CCBHash (Compound Code Block Hash) para Análisis de Malware

Pablo Pérez, José Antonio Onieva

Network, Information and Computer Security (NICS) Lab

Universidad de Málaga

Gerardo Fernández

VirusTotal

Málaga

Santander, Octubre 2022

NICS

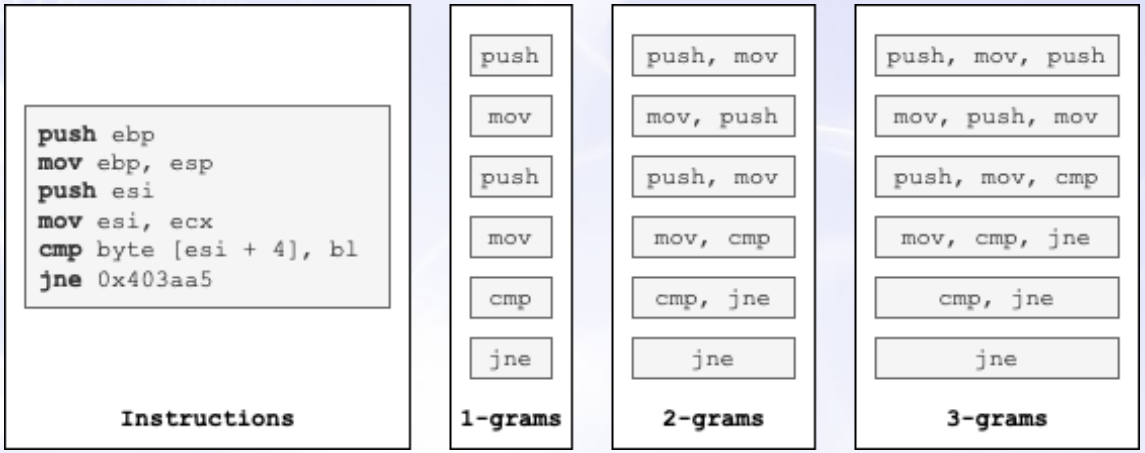
Buscadores de similitud

- Similitudes entre numerosos ficheros:
  - Estrategias basadas en n-grams
  - Estrategias basadas en grafos
  - Estrategias basadas en Fuzzy Hashing
  - …
- Similitudes entre funciones de dos ficheros:
  - BinDiff
  - Diaphora
- Similitudes entre numerosas funciones:
  - ?

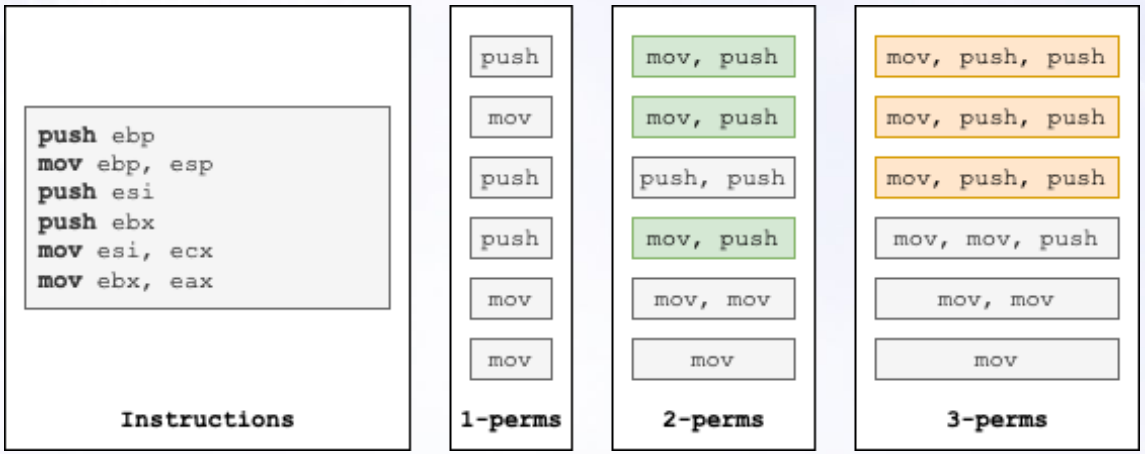
NICS

Similitudes entre numerosos ficheros

- N-grams



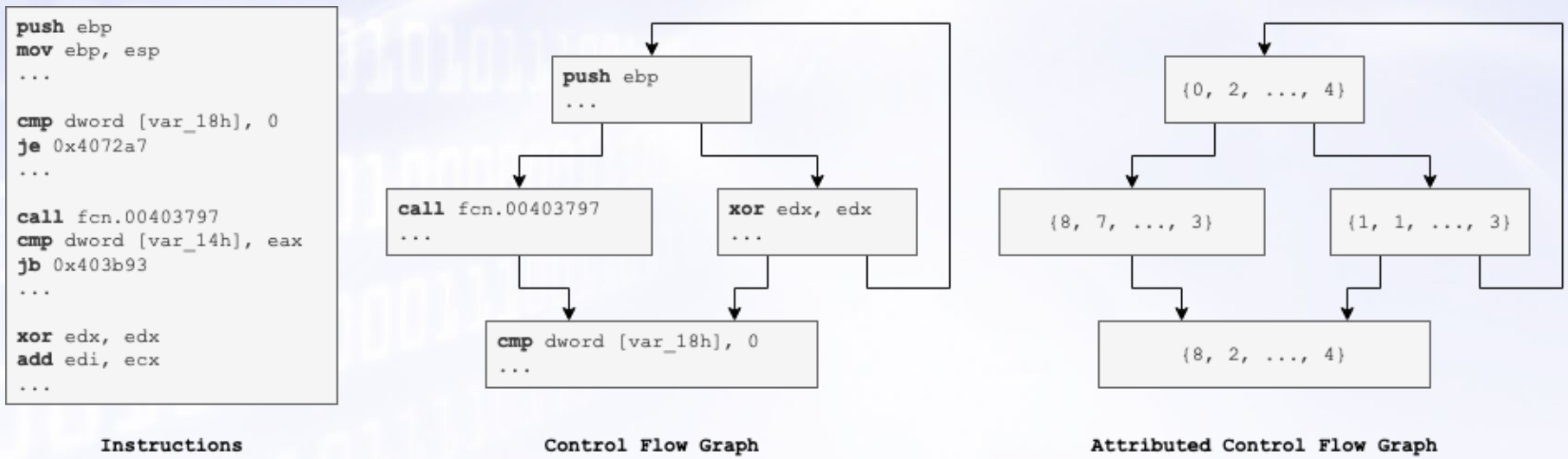
- N-perms



NICS

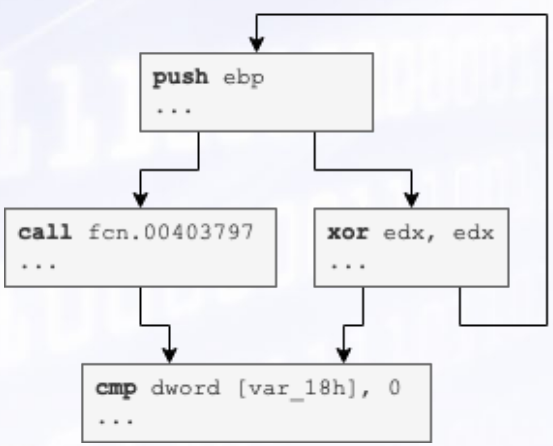
Similitudes entre numerosos ficheros

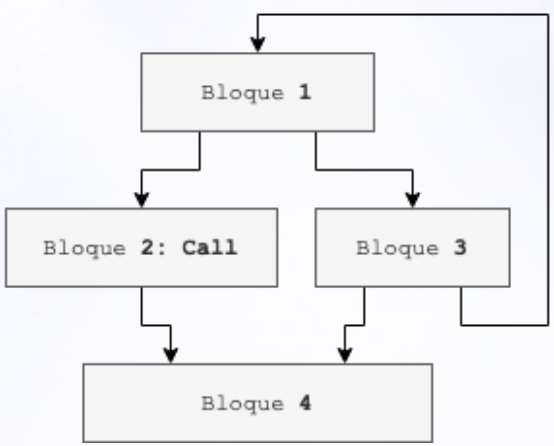
- Grafos
  - Callgraph
  - CFG (Control Flow Graph)
  - ACFG (Attributed CFG)

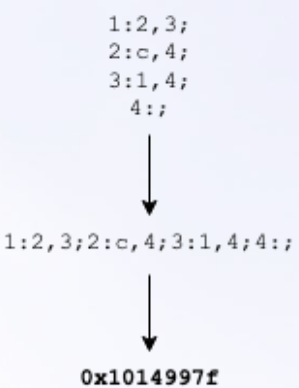


Similitudes entre numerosos ficheros

- Fuzzy Hashing
  - BBH (Block Based Hashing)
  - CTPH (Context Triggered Piecewise Hashing)
  - SIF (Statistically Improbable Features)
  - BBR (Block Based Rebuilding)
  - LSH (Locality Sensitive Hashing)
- Machoc y Machoke

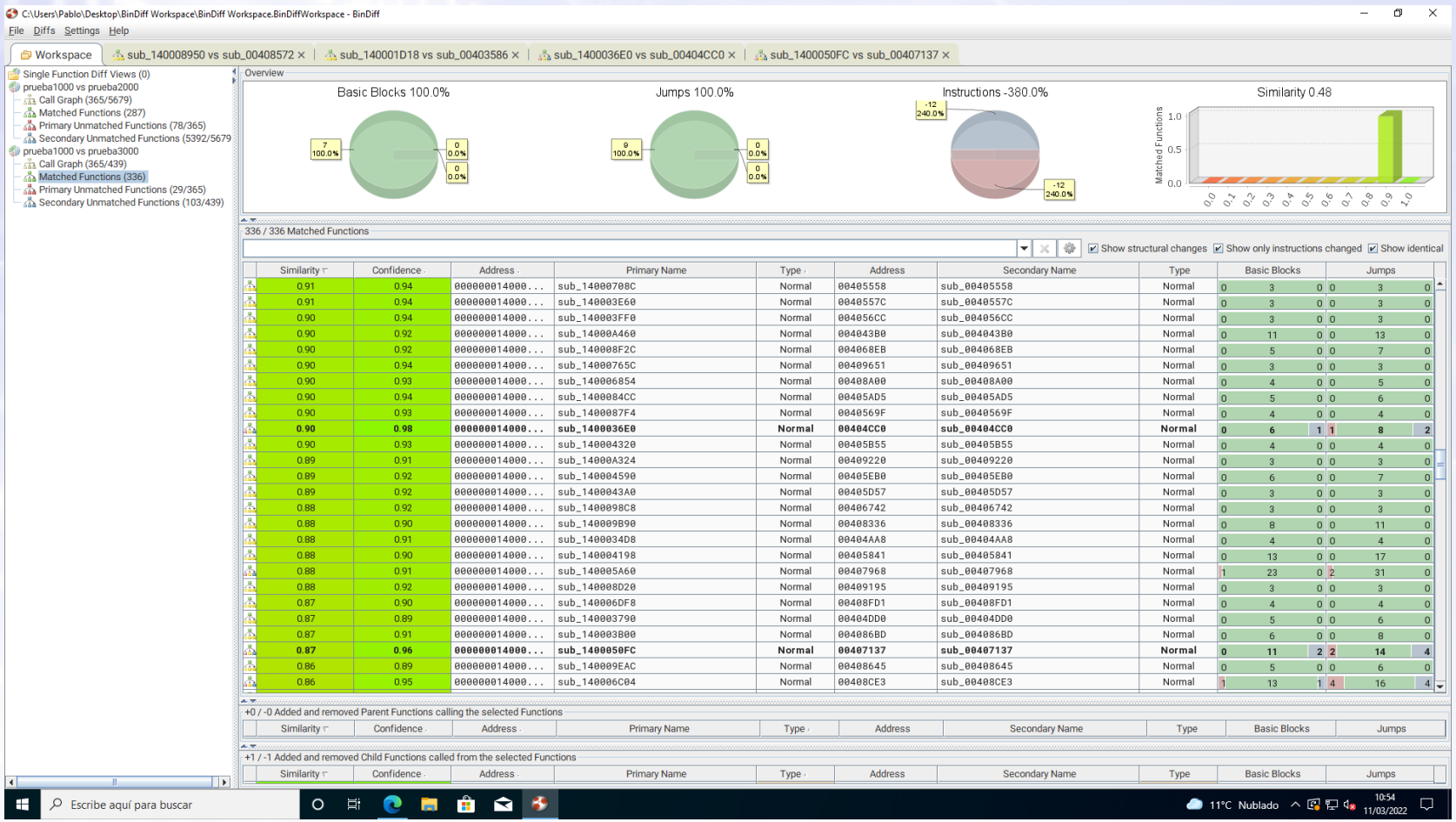






Similitudes entre funciones de dos ficheros

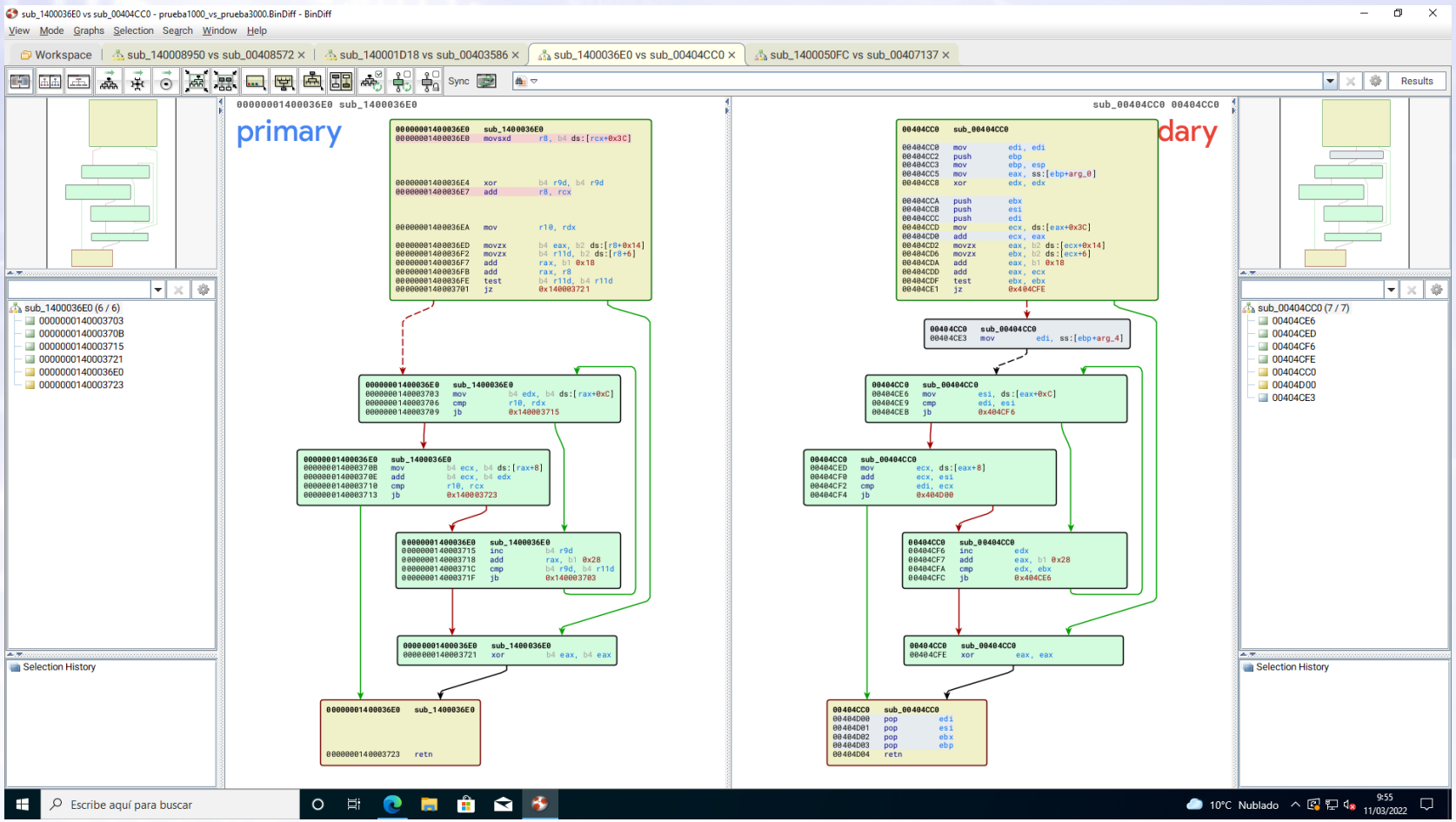
- BinDiff y Diaphora



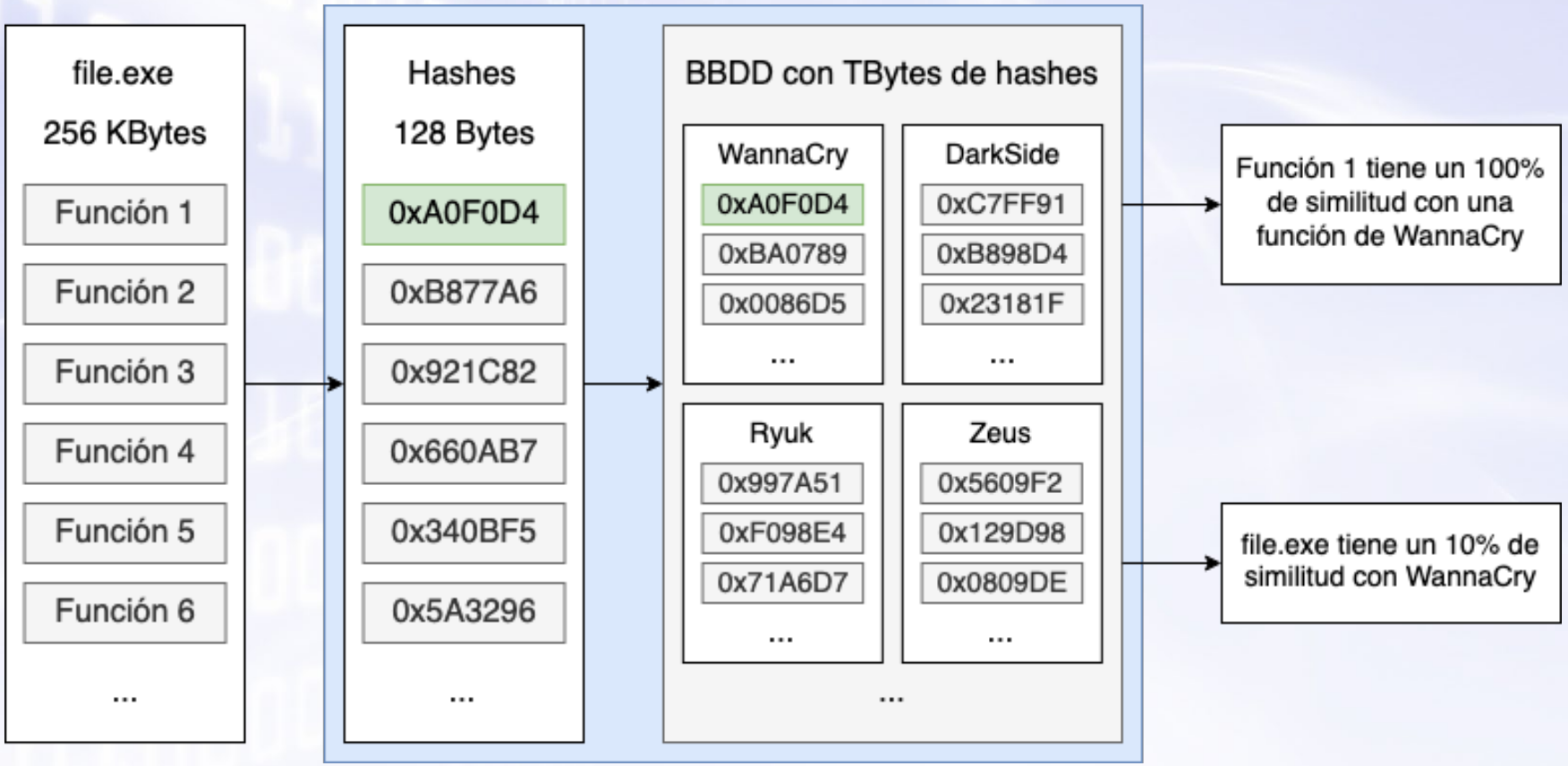
NICS

Similitudes entre funciones de dos ficheros

- BinDiff y Diaphora



CCBHash (Compound Code Block Hash)



NICS

Diseño de CCBHash

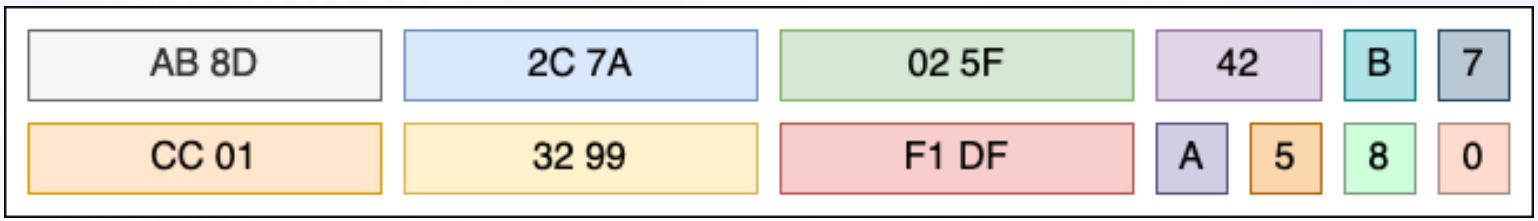
- Nombre de la función (2 bytes)
- CFG de la función (2 bytes)
- Callgraph de la función (2 bytes)
- Tipos de opcodes (2 bytes)
- Tipos de argumentos (2 bytes)
- Tipos de variables locales (2 bytes)

Se almacenan hashes (usando blake2b).

- Número de instrucciones (1 byte)
- Número de bloques (4 bits)
- Número de entradas (4 bits)
- Número de salidas (4 bits)
- Número de args + vars (4 bits)
- Complejidad ciclomática (4 bits)
- Tamaño de la pila (4 bits)

Se almacenan valores numéricos.

16 bytes



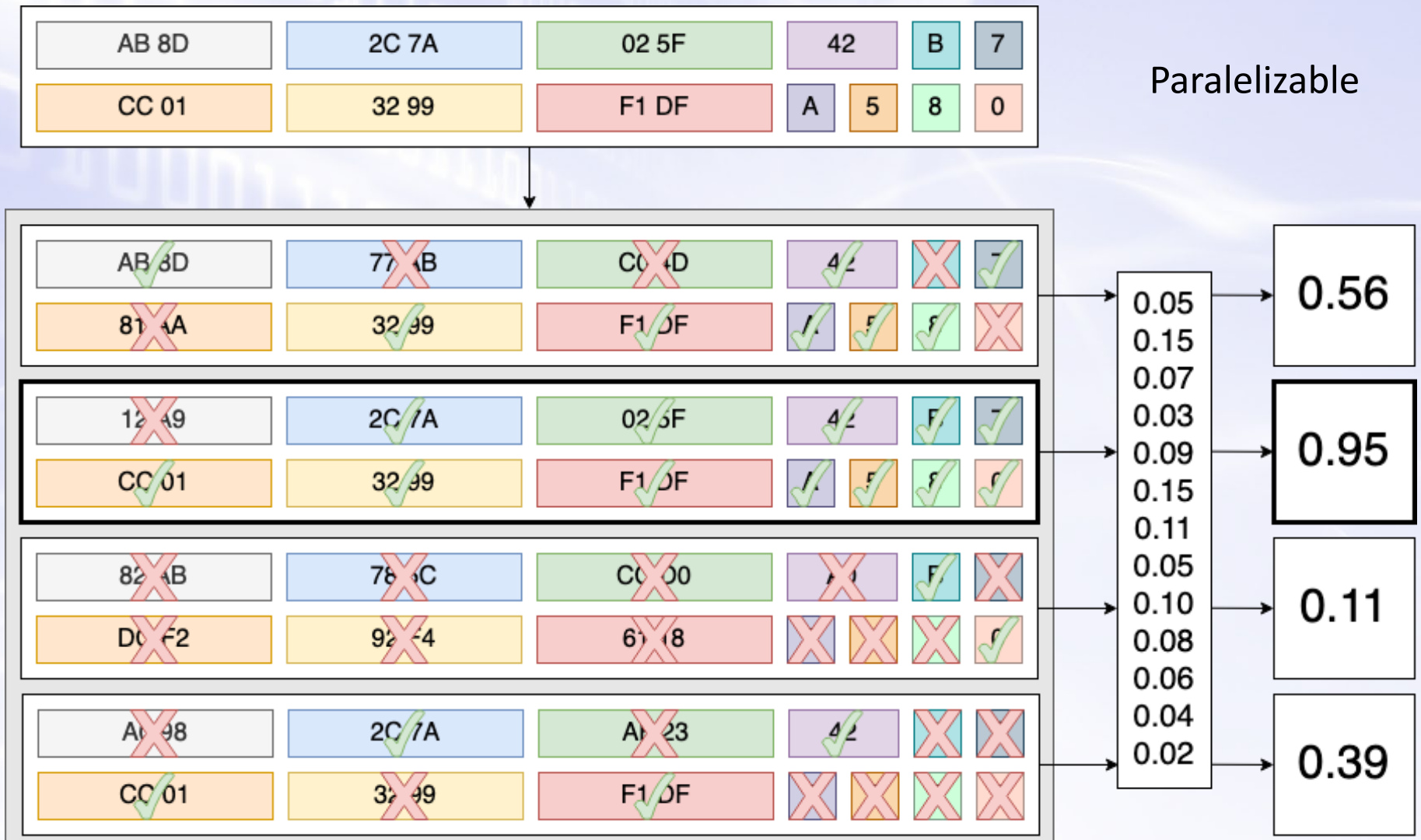
NICS

Diseño de CCBHash

Atributo	Puntuación
Tipos de opcodes	0.25
CFG de la función	0.16
Complejidad ciclomática	0.15
Número de salidas (outdegree)	0.09
Número de instrucciones	0.07
Callgraph de la función	0.06
Tipo de argumentos de la función	0.04
Tipo de variables locales	0.04
Cantidad de argumentos y variables locales	0.04
Número de bloques	0.04
Número de entradas (indegree)	0.03
Tamaño de la pila	0.02
Nombre de la función	0.01

NICS

Diseño de CCBHash



Paralelizable

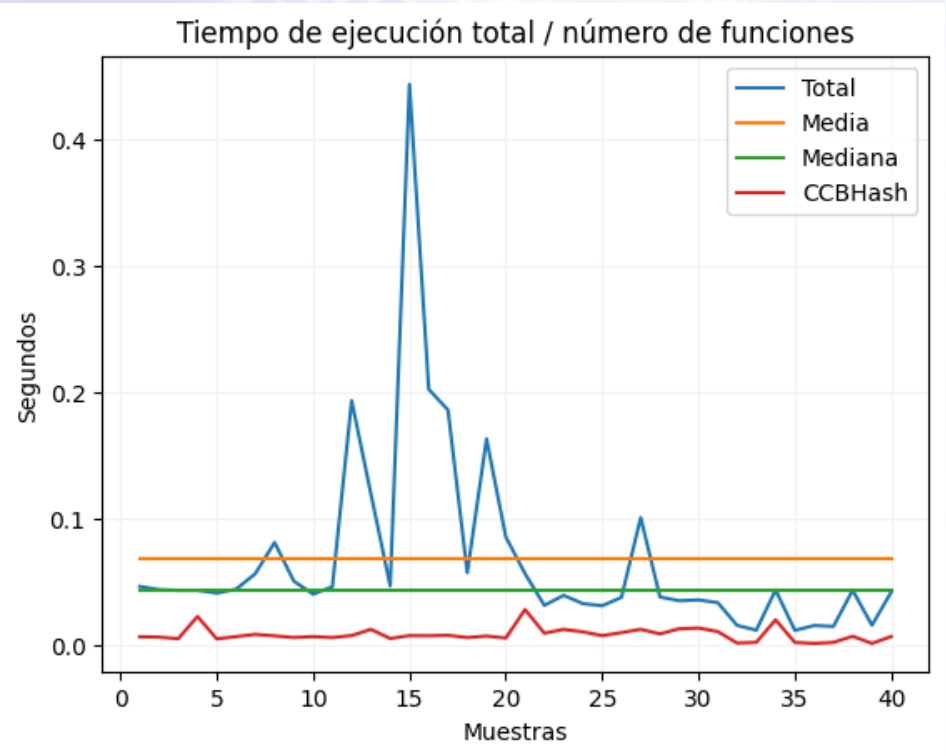
NICS

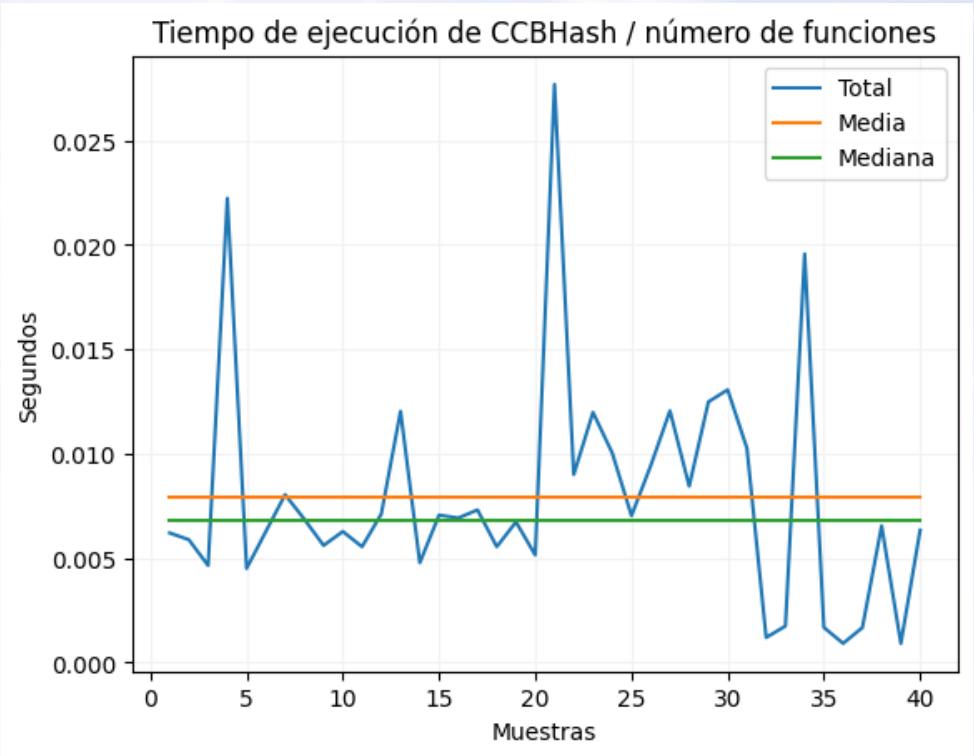
Implementación de CCBHash

NICS

Análisis de resultados

- Análisis de tiempos:





Análisis de resultados

- Análisis de precisión:

Familia de malware	Número medio de ficheros con dicha similitud					
	CCBHash			ssdeep		
	$> 90\,\%$	$> 75\,\%$	$> 50\,\%$	$> 90\,\%$	$> 75\,\%$	$> 50\,\%$
WannaCry	4.5	0	0	0.4	1.1	2.3
DarkSide	2	2	2.3	0.2	0.2	0.7
Ryuk	6	0	2*	0	0	0
Zeus	3.3	0	0	2.2	0	0
Total	4	0.5	1.1	0.7	0.3	0.8

Familia de malware	Número medio de ficheros con dicha similitud					
	CCBHash			TLSH		
	$> 90\,\%$	$> 75\,\%$	$> 50\,\%$	< 10	< 50	< 100
WannaCry	4.5	0	0	0	0	0.6
DarkSide	2	2	2.3	0.2	0.8	0.5
Ryuk	6	0	2*	0	0.2	1
Zeus	3.3	0	0	2	0	0
Total	4	0.5	1.1	0.6	0.3	0.5

NICS

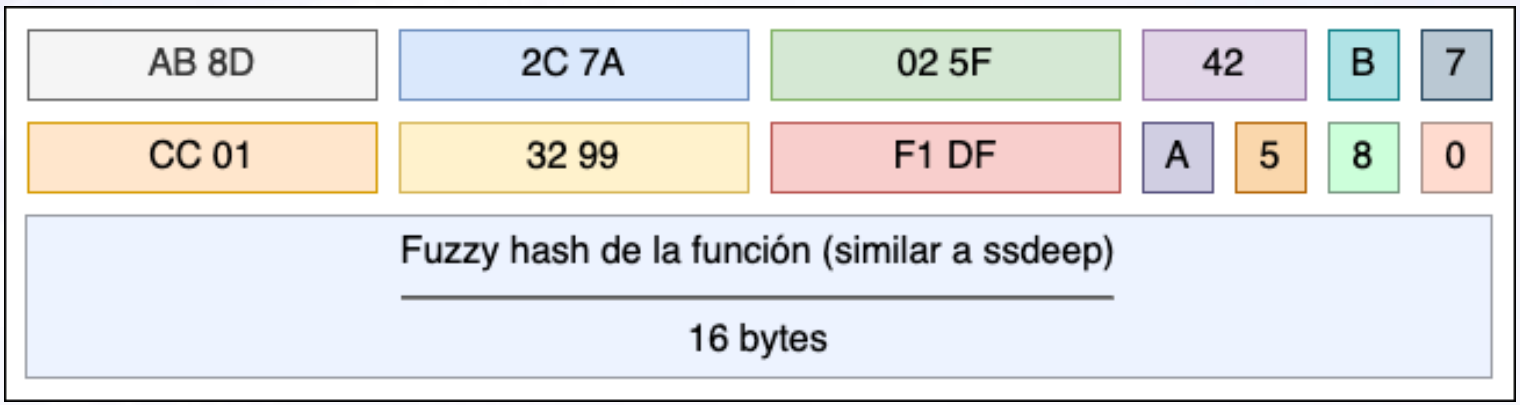
Conclusiones

- Solución a un problema actual y complejo
- Comparación rápida de funciones
- Comparación rápida de ejecutables
- Mejora a fuzzy hashes como ssdeep o TLSH

NICS

Trabajo futuro

- Mejora de atributos
- Optimización de filtros de funciones
- Plugin para IDA Pro
- Estudio de colisiones para atributos con hashes
- Análisis con dataset más amplio
- Versión extendida de 32 bytes



CCBHash (Compound Code Block Hash) para Análisis de Malware

Pablo Pérez, José Antonio Onieva

Network, Information and Computer Security (NICS) Lab

Universidad de Málaga

Gerardo Fernández

VirusTotal

Málaga

Santander, Octubre 2022

NICS