Estudio y comparación de métodos de tokenización

Daniel Ayala Zamorano, Laura Natalia Borbolla Palacios, Ricardo Quezada Figueroa

Escuela Superior de Cómputo, Instituto Politécnico Nacional
daniel@ejemplo.com, laura@ejemplo.com, qf7.ricardo@gmail.com

Resumen La tokenización consiste en el reemplazo de información sensible por valores sustitutos, llamados tokens, en donde el camino de regreso, del token a la información sensible, no es factible. En los últimos años este proceso se ha vuelto muy popular entre los comercios en línea, pues les permite descargar parte de las responsabilidades de seguridad adquiridas al manejar números de tarjetas de crédito en un tercero, proveedor de servicios de tokenización. Lamentablemente, existe una gran cantidad de desinformación alrededor de cómo generar los tokens, principalmente producida por las estrategias publicitarias de las empresas tokenizadoras, en donde cada una intenta convencer al comprador de que su sistema es el mejor, sin explicar realmente qué es lo que hacen para generar tokens. Uno de los mensajes más comúnes entre la publicidad es que la criptografía y la tokenización son cosas distintas, y la segunda es mucho más segura. En este trabajo se explica a detalle en qué consiste la tokenización y cuál es su relación con la criptografía; se revisan y comparan los desempeños de los métodos más comunes para tokenizar; para terminar se concluye con una discusión alrededor de las ventajas y desventajas de cada uno.

1. Introducción

2. Preliminares

3. Clasificación de los algoritmos tokenizadores

4. Métodos reversibles

4.1. FFX (*Format-preserving Feistel-based Encryption*)

4.2. BPS (Brier, Peyrin, Stern)

5. Métodos irreversibles

5.1. TKR

5.2. RHA (*Reversile Hybrid Algorithm*)

5.3. UTO (*Updatable Tokenization*)

5.4. Basados en DRBG (*Deterministic Random Bit Generator*)

6. Resultados de comparaciones de desempeño

Todos los resultados presentados en esta sección se llevaron a cabo en una computadora con las siguientes características:

Procesador: Intel i5-7200U (2.5 GHz) de 4 núcleos.
Sistema operativo: Arch Linux, kernel 4.17.
Base de datos: MariaDB 10.1.
Compilador: GCC 8.1.1

Algoritmo	Tokenización (μs)	Detokenización (μs)
FFX	82	63
BPS	169	103
TKR	4341	374
AHR	4795	487
DRBG	2619	269

Tabla 1. Comparación de tiempos de tokenización.

7. Conlusiones

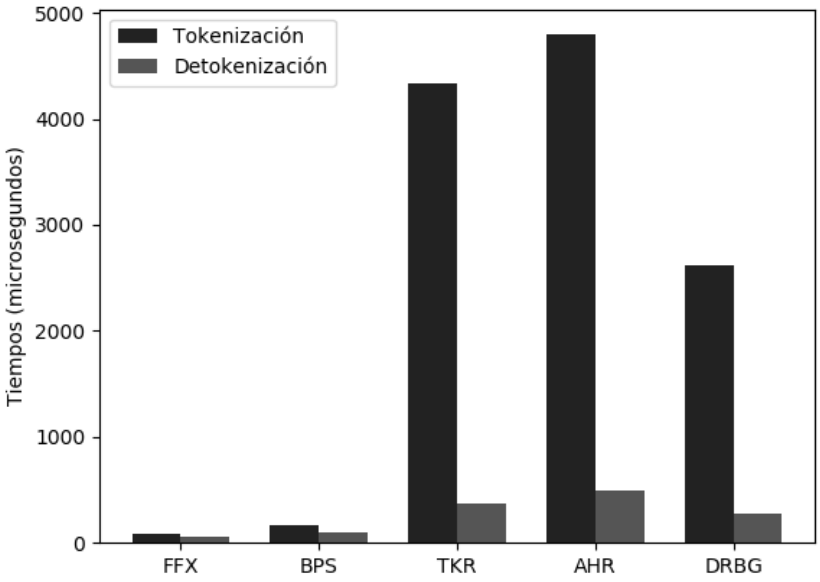

Figura 1. Tiempos de tokenización generales.