AnoMark

Anomaly detection in command lines with Machine Learning using Markov Chains

Alexandre Junius

ANSSI - French National Cybersecurity Agency

FIRST - June 2023

TLP:CLEAR

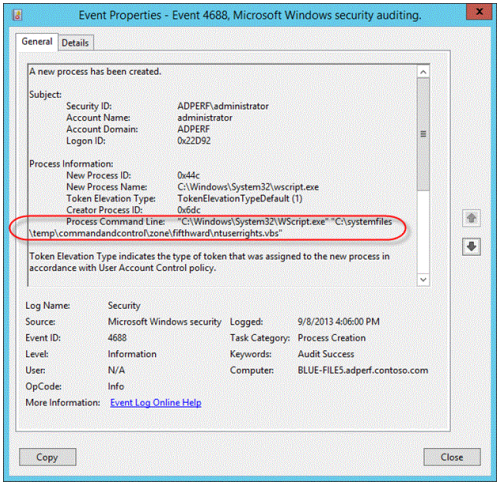
Table of contents

Table of contents

About me

- Formerly studied statistics in engineer school
- 3 years as Data Scientist at ANSSI (French National Cybersecurity Agency), part of a team of cybersecurity specialists
- Focusing on detecting intrusion in **endpoint logs**

Windows Security log sample



Windows Security Event ID 4688 : A new process has been created

Where to find the data ?

Command lines from processes can be found by :

- Enabling the "Audit Process Creation" audit policy, and the command line logging in Windows Security 4688
- Deploying Sysmon, the event ID 1 also tracks process creation and adds the parent process command line

Common methods in Intrusion Detection on event logs

Commonly intrusion detection on endpoints relies on analyzing event logs:

- Searching for IOCs (Indicators of Compromise)
- Creating signatures for known behaviors (example: SIGMA framework)
- Crafting custom alerts in a SIEM

Common methods in Intrusion Detection on event logs

Commonly intrusion detection on endpoints relies on analyzing event logs:

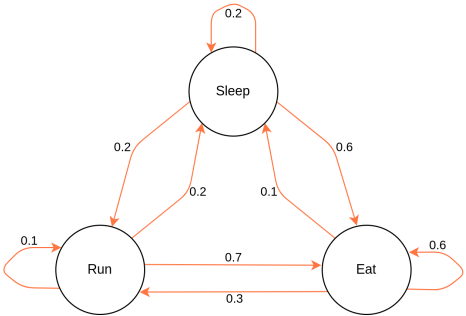
- Searching for IOCs (Indicators of Compromise)
- Creating signatures for known behaviors (example: SIGMA framework)
- Crafting custom alerts in a SIEM

But it is also a great field of application for statistical learning algorithms, particularly in the detection of anomalies. It can make it possible to move towards so far unknown behaviors.

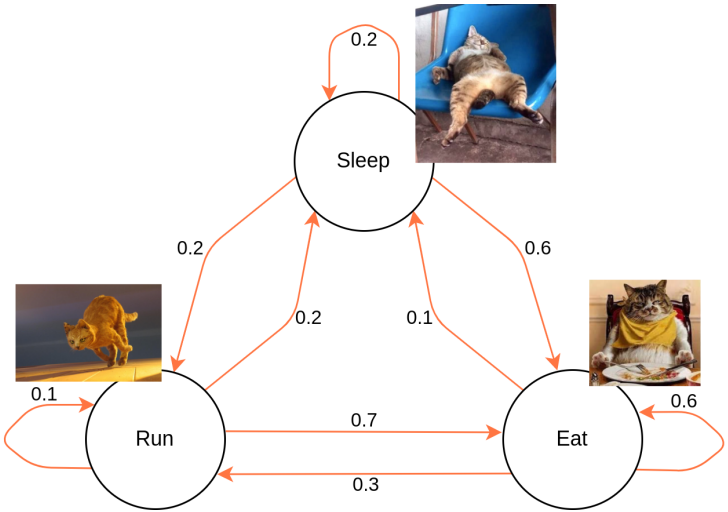
Table of contents

Markov Chains

The expression *Markov chains* refers to a mathematical concept allowing to model the transitions between states independently of the past. It is a stochastic process whose prediction of the future from the present is not made more accurate by the past.



Markov Chains - Cats version



https://t.me/learningnets

Ngrams of letters

We call cutting into ngrams of the command lines the fact of cutting them into groups of n letters.

» cmd.exe /c handle.exe

Model:

```
{"cmd.":  {"e":  100%}}
```

Ngrams of letters

We call cutting into ngrams of the command lines the fact of cutting them into groups of n letters.

```
» cmd.exe /c handle.exe
```

Model:

```
{"cmd.":  {"e":  100%},
"md.e":  {"x":  100%}}
```

Ngrams of letters

We call cutting into ngrams of the command lines the fact of cutting them into groups of n letters.

```
» cmd.exe /c handle.exe
```

Model:

```
{"cmd.":  {"e":  100%},
"md.e":  {"x":  100%},
"d.ex":  {"e":  100%} }
```

Ngrams of letters

etc.

Ngrams of letters

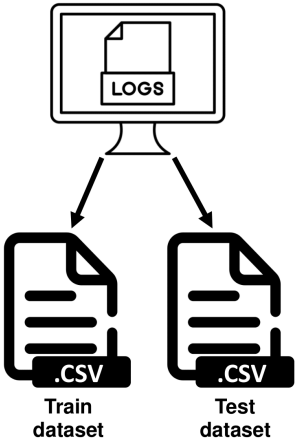
We call cutting into ngrams of the command lines the fact of cutting them into groups of n letters.

```
» cmd.exe /c handle.exe
» cmd.jar /c something.exe
```

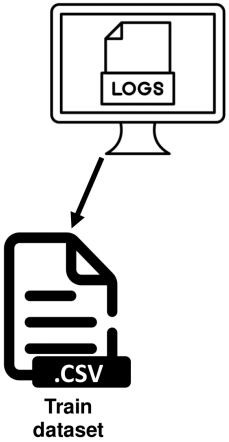
Model:

```
{"cmd.":  {"e":  50%, "j":  50%},
"md.e":  {"x":  100%},
"d.ex":  {"e":  100%},
...}
```

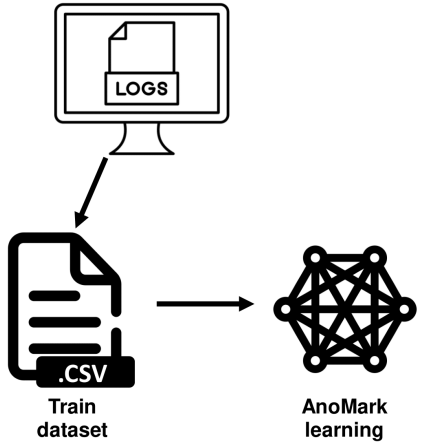
Application



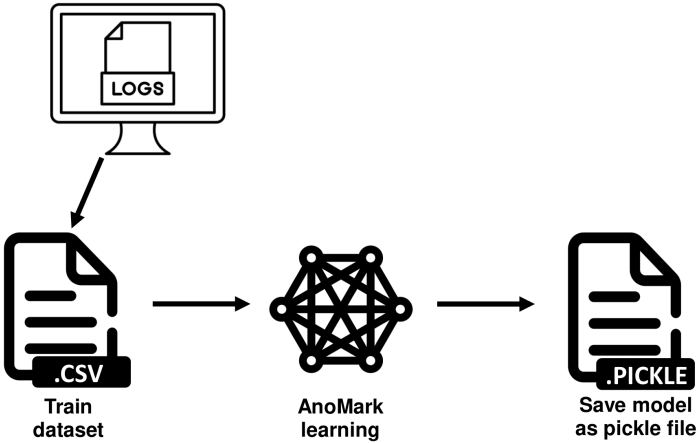
Application



Application

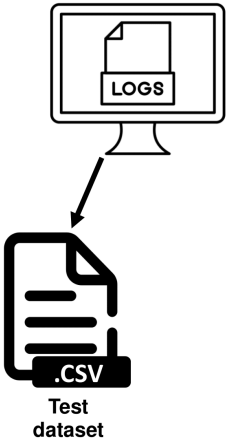


Application

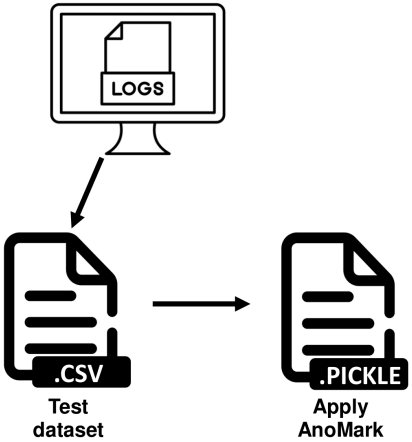


https://t.me/learningnets

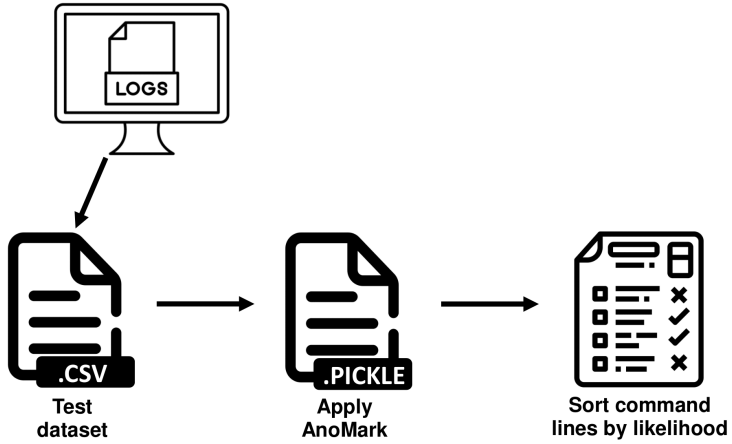
Application



Application



Application



Typical command lines detected by AnoMark

Examples of command lines detected by AnoMark:

Typical command lines detected by AnoMark

Examples of command lines detected by AnoMark:

- encoded command lines:

```
» powershell -EncodedCommand Rm9jdXMub24ucHJlc2VudGF0aW9uIQ==
```

Typical command lines detected by AnoMark

Examples of command lines detected by AnoMark:

- encoded command lines:

» `powershell -EncodedCommand Rm9jdXMub24ucHJlc2VudGF0aW9uIQ==`

- *ping* towards unusual domains:

» `ping heeeeeeeey.com`

Typical command lines detected by AnoMark

Examples of command lines detected by AnoMark:

- encoded command lines:

» `powershell -EncodedCommand Rm9jdXMub24ucHJlc2VudGF0aW9uIQ==`

- *ping* towards unusual domains:

» `ping heeeeeeeey.com`

- unknown process execution :

» `iWillPwnYou.exe /user adminAccount`

Typical command lines detected by AnoMark

And also:

Typical command lines detected by AnoMark

And also:

- unusual flags:

» `legit.exe -newflag newdata`

Typical command lines detected by AnoMark

And also:

- unusual flags:

» `legit.exe -newflag newdata`

- small changes in letters:

» `CmD.eXe -someflag -someparam`

Typical command lines detected by AnoMark

And also:

- unusual flags:

» `legit.exe -newflag newdata`

- small changes in letters:

» `CmD.eXe -someflag -someparam`

- known process executions from unknown paths:

» `C:\newfolder\myproc.exe`

Table of contents

GitHub project

- AnoMark is available on ANSSI's Github page
- Written in python
- Splunk *custom command* provided

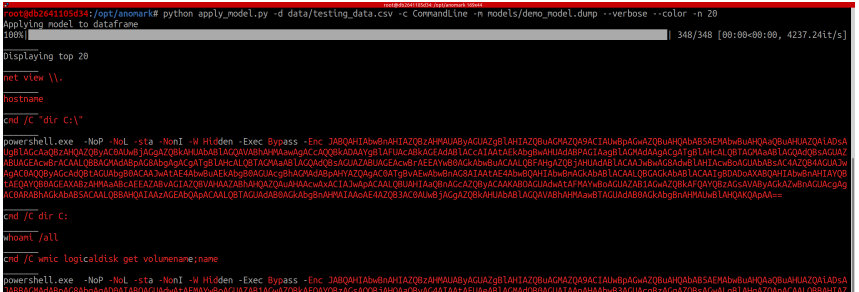
splunk>

Helping investigations

The algorithm can be used both in detection and in threat hunting, while being quick to set up (1 to 2 days, training included), which makes it an asset for investigations:

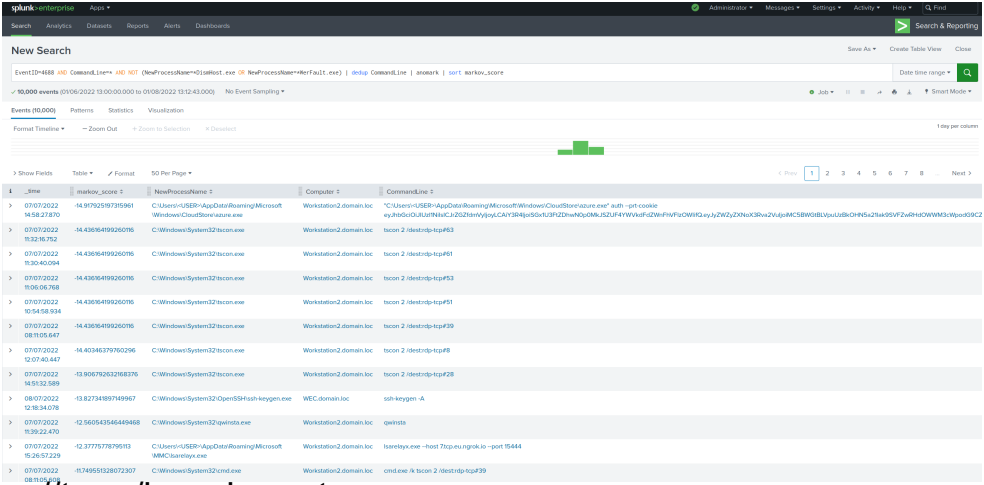
- *Post mortem* analysis
- Helping SOC team during their *live* monitoring

Demo: Creating a model and applying it to a sample inside terminal



https://t.me/learningnets

Demo: Custom command in Splunk



https://t.me/learningnets

Table of contents

Alerting

1. Each day, launch AnoMark on the data indexed the day before

Alerting

1. Each day, launch AnoMark on the data indexed the day before
2. Select the most unusual command lines (top 100)

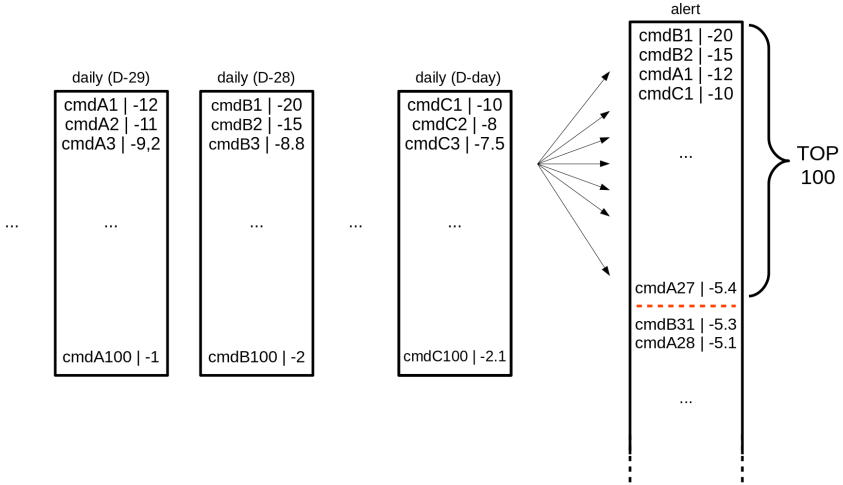
Alerting

1. Each day, launch AnoMark on the data indexed the day before
2. Select the most unusual command lines (top 100)
3. Compare this top with the most unusual command lines identified along the 30 previous days

Alerting

1. Each day, launch AnoMark on the data indexed the day before
2. Select the most unusual command lines (top 100)
3. Compare this top with the most unusual command lines identified along the 30 previous days
4. What can enter in the *historic* top is an alert

Alerting - Schema



Conclusion

This algorithm proves to us that statistical learning is a useful source of additional information. It opens the way to other anomaly detection algorithms, in the field of language processing or for other use cases that can be modeled by Markov Chains.

Time for questions, if we have time

Merci beaucoup !