Malaware: the beginning of a behavioral malware detection

Estelle Hotellier[1], Elise Klein[2]
Dylan Marinho[2], Camille Moriot[3] and Guillaume Roumage[4]

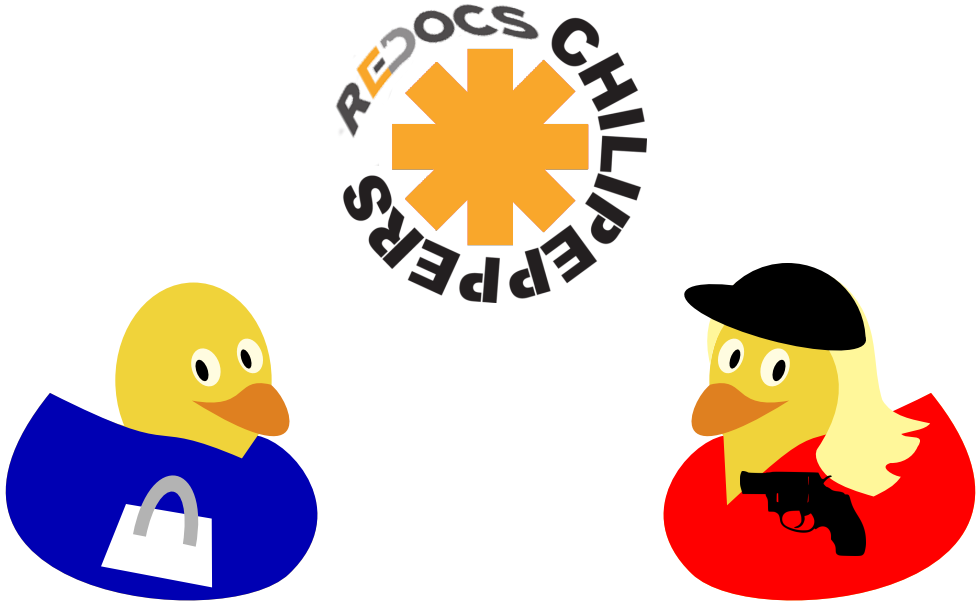
[1] Naval Group, Naval Cyber Laboratory (NCL), Inria, LIG, Grenoble INP, UGA, 38000 Grenoble, France
[2] Université de Lorraine, CNRS, Inria, LORIA, 54000 Nancy, France
[3] Univ Lyon, INSA Lyon, Inria, CITI, EA3720, 69621 Villeurbanne, France
[4] Université Paris-Saclay, CEA, List, 91120 Palaiseau, France

Research topic proposed and supervised by
Ludovic Robin, CyberDetect

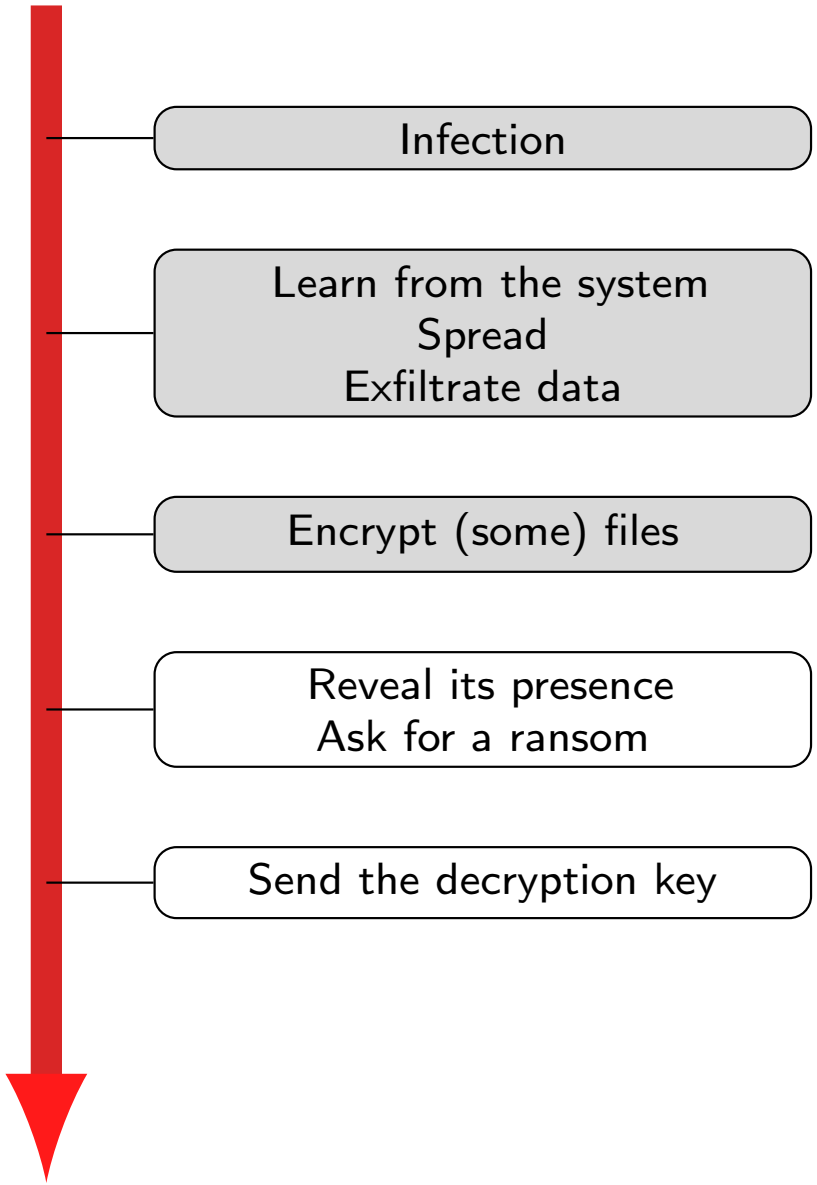
December 2nd 2022
Luminy

What is a ransomware?

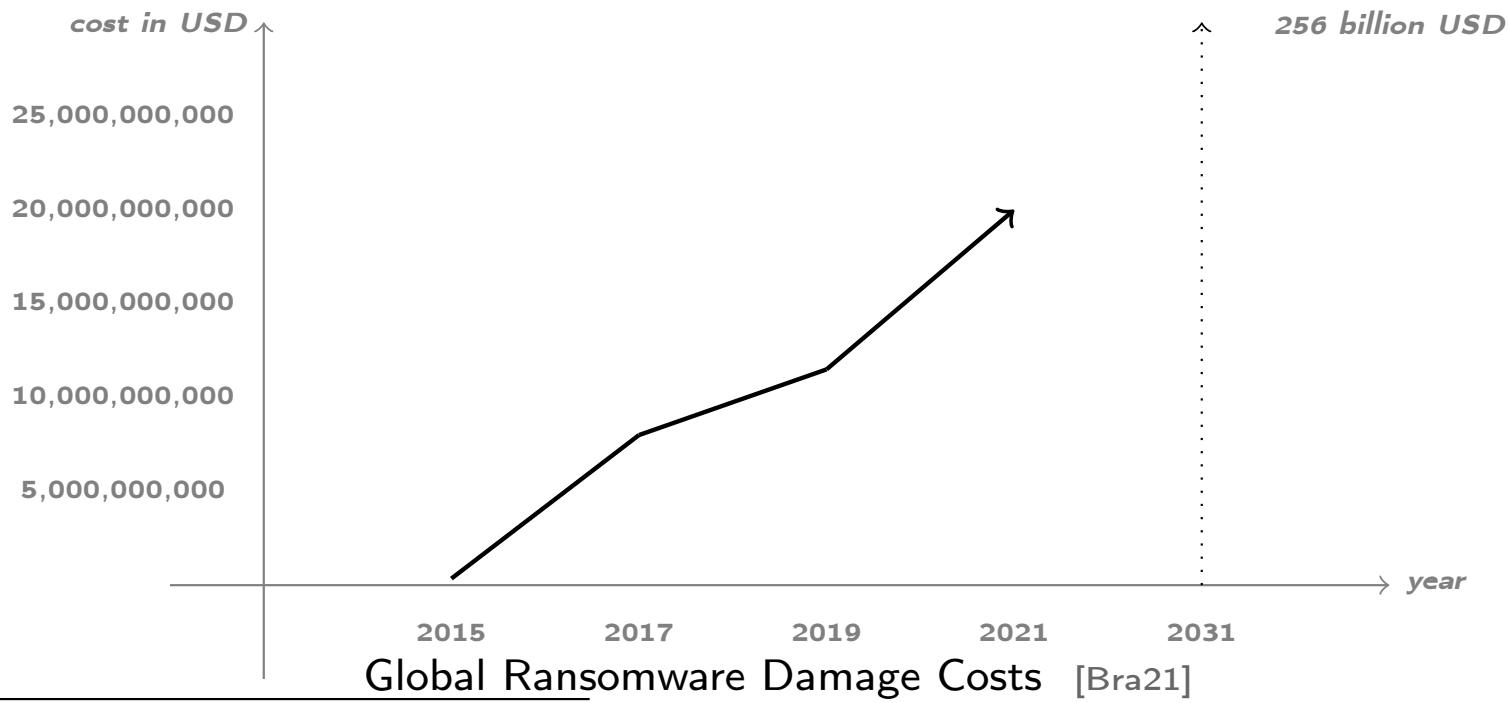
What is a ransomware?



How does a ransomware work?



A story of money



Global Ransomware Damage Costs [Bra21]

[Bra21] David Braue. Global Ransomware Damage Costs Predicted To Exceed 265 Billion USD By 2031. June 2021

A cat and mouse game

[Bio+18]

	Simple malware
No defense	√

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

A cat and mouse game

[Bio+18]

	Simple malware
No defense	√
Signature analysis	×

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

A duck and dodge game

[Bio+18]

	Simple malware	Small variations
No defense	√	√
Signature analysis	×	√

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

A duck and dodge game

[Bio+18]

	Simple malware	Small variations
No defense	√	√
Signature analysis	×	√
Dynamic analysis	×	×

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

A duck and duck game

[Bio+18]

	Simple malware	Small variations	Anti-sandboxing
No defense	√	√	√
Signature analysis	×	√	√
Dynamic analysis	×	×	√

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

A duck and duck game

[Bio+18]

	Simple malware	Small variations	Anti-sandboxing
No defense	√	√	√
Signature analysis	×	√	√
Dynamic analysis	×	×	√
Concolic analysis	×	×	×

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

Quaaaaack

[Bio+18]

	Simple malware	Small variations	Anti-sandboxing	Symbolic explosion
No defense	√	√	√	√
Signature analysis	×	√	√	√
Dynamic analysis	×	×	√	√
Concolic analysis	×	×	×	√

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

Quaaaaack

[Bio+18]

	Simple malware	Small variations	Anti-sandboxing	Symbolic explosion
No defense	√	√	√	√
Signature analysis	×	√	√	√
Dynamic analysis	×	×	√	√
Concolic analysis	×	×	×	√

Most of the widespread anti-malwares only uses signature analysis

[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

Quaaaaack

[Bio+18]

	Simple malware	Small variations	Anti-sandboxing	Symbolic explosion
No defense	√	√	√	√
Signature analysis	×	√	√	√
Dynamic analysis	×	×	√	√
Concolic analysis	×	×	×	√

Most of the widespread anti-malwares only uses signature analysis

But, you can imagine more advanced statistical analysis

e.g. checking the imported libraries

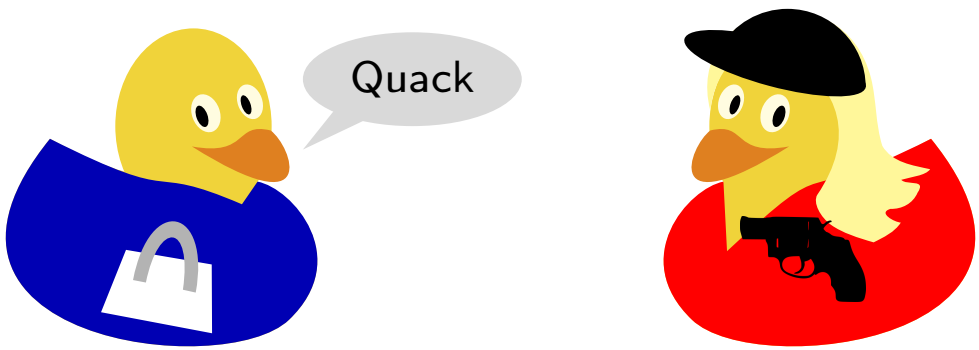
[Bio+18] Fabrizio Biondi et al. "Tutorial: an Overview of Malware Detection and Evasion Techniques". In: ISoLA 2018 - 8th International Symposium On Leveraging Applications of Formal Methods, Verification and Validation. Oct. 2018, pp. 1–23

An overview of malware detection techniques

Signature-based analysis

Classify binaries by looking at particular patterns in their code

"Have I already seen this binary?"



Behavioral-based analysis

Detect malware based on their behavior

"What does it want to do?"

Outline

Outline

How to detect if a file is encrypted ?

Measuring the byte's coherence in a file

According to the literrature, The Shannon method of entropy calculation is the most commonly-used technique when it comes to file encryption identification in crypto-ransomware detection techniques. [Pal+17][a] [DMB22][b]

encrypted data are similar to random data

[a] [Pal+17] Aurélien Palisse et al. "Data aware defense (DaD): towards a generic and practical ransomware countermeasure". In: Nordic Conference on Secure IT Systems. Springer. 2017, pp. 192–208

[b] [DMB22] Simon R Davies, Richard Macfarlane, and William J Buchanan. "Comparison of Entropy Calculation Methods for Ransomware Encrypted File Identification". In: Entropy 24.10 (2022), p. 1503

Entropy related functions

Shannon Entropy

$$H(X) = -\sum_{i=1}^{n} P(x_i) \log_2 P(x_i)$$

Where H is the entropy (in bits)
n is the number of bytes
$P(x_i)$ probability of byte i

Monte Carlo

$$E(X) \approx \frac{1}{N} \sum_{n=1}^{N} x_n$$

Where E is the result of the approximation
x_n is a randomly choosen value

Chi-quare

$$\chi^2 = \sum_{i=0}^{255} \frac{(O_i - E_i)^2}{E_i}$$

Where O_i is the observed value
E_i is the expected value

Arithmetic mean

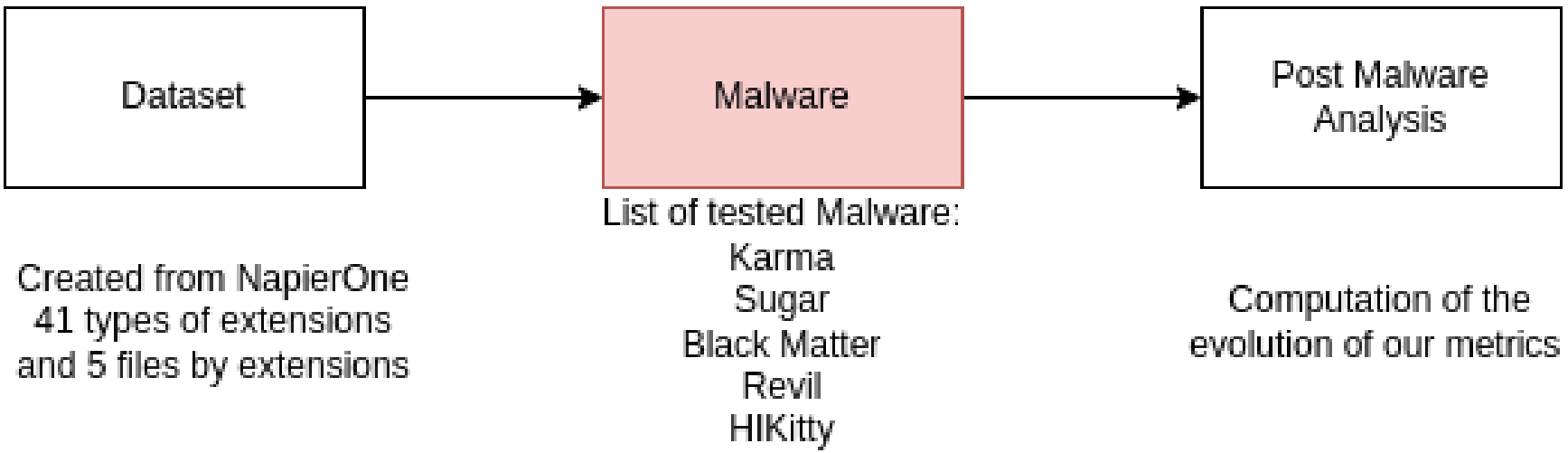
$$M = \frac{S}{T}$$

Where M is the arithmetic mean
S is the sum of the osbervations
T is the number of values

Serial Byte Correlation Coefficient

$$C = \frac{n(U_0U_1 + U_1U_2 + \cdots + U_{n-2}U_{n-1} + U_{n-1}U_0) - (U_0 + U_1 + \cdots + U_{n-1})^2}{n(U_0^2 + U_1^2 + \cdots + U_{n-1}^2) - (U_0 + U_1 + \cdots + U_{n-1})^2}$$

How the entropy-related analysis is conducted?



Which files are modified ?

Not Modified by Karma

Exe, DLL

Not Modified by REvil

Exe, DLL, ICS

Not Modified by Black Matter

Exe, DLL, ICS

Results by type of measure

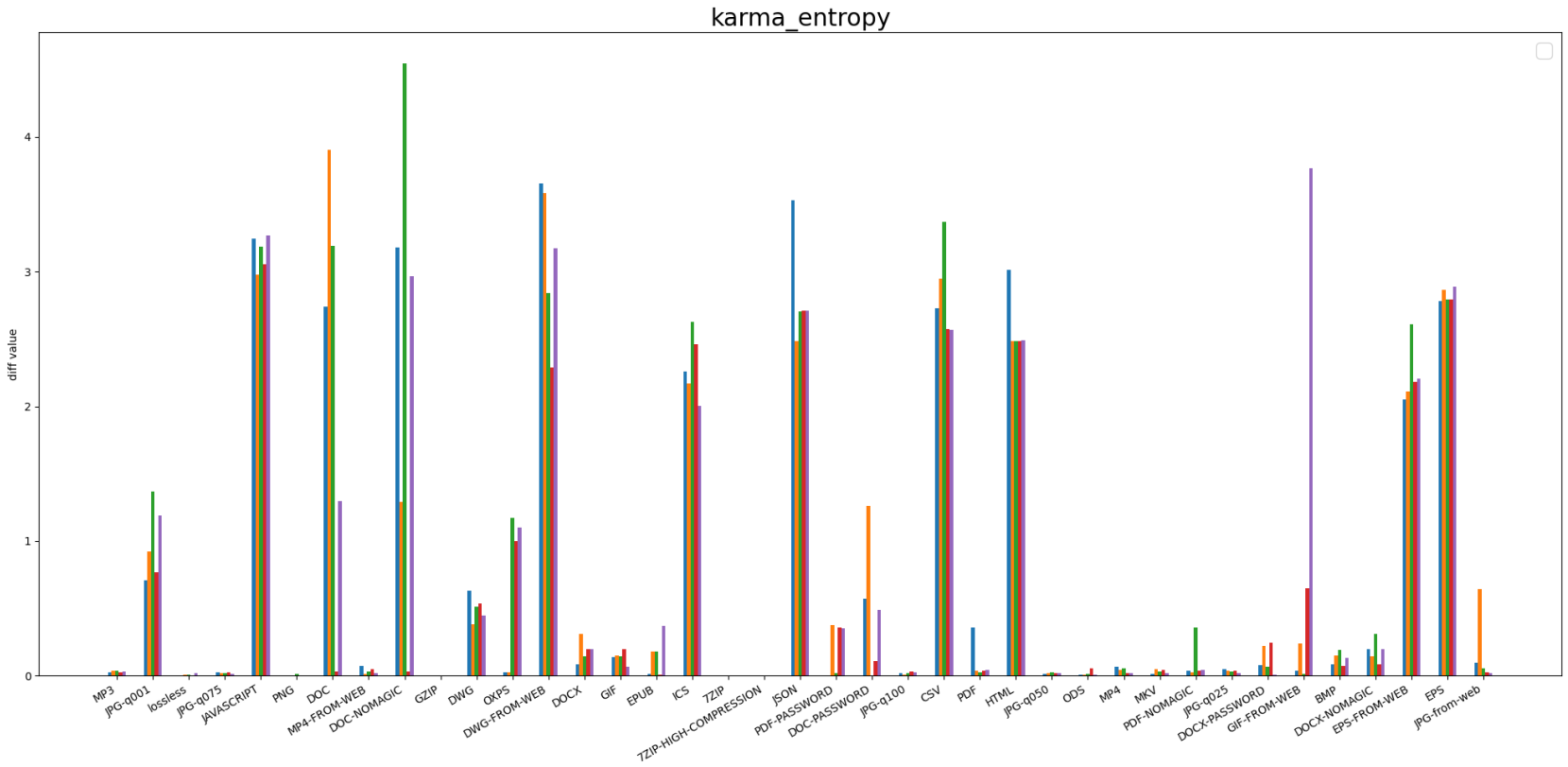


Figure: Entropy Measure for Karma

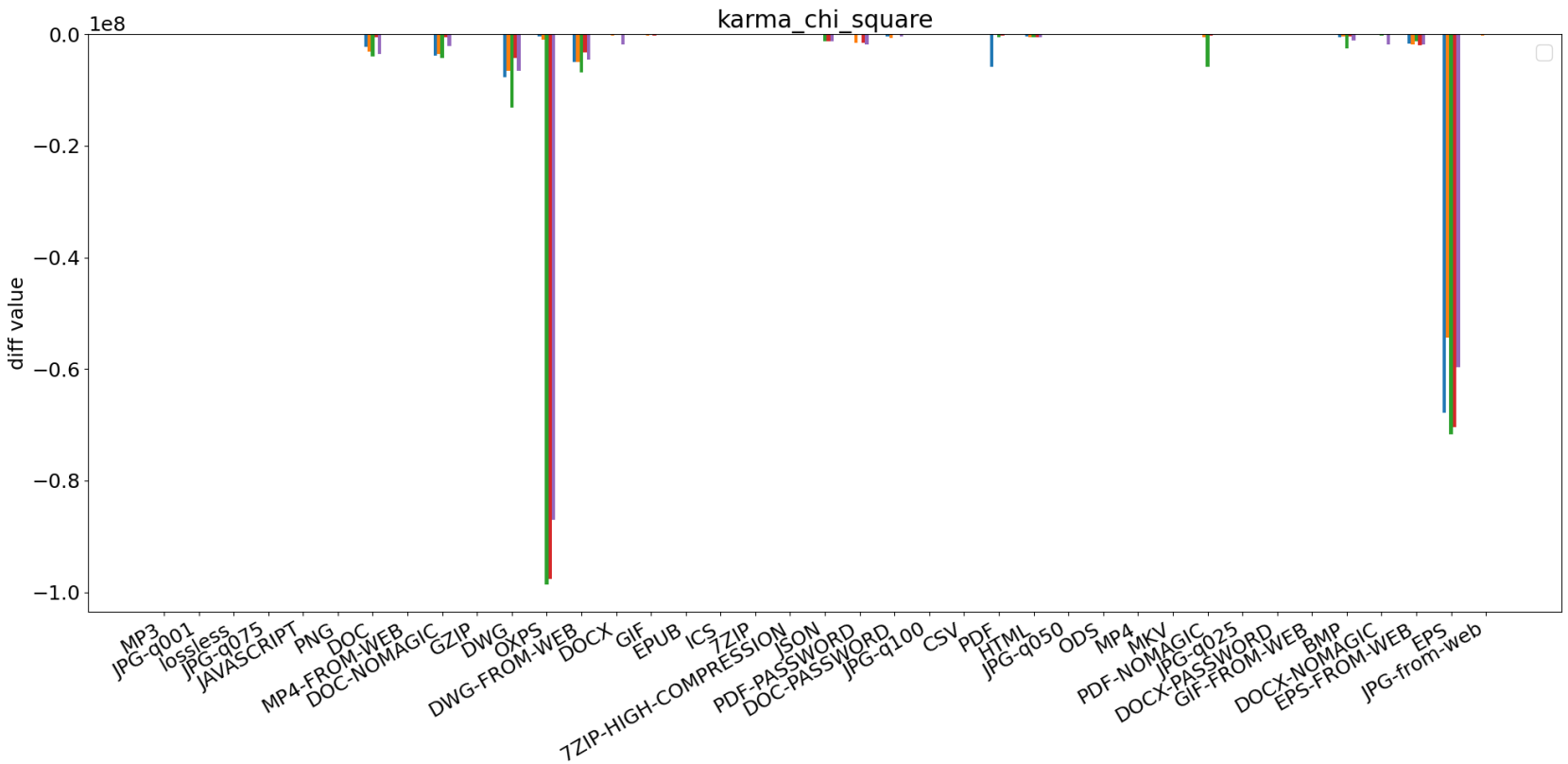


Figure: Chi Square Measure for Karma

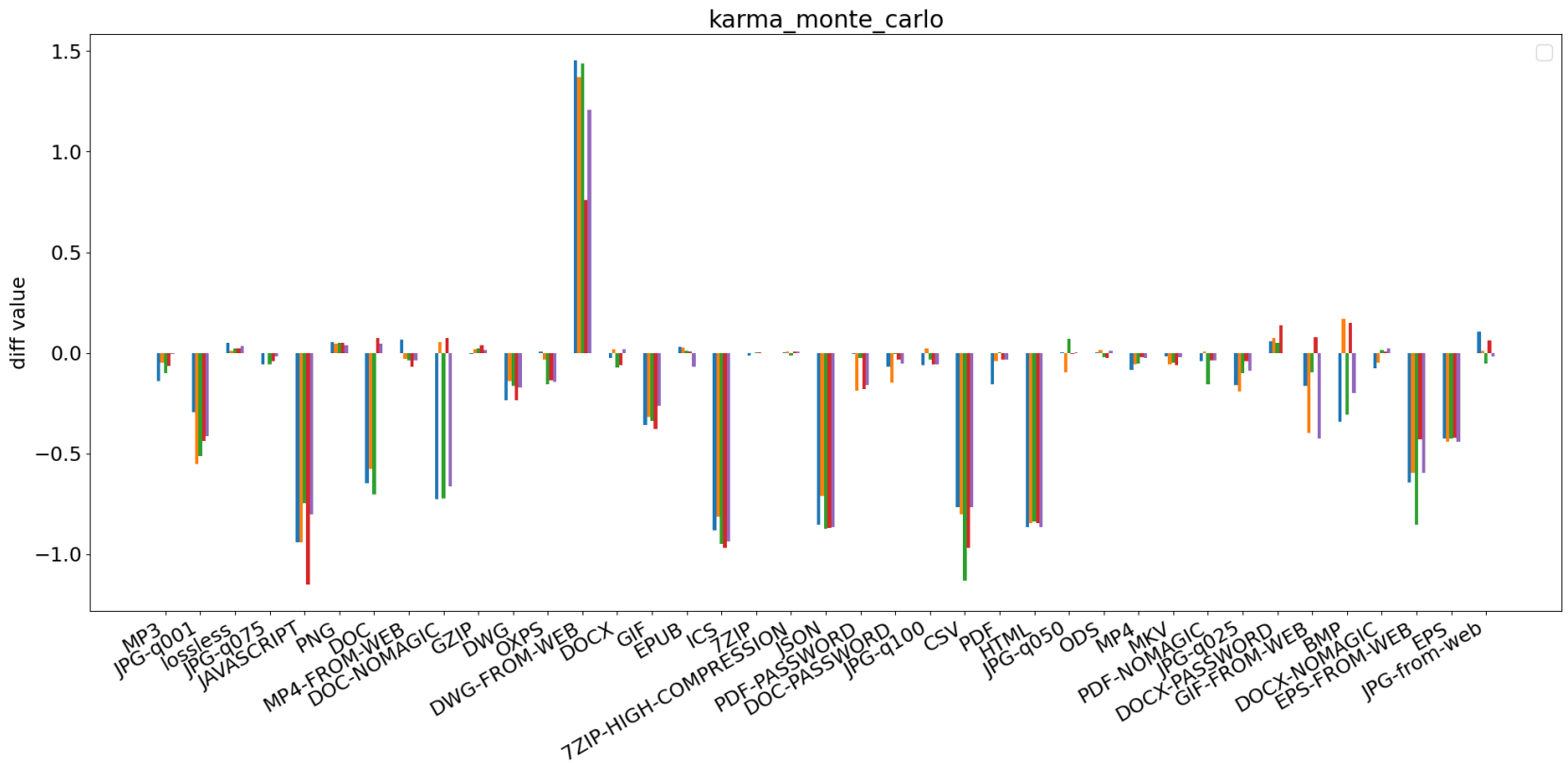


Figure: Monte Carlo Measure for Karma

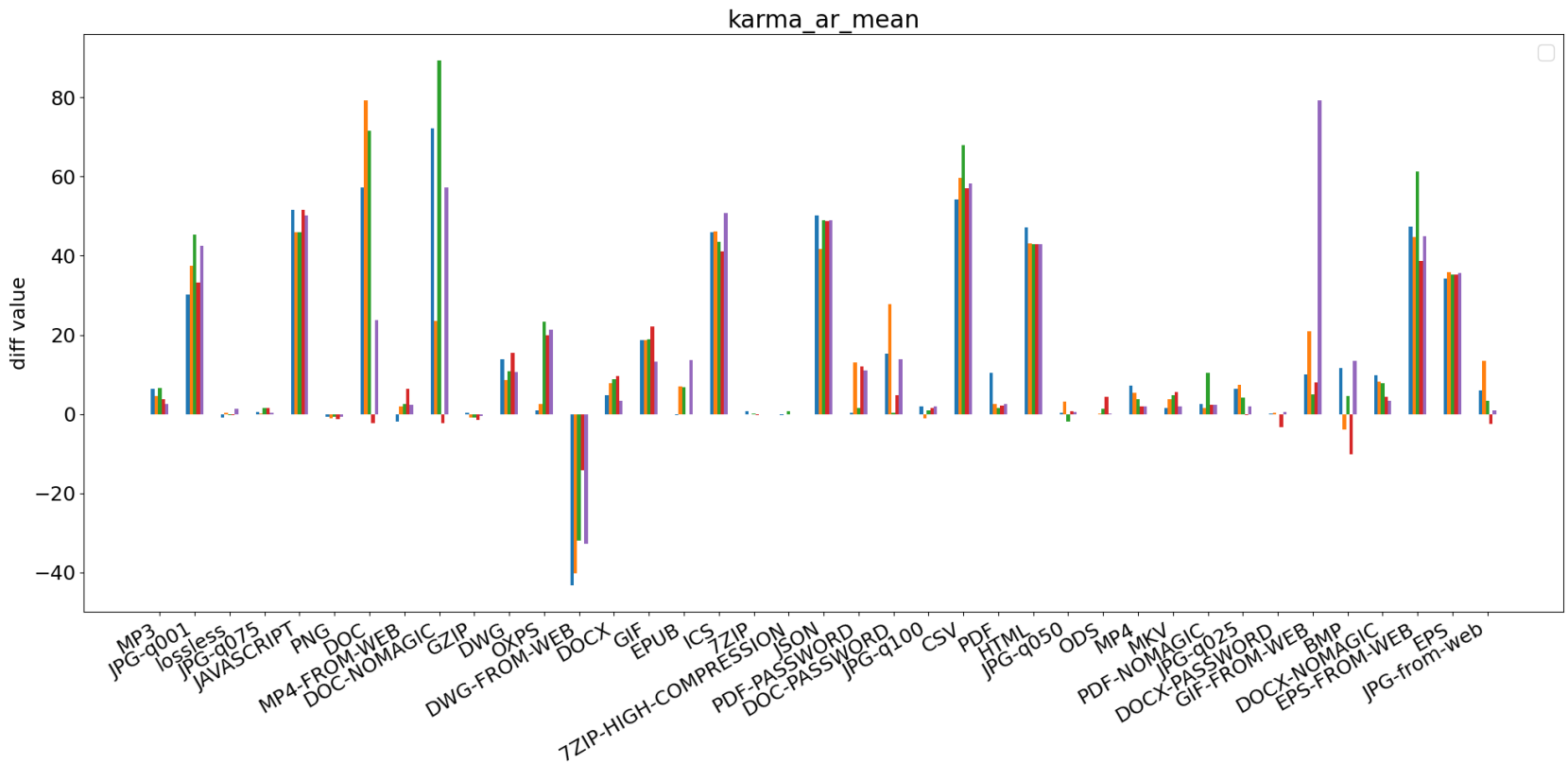


Figure: Arithmetic Mean Measure for Karma

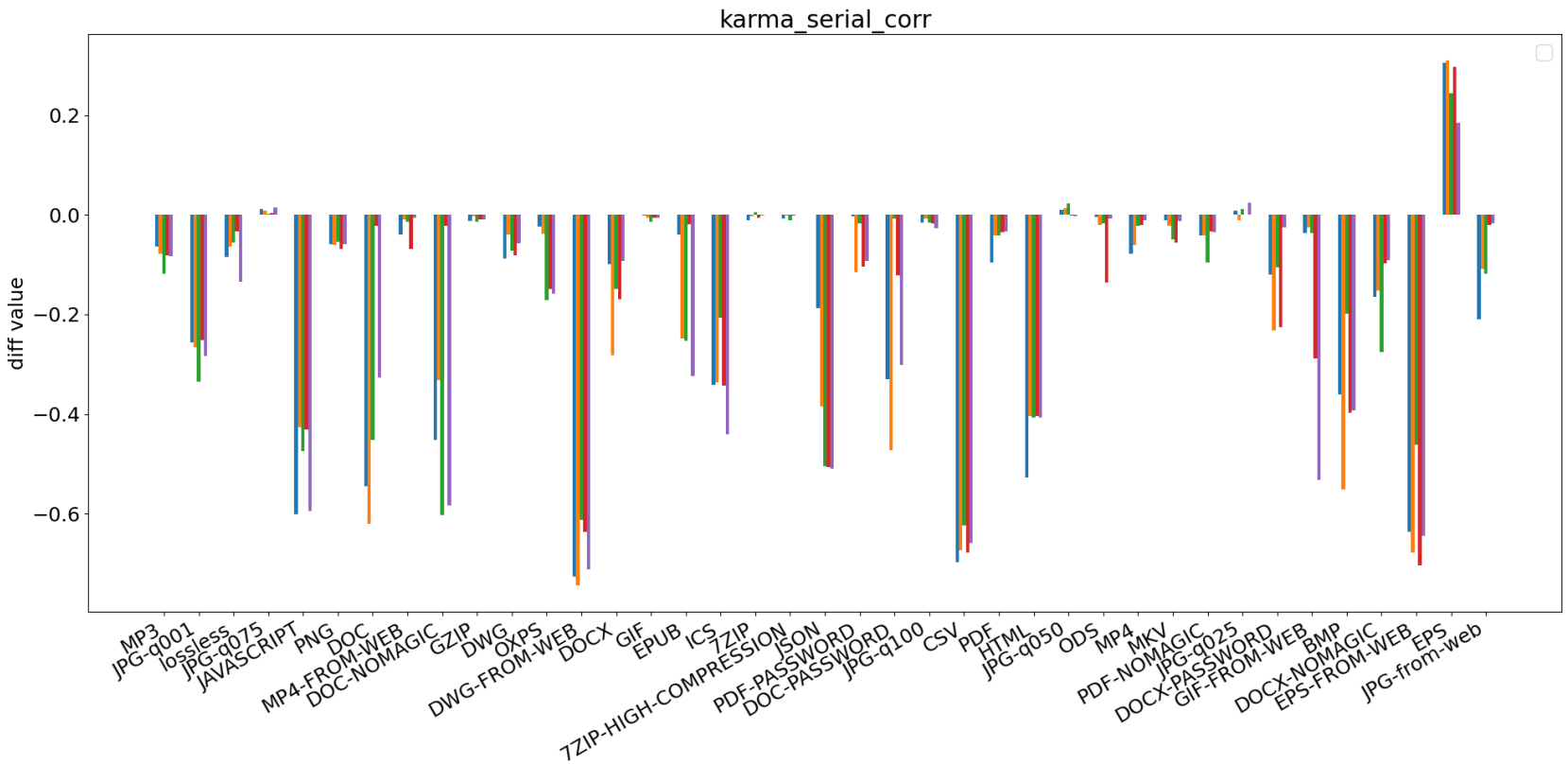
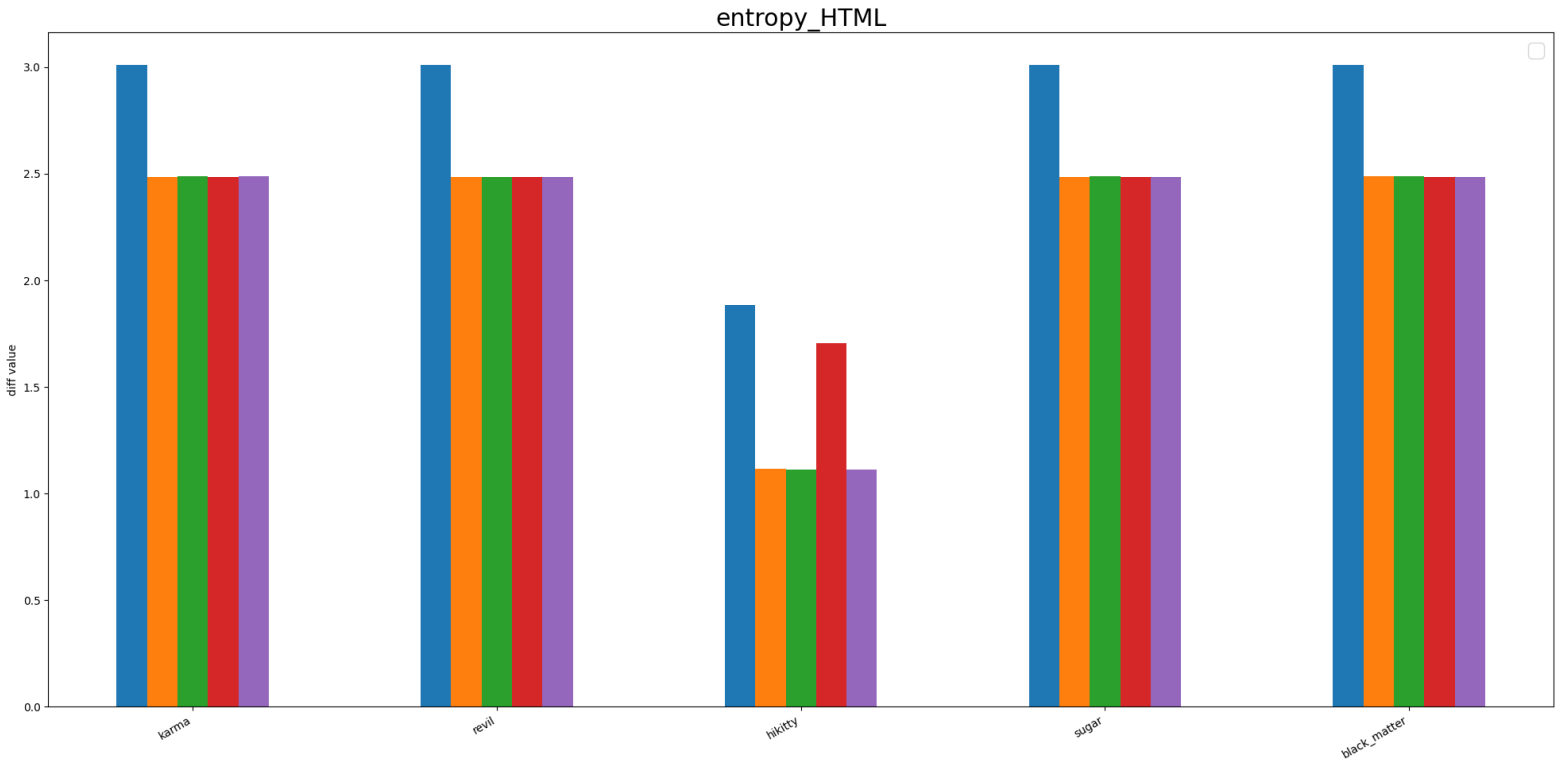
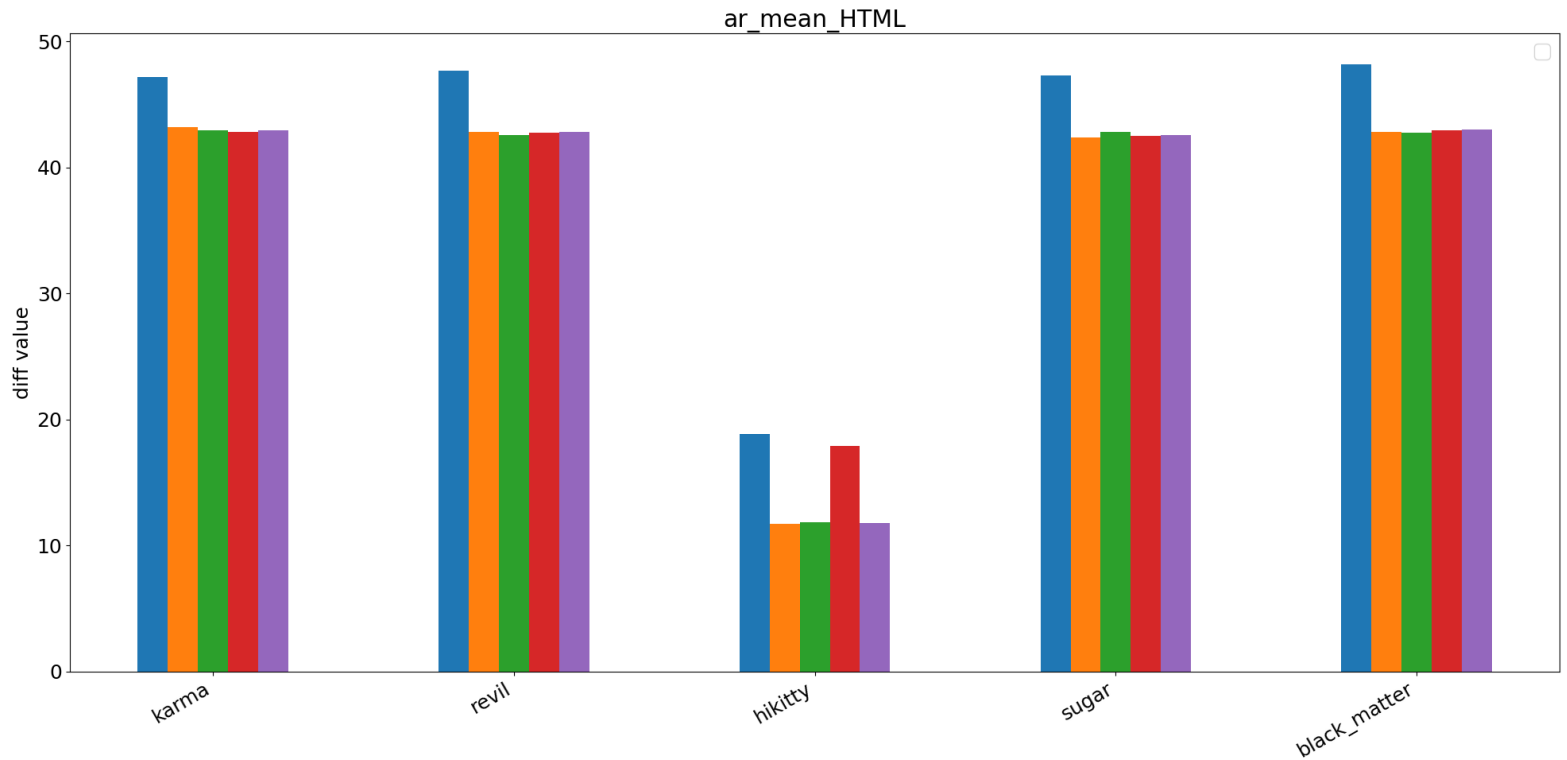


Figure: Serial Correlation Measure for Karma

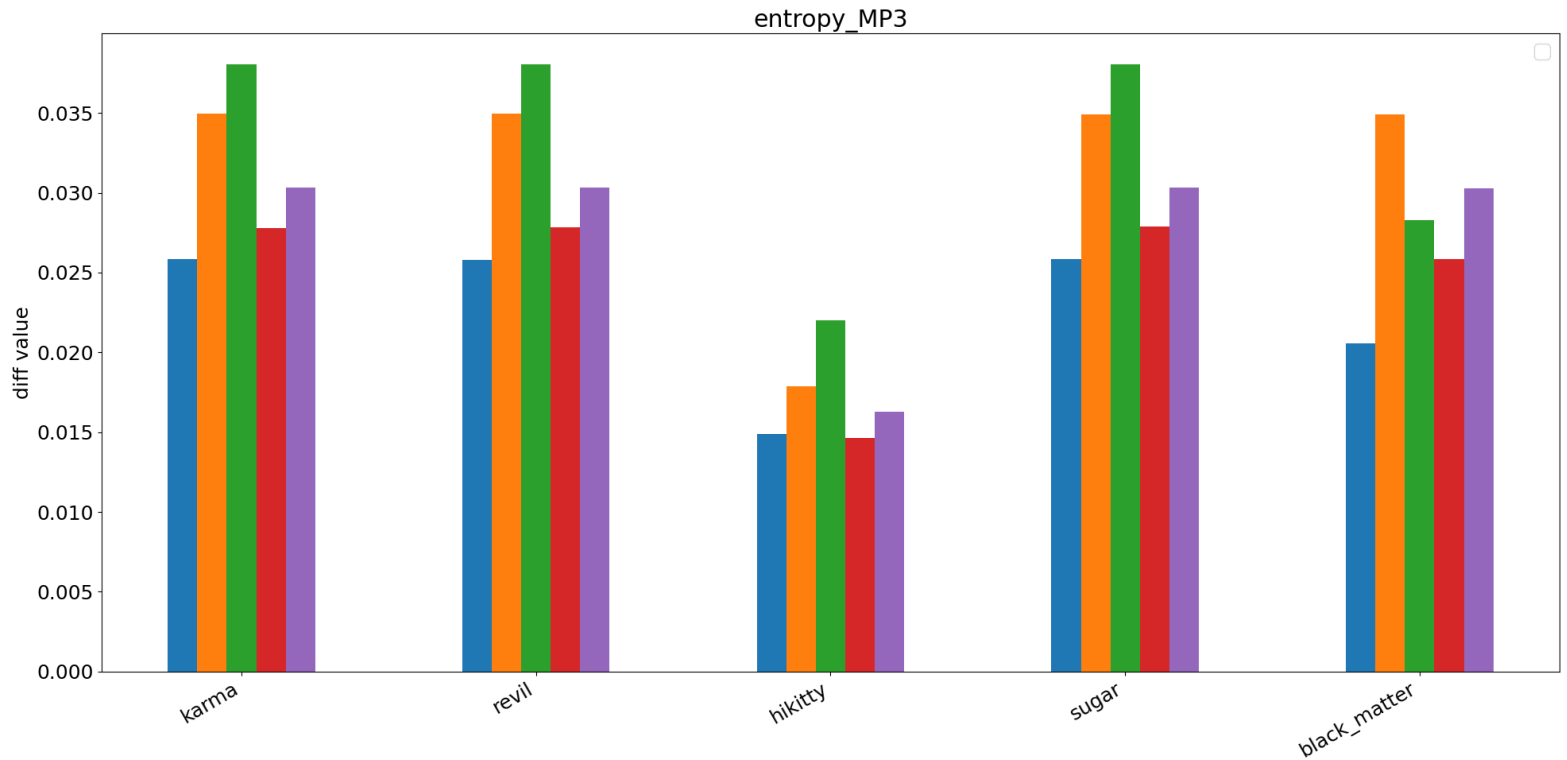
Result for low entropy file



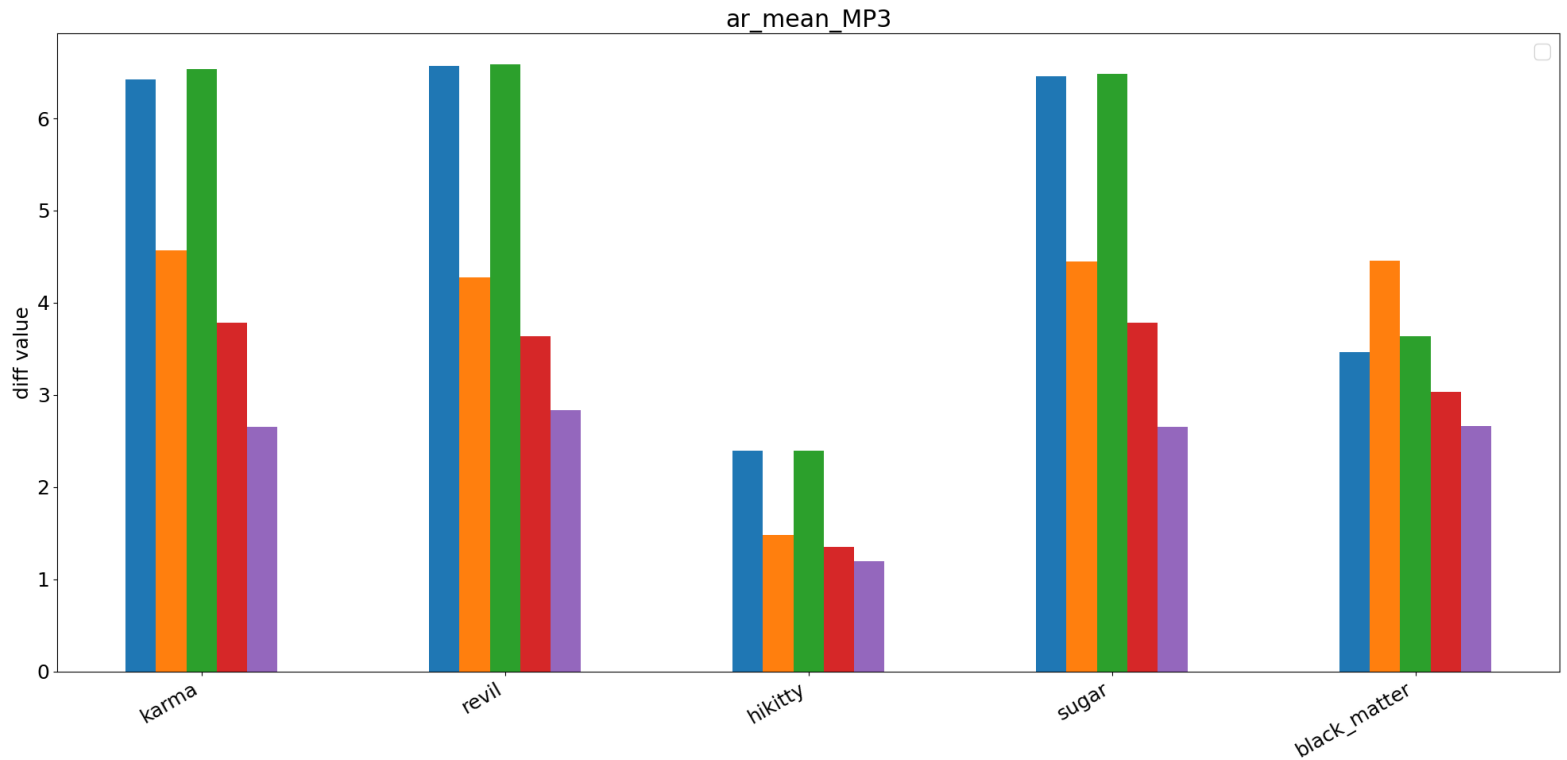
Result for low entropy file



Result for high entropy file



Result for high entropy file



System Call

How to act on a system ?

Everything is OS, OS is everything :

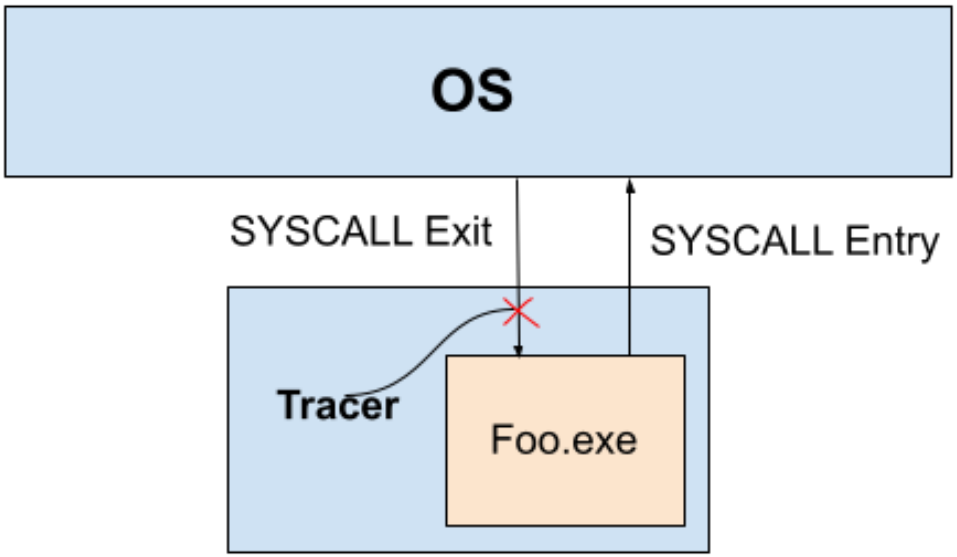
- Act on file, process, device, network...
- Open, read, write, delete...

How to manage with OS ?

Just ask and pray
Asking with signal called **System Call**

If we can see the System Call send by a ransomware, we know its behaviour !

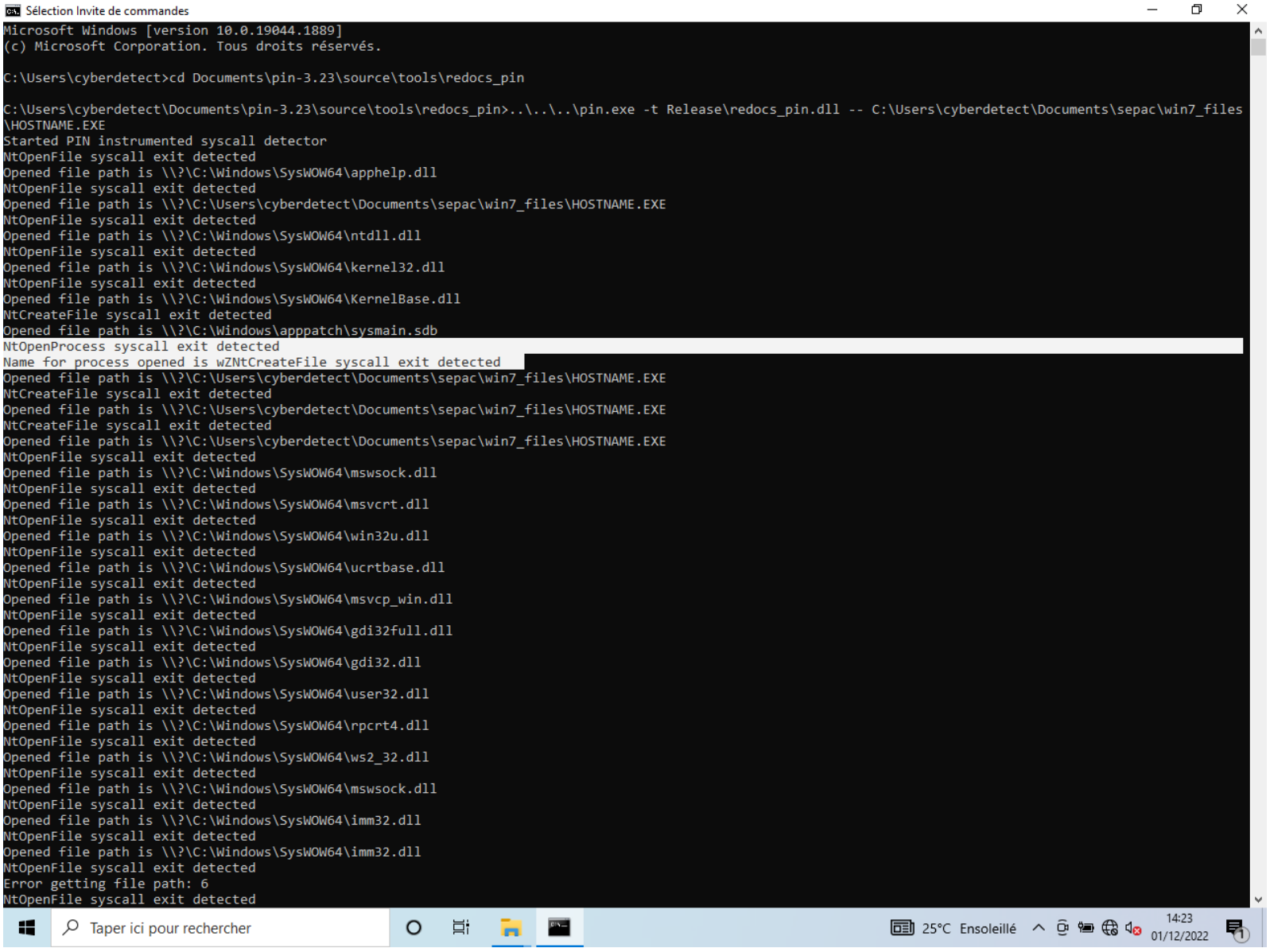
A useful tool : The Tracer



Tracer adapted to our project : Pin-tool

- NtOpenFile **(1)**
- NtCreateFile **(2)**
- NtWriteFile **(2)**
- NtDeleteFile **(3)**
- NtOpenProcess **(3)**
- NtTerminateProcess **(3)**

A little illustration



Disclaimer

Be careful with the results !

Non neutral impact on the ransomware's execution !

- Impact on the performance
- Impact on the behaviour of the ransomware

Crafty ransomware : tricks for hiding from the tracer
→ Not seeing a system call doesn't mean it doesn't exist !

Dataset

Study of 15 ransomwares : tracer + empirical observation on the state of the environment

SYSCALL observed by the tracer

	Karma (Normal behaviour)	Hentai Oniichan (Tracer killed !)	Sugar (Altered behaviour)
OpenFile	X	X	X
CreateFile	X		X
WriteFile	X		X
DeleteFile			
OpenProcess		X	
TerminateProcess	X		X

No SYSCALL observed, but interessant observations

Observations	Ransomware
Normal run	Wannacry gandcrab Zeoticus Blackmatter Hi_Kitty 2 Mallox
Run but no effect	CNH Hello LockBit Chaos
Altered behaviour	Ranzy Revil

Empirical observations without tracer

Observations	Ransomware
Add a new extension	Majority
No encryption for app	Blackmatter Karma Mallox REvil ...
Kind of replacing of file by evil README	gandcrab REvil
Open terminal for killing process	Hi_Kitty 2 (with deletion of file) Mallox
If no extension, no encryption	Chaos
New local disk, but no user access	Zeoticus Lockbit

Outline

Contexts

Simulated

Simulated The analysis is performed in a virtual machine
→ We can break anything

Not simulated The analysis is not performed in a virtual machine
→ We can not break anything
→ If there is a ransomware, it has to be detected/killed as soon as possible

Contexts

Controlled

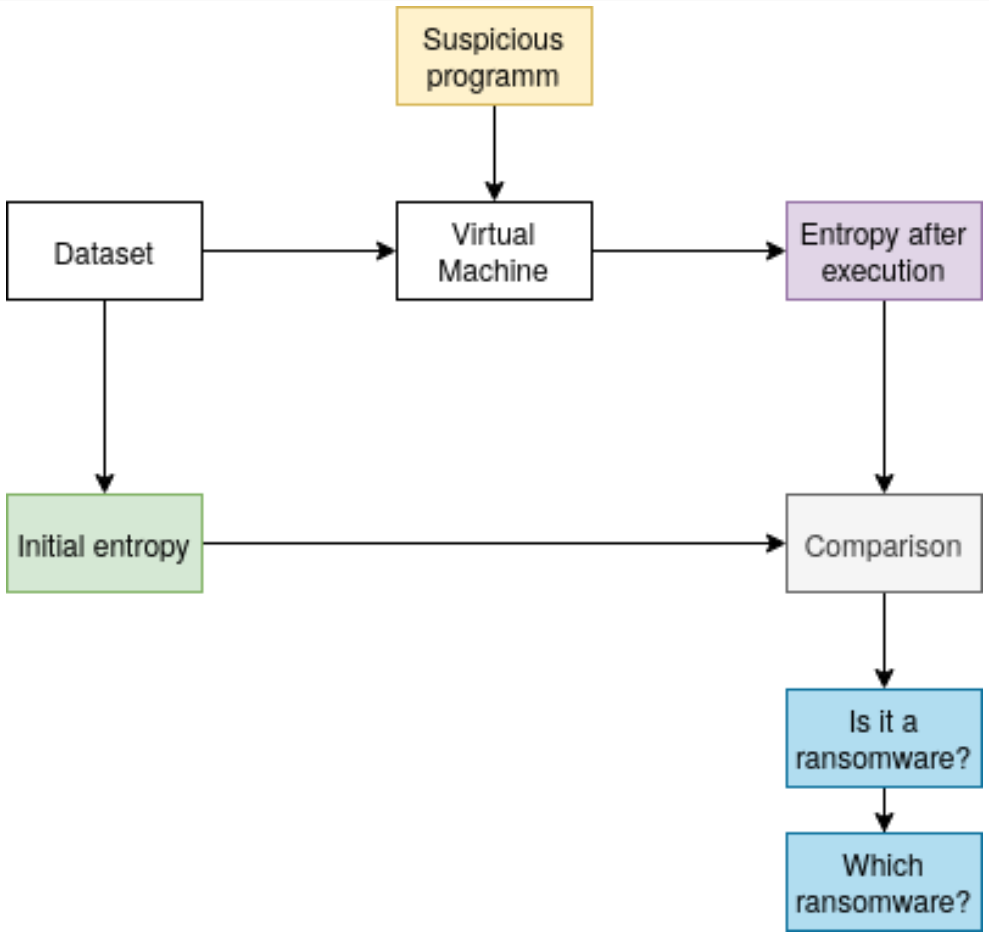
Controlled All the updates of the file system comes from the observed programm
→ If a change is done, it is done by the observed programm

Not controlled Other programms or the user can use the machine during the analysis

Case 1: Studying the entropy of the files

Context

Simulated | Controlled



A watcher: Mal-Aware

Monitoring

We can monitor each update in the file system

- Creation and deletion of files
- Modifications in a file
- Moves of files

We generate

- all the updates (in chronological order) of the filesystem
- the history of each file

A watcher: Mal-Aware

Analysis

We can raise alerts

- when the timestamps are manipulated
- when a lot of files are modified too quick
- when a lot of files are encrypted

Case 2: With the history of the file system

Context

Simulated | Controlled

Idea

- Monitoring the file system
- Running a suspicious programm
- Analysing the history of the file system

Case 2: With the history of the file system

Context

Simulated | Controlled

Idea

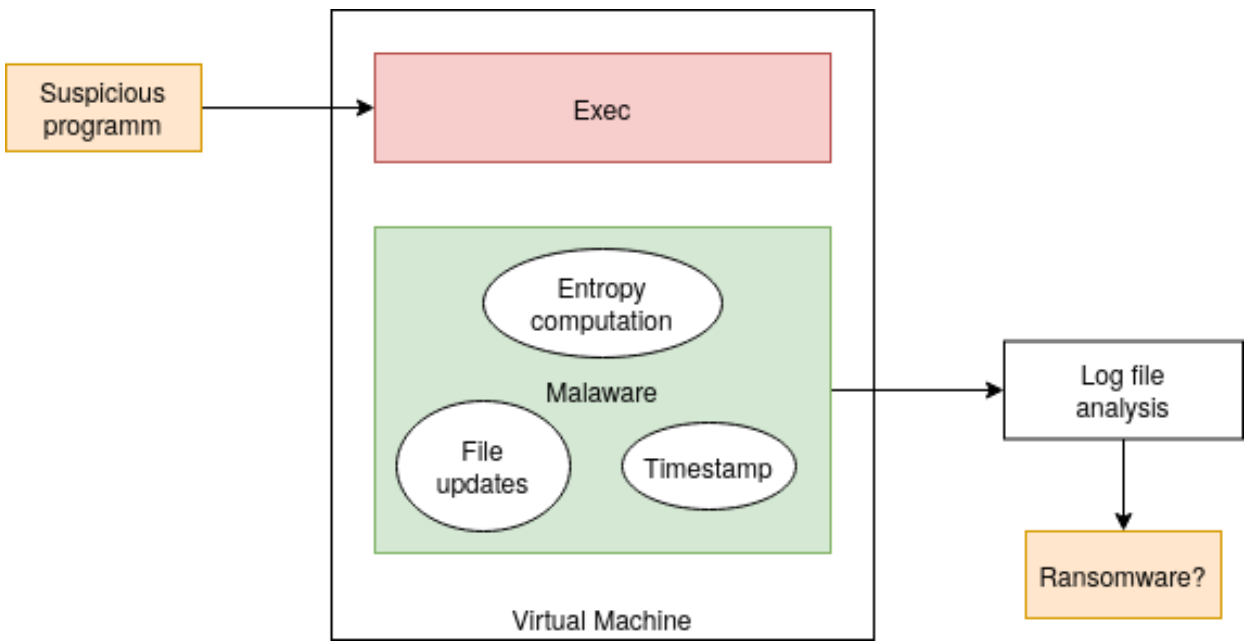
- Monitoring the file system
- Running a suspicious programm
- Analysing the history of the file system

- Are some files encrypted during the execution?
- Are suspicious actions performed? (eg. manipulation of timestamps)

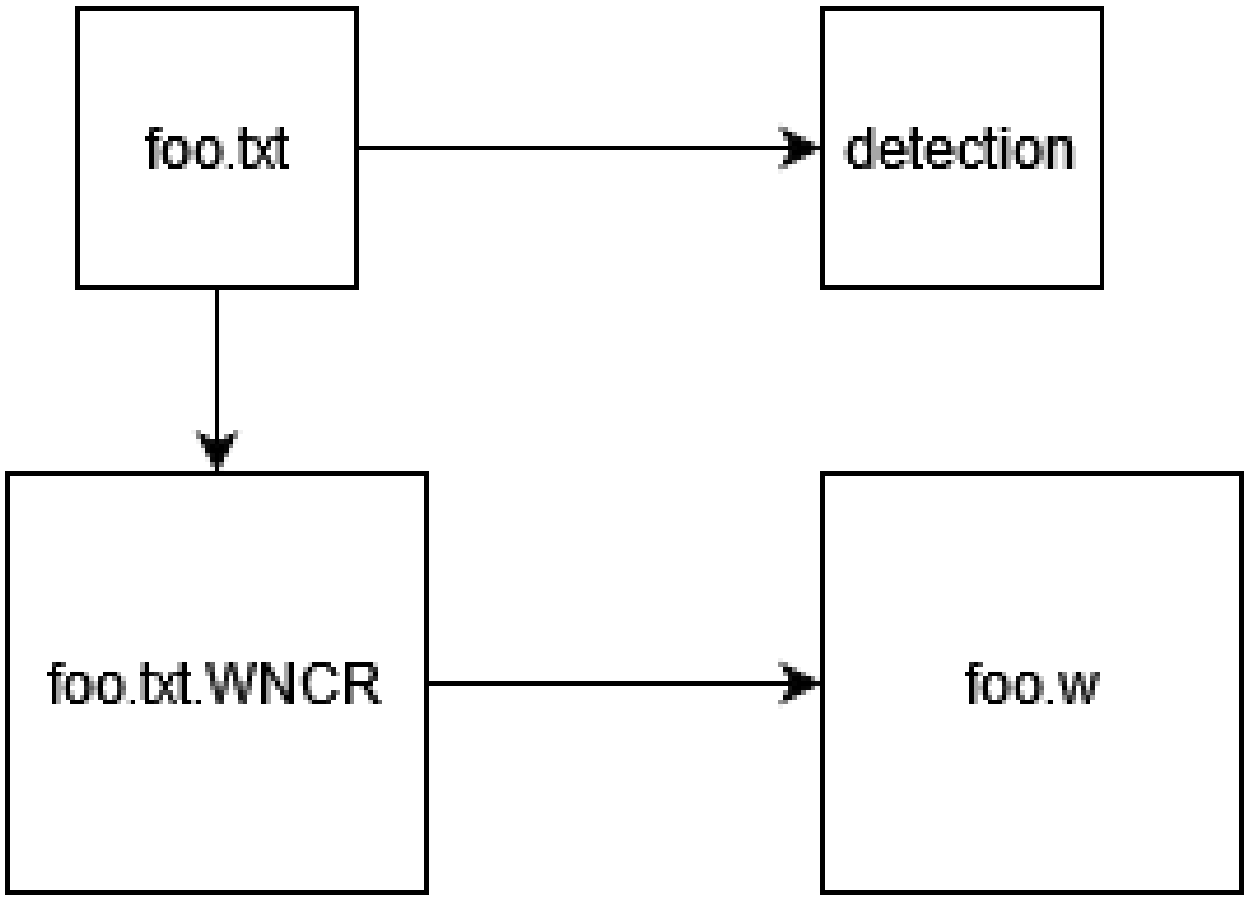
Case 2: With the history of the file system

Context

Simulated | Controlled



Case 2: With the history of the file system



Idea for case 3: Detection on-the-fly, as soon as possible

Context

Simulated - Controlled or Not simulated - not controlled

Idea

- Monitoring the file system
- Running a suspicious programm
- Analysing on-the-fly the history of the file system

Idea for case 3: Detection on-the-fly, as soon as possible

Context

Simulated - Controlled or Not simulated - not controlled

Idea

- Monitoring the file system
- Running a suspicious programm
- Analysing on-the-fly the history of the file system

- Are some files encrypted during the execution?
- Are suspicious actions performed? (eg. manipulation of timestamps)

Idea for case 3: Detection on-the-fly, as soon as possible

Context

Simulated - Controlled or Not simulated - not controlled

Idea

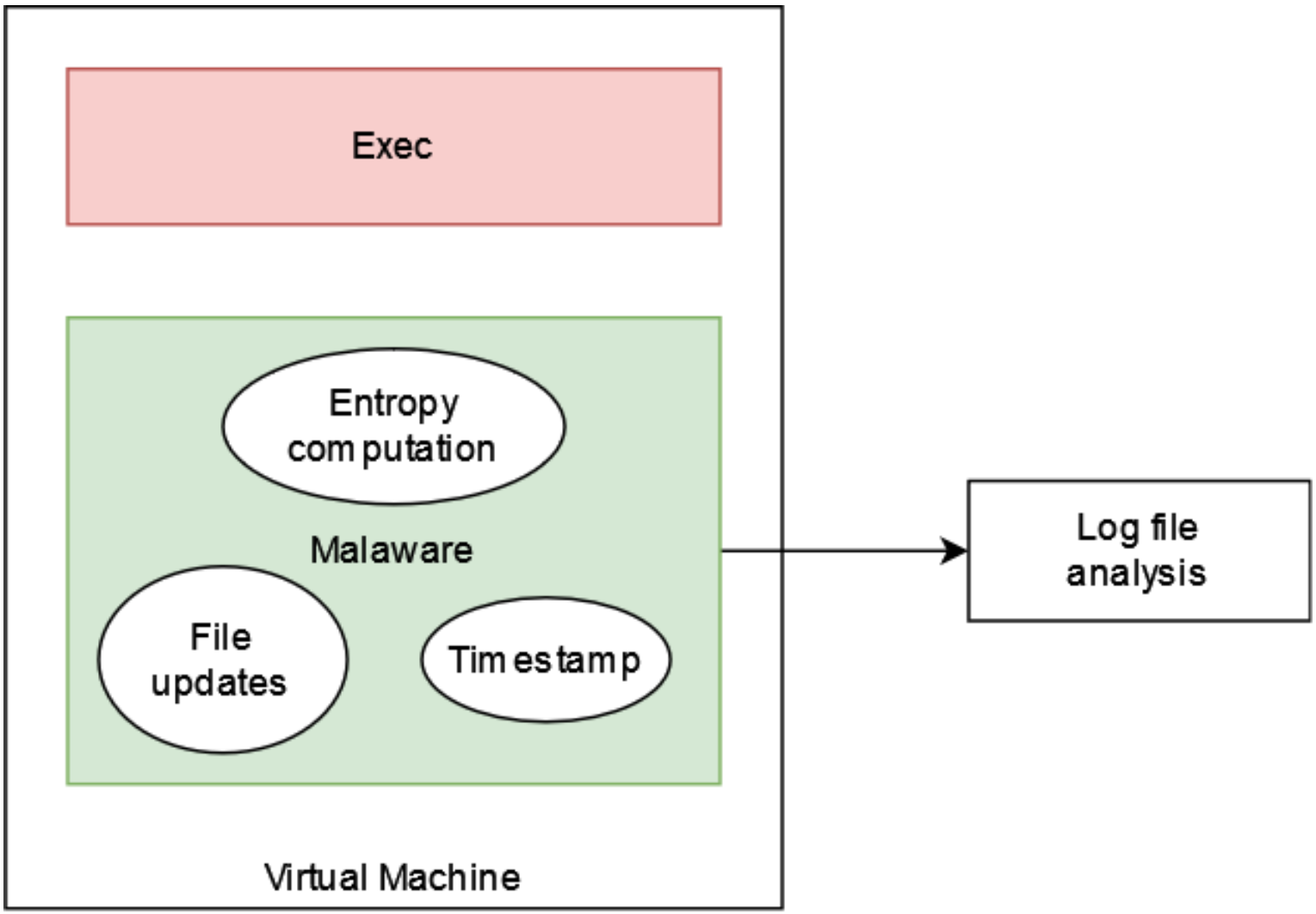
- Monitoring the file system
- Running a suspicious programm
- Analysing on-the-fly the history of the file system

- Are some files encrypted during the execution?
- Are suspicious actions performed? (eg. manipulation of timestamps)

We want to detect, as soon as possible, if the programm is a ransomware

Outline

Workflow


Exec
Entropy computation
Malaware
File updates
Timestamp
Virtual Machine
Log file analysis

Results

Entropy computation

- Study of different functions to compute files' entropy
- How malware execution affect files' entropy (or not) depending on file type

Empirical and tracer observations

- Monitoring of malwares' system calls with the tracer
- Empirical observations of malwares' execution
- Some pertinent events to monitor in order to detect a malware

Watcher

- Monitoring of creation, deletion, modifications and moves of files
- Focus on some directories
- Track files' history

Future works

Entropy computation

- Determining the best function to compute file entropy depending on file type
- Study a combination of entropy measurements

Tracer

- Enhance information received from events
- Intercept terminate process when tracer terminal is involved
- Print tracer information in a log file

Watcher

- Determine a pertinent set of directories to monitor
- Additional functions in the watcher