}, \{p_{3_1}, h(p_{3_t})\}, \ldots, \{p_{m_1}, h(p_{m_t})\}$

1. Comprobar las últimas entradas de la tabla. Detener la búsqueda con éxito si se encuentra la colisión en una de ellas.

Búsqueda de colisiones con una tabla del arco iris

Teniendo $\{p_{1_1}, h(p_{1_t})\}, \{p_{2_1}, h(p_{2_t})\}, \{p_{3_1}, h(p_{3_t})\}, \ldots, \{p_{m_1}, h(p_{m_t})\}$

1. Comprobar las últimas entradas de la tabla. Detener la búsqueda con éxito si se encuentra la colisión en una de ellas.
2. Aplicar la función de reconstrucción seguida de la función resumen desde las últimas entradas

Búsqueda de colisiones con una tabla del arco iris

Teniendo $\{p_{1_1}, h(p_{1_t})\}, \{p_{2_1}, h(p_{2_t})\}, \{p_{3_1}, h(p_{3_t})\}, \ldots, \{p_{m_1}, h(p_{m_t})\}$

1. Comprobar las últimas entradas de la tabla. Detener la búsqueda con éxito si se encuentra la colisión en una de ellas.
2. Aplicar la función de reconstrucción seguida de la función resumen desde las últimas entradas
3. Repetir t veces por cada fila hasta encontrar (o no) la colisión.

Implementación desarrollada

Implementación desarrollada

Objetivo: adivinar las contraseñas de acceso a un sistema operativo.

Implementación desarrollada

Objetivo: adivinar las contraseñas de acceso a un sistema operativo.

Dominio de colisiones: contraseñas de seis dígitos, desde '000000' hasta '999999'.

Implementación desarrollada

Objetivo: adivinar las contraseñas de acceso a un sistema operativo.

Dominio de colisiones: contraseñas de seis dígitos, desde '000000' hasta '999999'.

Función resumen atacada: CRC-32, la cual genera valores resumen de 32 bits de longitud.

Función de reconstrucción **R1**

$$p = \text{'112233'} \xrightarrow{CRC-32} h = \text{'3570}\underline{655599}\text{'} \xrightarrow{\mathbf{R1}} r = \text{'655599'}$$

t	1												100 %
	2											91.9 %	
	4										90.6 %		
	5									90.6 %			
	10								95.2 %				
	20							96.1 %					
	40						97.1 %						
	50					97.4 %		99.3 %	99.6 %		99.6 %	100 %	100 %
	100					96.4 %	98.0 %	98.4 %	99.0 %	99.8 %		100 %	100 %	100 %
	200				95.2 %								
	400			91.8 %									
	500		84.1 %										
	1000	63.5 %											
		1000	2000	2500	5000	10000	20000	25000	50000	100000	200000	250000	500000	1000000
		m												

Tabla: Porcentajes de éxito para las tablas empleando **R1**

Función de reconstrucción **R2**

$$p = \text{'112233'} \xrightarrow{CRC-32} h_d = \text{'3570655}\underline{\text{599}}\text{'},\ h_h = \text{'}\cancel{\text{D}}\text{4D3E16F'} \xrightarrow{\textbf{R2}}$$

$$\xrightarrow{\textbf{R2}} r = \text{'939165'}$$

t	1												100 %
	2											91.1 %	
	4										89.4 %		
	5									88.8 %			
	10								93.0 %				
	20							95.5 %					
	40						97.1 %						
	50					97.1 %		98.8 %	99.7 %		100 %	100 %	100 %
	100				97.1 %	98.8 %	99.2 %	99.6 %	99.9 %		99.9 %	99.9 %	100 %
	200			94.3 %									
	400		82.1 %										
	500	74.9 %											
	1000	57.1 %											
		1000	2000	2500	5000	10000	20000	25000	50000	100000	200000	250000	500000	1000000
		m												

Tabla: Porcentajes de éxito para las tablas empleando **R2**

t	1												100 %
	2											91.9 %	
	4										90.6 %		
	5									90.6 %			
	10								95.2 %				
	20							96.1 %					
	40						97.1 %						
	50					97.4 %		99.3 %	99.6 %		99.6 %	100 %	100 %
	100					96.4 %	98.0 %	98.4 %	99.0 %	99.8 %		100 %	100 %	100 %
	200				95.2 %								
	400			91.8 %									
	500		84.1 %										
	1000	63.5 %											
		1000	2000	2500	5000	10000	20000	25000	50000	100000	200000	250000	500000	1000000
		m												

Tabla: Porcentajes de éxito para las tablas empleando **R1**

Combinando funciones de reconstrucción

Alternancia de funciones de reconstrucción

$$p_{1_c} \xrightarrow{CRC-32} h_{1_c} \xrightarrow{\mathbf{R1}} r_{1_c} \xrightarrow{CRC-32} h_{2_c} \xrightarrow{\mathbf{R2}} r_{2_c} \xrightarrow{CRC-32} h_{3_c} \ldots$$

	1												100 %
	2											97.0 %	
	4										96.2 %		
	5									97.0 %			
	10								98.3 %				
	20							99.5 %					
t	40						99.7 %						
	50					99.6 %		100 %	100 %		100 %	100 %	100 %
	100					99.9 %	99.9 %	99.9 %	100 %	100 %		100 %	100 %	100 %
	200				99.9 %								
	400			98.1 %									
	500		95.4 %										
	1000	63.1 %											
		1000	2000	2500	5000	10000	20000	25000	50000	100000	200000	250000	500000	1000000
		m												

Tabla: Porcentajes de éxito para las tablas empleando la alternancia de las

Combinando funciones de reconstrucción

Combinación mediante un patrón

$$p_{1_c} \xrightarrow{CRC-32} h_{1_c} \xrightarrow{\mathbf{R1}} r_{1_c} \xrightarrow{CRC-32} h_{2_c} \xrightarrow{\mathbf{R2}} r_{2_c} \xrightarrow{CRC-32} h_{3_c} \xrightarrow{\mathbf{R2}} r_{3_c} \ldots$$

t	1												100 %
	2											98.1 %	
	4										98.9 %		
	5									99.4 %			
	10								99.8 %				
	20							100 %					
	40						100 %						
	50					100 %		100 %	100 %		100 %	100 %	100 %
	100				100 %	100 %	100 %	100 %	100 %		100 %	100 %	100 %
	200			100 %									
	400		99.5 %										
	500	99.3 %											
	1000	80.5 %											
		1000	2000	2500	5000	10000	20000	25000	50000	100000	200000	250000	500000	1000000
		m												

Tabla: Porcentajes de éxito para las tablas empleando el patrón reducido

Combinando funciones de reconstrucción

Un patrón más extenso

$$p_{1_c} \xrightarrow{CRC-32} h_{1_c} \xrightarrow{\mathbf{R1}} r_{1_c} \xrightarrow{CRC-32} h_{2_c} \xrightarrow{\mathbf{R2}} r_{2_c} \xrightarrow{CRC-32} h_{3_c} \xrightarrow{\mathbf{R2}} r_{3_c} \ldots$$

$$\ldots \xrightarrow{CRC-32} h_{4_c} \xrightarrow{\mathbf{R1}} r_{4_c} \xrightarrow{CRC-32} h_{5_c} \xrightarrow{\mathbf{R1}} r_{5_c} \xrightarrow{CRC-32} h_{6_c} \xrightarrow{\mathbf{R2}} r_{6_c} \ldots$$

$$\ldots \xrightarrow{CRC-32} h_{7_c} \xrightarrow{\mathbf{R2}} r_{7_c} \xrightarrow{CRC-32} h_{8_c} \xrightarrow{\mathbf{R2}} r_{8_c} \xrightarrow{CRC-32} h_{9_c} \ldots$$

t														
	1													100 %
	2												98.5 %	
	4											99.9 %		
	5										100 %			
	10									100 %				
	20								100 %					
	40							100 %						
	50						100 %		100 %	100 %		100 %	100 %	100 %
	100					100 %	100 %	100 %	100 %	100 %		100 %	100 %	100 %
	200				100 %									
	400			100 %										
	500		100 %											
	1000	63.0 %												
		1000	2000	2500	5000	10000	20000	25000	50000	100000	200000	250000	500000	1000000
		m												

Tabla: Porcentajes de éxito para las tablas empleando el patrón extenso

t	50	14.6 %	27.9 %	39.0 %	46.6 %	59.4 %	67.8 %	73.8 %	87.4 %	94.7 %
	100	22.1 %	41.9 %	55.8 %	65.8 %	77.2 %	81.6 %	85.9 %	92.3 %	96.4 %
	200	24.0 %	46.2 %	61.2 %	71.3 %	81.6 %	87.7 %	91.0 %	95.2 %	97.7 %
	400	23.7 %	45.0 %	60.5 %	69.9 %	82.1 %	88.5 %	91.8 %	96.0 %	98.3 %
	500	23.5 %	44.3 %	58.2 %	67.0 %	78.3 %	84.1 %	87.9 %	94.6 %	97.9 %
	1000	22.7 %	42.4 %	55.1 %	63.5 %	72.7 %	77.6 %	80.7 %	86.2 %	89.3 %
		250	500	750	1000	1500	2000	2500	5000	10000
		m								

Tabla: Porcentajes de éxito para las tablas que emplean tan sólo la función de reconstrucción **R1**

t	50	15.9 %	29.7 %	40.2 %	48.8 %	61.2 %	69.7 %	76.0 %	88.6 %	94.4 %
	100	22.2 %	41.7 %	56.2 %	66.2 %	76.9 %	82.8 %	87.0 %	93.4 %	97.1 %
	200	24.1 %	44.4 %	61.2 %	70.3 %	82.1 %	87.3 %	90.2 %	94.3 %	97.4 %
	400	22.1 %	41.0 %	57.0 %	64.7 %	74.6 %	78.9 %	82.1 %	90.1 %	93.8 %
	500	21.1 %	38.2 %	53.1 %	60.8 %	69.3 %	74.9 %	77.6 %	81.9 %	85.3 %
	1000	20.5 %	35.6 %	48.0 %	57.1 %	69.4 %	73.8 %	76.1 %	79.3 %	79.6 %
		250	500	750	1000	1500	2000	2500	5000	10000
		m								

Tabla: Porcentajes de éxito para las tablas que emplean tan sólo la función de reconstrucción **R2**

t	50	15.5 %	29.4 %	42.5 %	51.5 %	65.8 %	75.9 %	82.4 %	96.1 %	98.8 %
	100	23.8 %	45.4 %	63.3 %	74.5 %	88.3 %	93.8 %	96.2 %	99.4 %	99.9 %
	200	24.4 %	47.3 %	68.2 %	83.6 %	94.4 %	97.2 %	98.1 %	99.9 %	100 %
	400	24.4 %	45.3 %	64.1 %	77.3 %	92.6 %	96.8 %	98.1 %	99.8 %	100 %
	500	24.2 %	44.5 %	62.3 %	74.4 %	89.9 %	95.4 %	97.7 %	99.6 %	100 %
	1000	22.6 %	39.3 %	53.1 %	63.1 %	77.3 %	85.9 %	89.9 %	95.4 %	97.9 %
		250	500	750	1000	1500	2000	2500	5000	10000
		m								

Tabla: Porcentajes de éxito para las tablas que emplean la alternancia de funciones de reconstrucción

t	50	17.0 %	32.0 %	45.6 %	55.5 %	71.0 %	82.0 %	88.2 %	98.0 %	99.8 %
	100	22.9 %	44.7 %	63.6 %	76.7 %	91.8 %	96.5 %	98.5 %	100 %	100 %
	200	24.4 %	48.0 %	69.4 %	85.8 %	97.3 %	99.3 %	99.7 %	100 %	100 %
	400	24.5 %	47.8 %	70.9 %	88.0 %	96.0 %	99.0 %	99.5 %	99.9 %	100 %
	500	24.6 %	48.0 %	70.5 %	88.0 %	97.2 %	99.3 %	99.7 %	100 %	100 %
	1000	23.6 %	43.9 %	62.7 %	80.5 %	95.6 %	99.1 %	99.8 %	99.9 %	100 %
		250	500	750	1000	1500	2000	2500	5000	10000
		m								

Tabla: Porcentajes de éxito para las tablas que emplean patrón reducido de funciones de reconstrucción

t	50	18.9 %	36.2 %	50.9 %	64.7 %	81.8 %	90.5 %	94.8 %	99.9 %	100 %
	100	23.6 %	47.0 %	67.0 %	83.1 %	96.9 %	99.9 %	100 %	100 %	100 %
	200	24.4 %	47.7 %	68.9 %	87.8 %	99.4 %	100 %	100 %	100 %	100 %
	400	24.0 %	45.9 %	65.3 %	83.1 %	99.9 %	100 %	100 %	100 %	100 %
	500	24.1 %	47.7 %	69.1 %	88.8 %	99.8 %	100 %	100 %	100 %	100 %
	1000	22.4 %	39.0 %	51.8 %	63.0 %	81.8 %	95.8 %	99.7 %	100 %	100 %
		250	500	750	1000	1500	2000	2500	5000	10000
		m								

Tabla: Porcentajes de éxito para las tablas que emplean patrón extenso de funciones de reconstrucción

Clasificación de configuraciones de tabla

1. Patrón extenso
2. Patrón reducido
3. Alternancia de funciones de reconstrucción
4. Uso exclusivo de **R1**
5. Uso exclusivo de **R2**

Conclusiones finales

- Mayor variedad en las funciones de reconstrucción brinda mejores resultados
- Tablas de tres ordenes de magnitud menores que el dominio
- Resultados escalables a contextos más complejos

Muchas gracias por su atención

Especial agradecimiento a Damián López