

CamBench - Cryptographic API Misuse Detection Tool Benchmark

Michael Schlichtig, Anna-Katharina Wickert,
Stefan Krüger, Eric Bodden, Mira Mezini



In-person MSR 2022



Zoom crypto misuse

63.7%
applications
 **≥ 1 crypto
misuse**



Photo by Chris Montgomery on Unsplash



**Comparison
crypto
misuse
detectors
difficult**

Crypto Benchmarks Meets Analyses

Preliminary Study

Analysis capabilities	Fix complexity					Real-world usages					Exploits							
Benchmark	A	AC	CC	CG	CO	DN	FD	FS	J	MA	MF	P	SB	SQ	V	X	Y	Z
<i>CryptoAPI-Bench</i> [6]	○	○	●	●	●	○	○	○	○	○	○	○	●	○	○	○	○	
<i>Braga Benchmark</i> [13]	○	○	○	○	○	○	○	●	○	○	○	○	○	●	●	●	●	
<i>Parametric Crypto Misuse Benchmark</i> [51]	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	
<i>MuBench</i> [8]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	
<i>Ghera Benchmark</i> [31]	●	●	○	○	○	●	●	○	○	●	●	○	○	○	○	○	○	
<i>OWASP Benchmark</i> [49]	○	○	○	○	○	○	○	●	●	○	○	●	○	●	○	○	●	

Crypto Benchmarks Meets Analyses

Preliminary Study

Benchmark	A	AC	CC	CG	CO	DN	FD	FS	J	MA	MF	P	SB	SQ	V	X	Y	Z
<i>CryptoAPI-Bench</i> [6]	○	○	●	●	●	○	○	○	○	○	○	○	●	○	○	○	○	○
<i>Braga Benchmark</i> [13]	○	○	○	○	○	○	○	●	○	○	○	○	●	●	●	●	●	○
<i>Parametric Crypto Misuse Benchmark</i> [51]	○	○	○	○	○	○	○	●	○	○	○	○	○	○	○	○	○	○
<i>MuBench</i> [8]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○
<i>Ghera Benchmark</i> [31]	●	●	○	○	○	●	●	○	○	●	●	○	○	○	○	○	○	○
<i>OWASP Benchmark</i> [49]	○	○	○	○	○	○	○	●	●	○	○	●	○	●	○	○	○	●

Only 11 % of the analyses were evaluated on more than one benchmark

Crypto Benchmarks Meets Analyses

Preliminary Study

Benchmark	A	AC	CC	CG	CO	DN	FD	FS	J	MA	MF	P	SB	SQ	V	X	Y	Z
<i>CryptoAPI-Bench</i> [6]	○														○	○	○	○
<i>Braga Benchmark</i> [13]	○														●	●	●	○
<i>Parametric Crypto Misuse Benchmark</i> [51]	○														○	○	○	○
<i>MuBench</i> [8]	○		○	○	○	○	○	○		○	○	○	○	○	○	○	○	○
<i>Ghera Benchmark</i> [31]	●	●	○	○	○	○	●	●	○	○	●	○	○	○	○	○	○	○
<i>OWASP Benchmark</i> [49]	○	○	○	○	○	○	○	○	●	○	○	●	○	●	○	○	○	●

Extend preliminary study with a systematic system literature review (Kitchenham)

Only 11 % of the analyses were evaluated on more than one benchmark

Design of CamBench



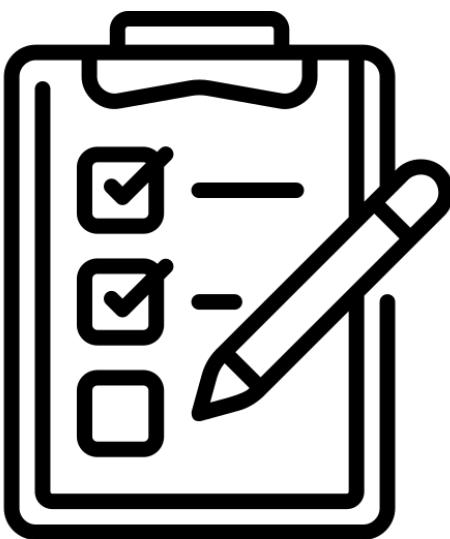
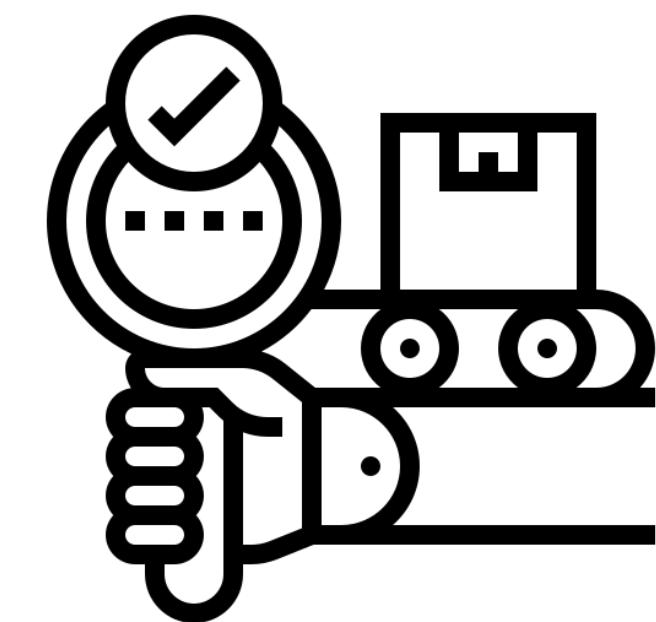
Real-world applications

- Mine Java applications on GitHub
- Compile projects with JCA
- Label usages manually



Analysis capabilities

- Synthetic test cases
- Cover flow-, context-, field-, object-, and path-sensitivity



Heuristic for crypto API coverage

Harness both benchmarks

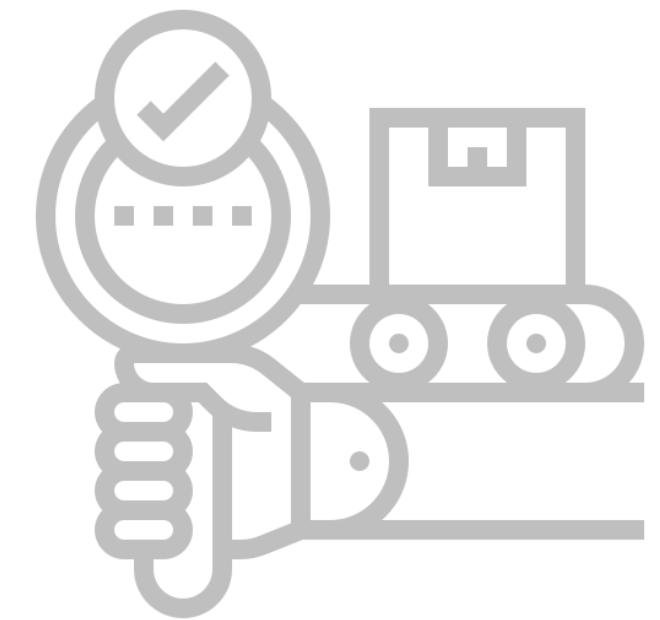
Design of CamBench



Real-world applications

- Mine Java applications on GitHub
- Compile projects with JCA
- Label usages manually

How do current misuse detection tools perform on CamBench?



How does CamBench compares to other crypto benchmarks?



**Comparison
crypto
misuse
detectors
difficult**

Crypto Benchmarks Meets Analyses

Preliminary Study

Benchmark	A	AC	CC	CG	CO	DN	FD	FS	J	MA	MF	P	SB	SQ	V	X	Y	Z
<i>CryptoAPI-Bench</i> [6]	○	○	●	●	●	○	○	○	○	○	○	○	●	○	○	○	○	
<i>Braga Benchmark</i> [13]	○	○	○	○	○	○	○	●	○	○	○	○	●	●	●	●	○	
<i>Parametric Crypto Misuse Benchmark</i> [51]	○	○	○	○	○	○	○	●	○	○	○	○	●	○	○	○	○	
<i>MuBench</i> [8]	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	○	
<i>Ghera Benchmark</i> [31]	●	●	○	○	○	○	○	●	●	○	●	●	●	○	○	○	○	
<i>OWASP Benchmark</i> [49]	○	○	○	○	○	○	○	●	○	●	●	●	●	●	●	●	●	

Only 11 % of the analyses were evaluated on more than one benchmark

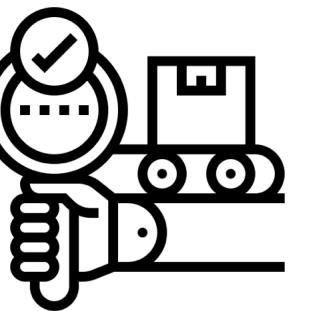
Our talk in one slide :)

Design of CamBench

- Real-world applications**
- Mine Java applications on GitHub
 - Compile projects with JCA
 - Label usages manually



- Analysis Capabilities**
- Synthetic test cases
 - Cover flow-, context-, field-, object-, and path-sensitivity



- Heuristic for crypto API coverage**
Harness both benchmarks



michael.schlichtig@uni-paderborn.de
@M_Schlichtig



wickert@cs.tu-darmstadt.de
@akwickert

