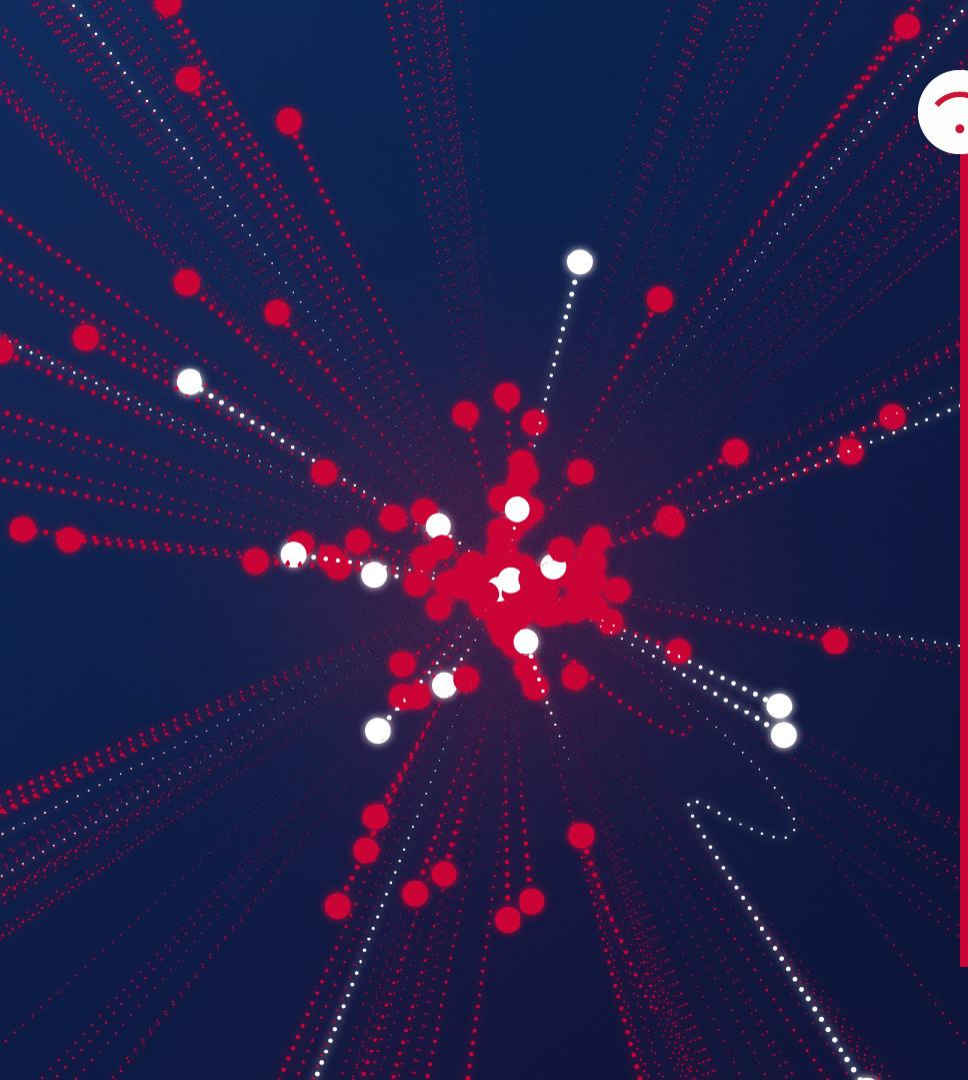


Attackers vs AI: how AI detects cyber threats

Daniele Ucci

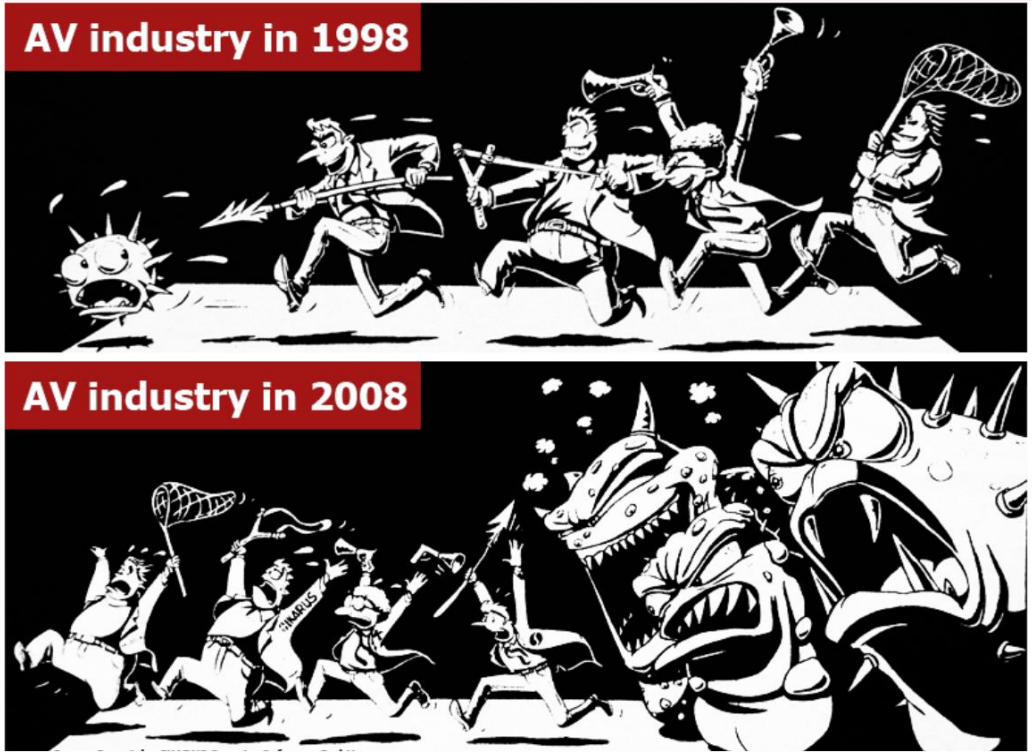
Molecon'21, December 2-4, Turin

aizoon® AUSTRALIA EUROPE USA
TECHNOLOGY CONSULTING

Outline

- Malware evolution
- Artificial Intelligence and Machine Learning applied to cybersecurity
- Common techniques used by attackers
- The infamous Wannacry ransomware
- The Astaroth malware family
- Offensive AI

Malware evolution



Source: IKARUS Software Security GmbH

Malware evolution

- The 2000s marked the transition from malware with either “noble” or “less-noble” purposes to:
  - recruit computers in a botnet in order to attack companies and organizations
  - industrial virtual espionage
  - mass surveillance
  - attack (h)ac(k)tivists
  - MaaS
- From single malicious software authors, now attackers are organized in cyber-criminal organizations structured as companies

Malware evolution

- The 2000s marked the transition from malware v noble" p
  in order to attack c

a cyber-
s

Sources – On the left: freebeacon.com. *Stolen F-35 Secrets Now Showing Up in China's Stealth Fighter,* 2014. On the right: quora.com. *US F-35 compared to Chinese J-31,* 2018.

Malware evolution

- The 2000s marked the transition from malware with either “noble” or “less-noble” purposes to:
  - recruit computers in a botnet in order to attack companies and organizations
  - industrial virtual espionage
  - attack (h)ac(k)tivists
  - MaaS
- From single malicious software authors, now attackers are organized in cyber-criminal organizations structured as companies

AI for cybersecurity

- Given the huge amount of information to process and the failure of classic approaches:
  - analysis processes should be automated as much as possible
  - security analysts should be supported by tools enabling the identification of possible cyber threats
- Artificial Intelligence (AI) can back their analyses highlighting suspicious events that are worth to be investigated by analysts

ML for cybersecurity

- Branch of artificial intelligence
- Allows an application to perform an activity without being explicitly programmed to do so by building a mathematical model
- Machine learning algorithms are responsible for creating mathematical models based on a data samples
- Differs from statistical approaches because:
  - there exist lots of statistical models able to make predictions, but accuracy is not one of their strengths

ML for cybersecurity

- provides more predictional power at the cost of a less interpretable mathematical model
- needs for a data sample for model creation

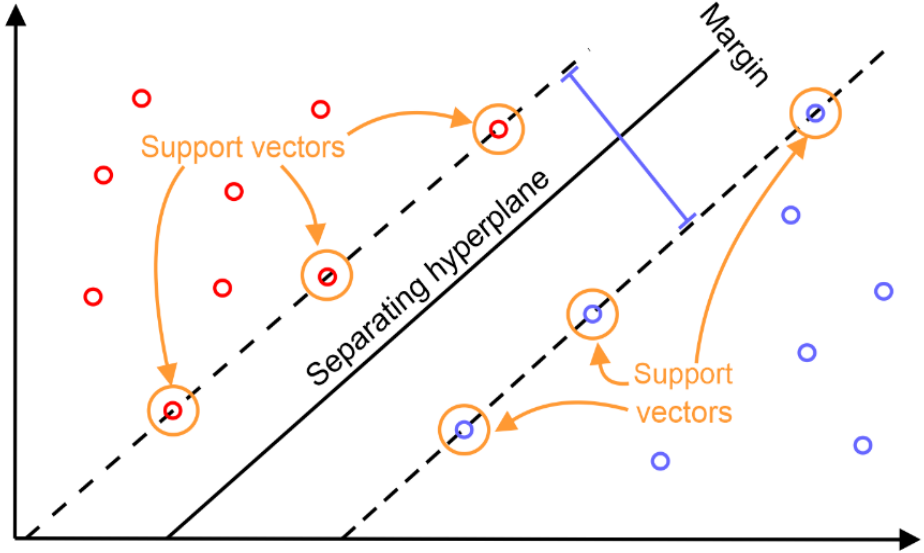
- ML-driven investigations can leverage different approaches:
  - supervised
  - unsupervised
  - semi-supervised

Overview on ML

- Classification is a typical example of supervised learning:

  Process for assigning an observation to a specific class on the basis of a labeled knowledge base, namely

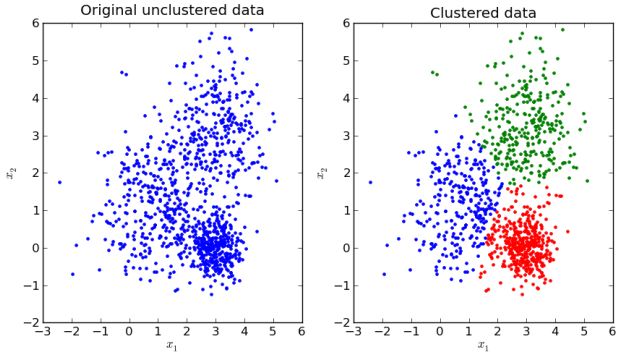
  - observation: bytes sent by a monitored machine → class: legit data transfer/data exfiltration
  - observation: HTTP traffic → suspicious HTTP traffic/ legit HTTP traffic



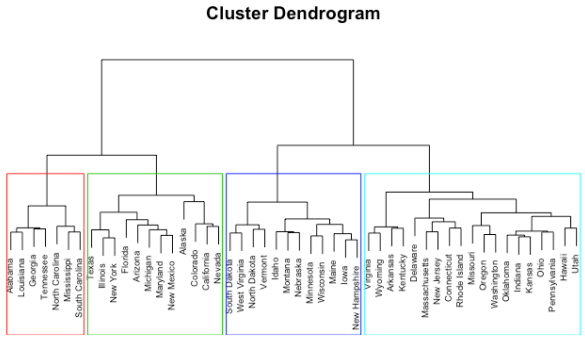
Source: K. Rezaul. *Machine Learning with Scala Quick Start Guide*. Packt Publishing, 2019.

Overview on ML

- Clustering is a typical example of unsupervised learning:
  - Task of grouping similar observations, like
    - observation: DNS query → similar to legit queries/similar to suspicious queries



Source: Medium.com. *A Beginner's Guide to Customer Segmentation with k-Means Clustering*. 2020. https://medium.com/@BryanJr/customer-segmentation-with-k-means-clustering-8167c8bd833b



Source: UC Business Analytics R Programming Guide. *Hierarchical Cluster Analysis*. 2018. https://uc-r.github.io/hc_clustering

Common techniques spotted into the wild

- DGA
- IP Flux
- Covert channel over DNS
- Covert channel over HTTP
- Host/Port scan
- Encrypted communications
- …

The infamous Wannacry ransomware

- Attributed[1,2] to the APT Lazarus Group:
  - a North Korean state-sponsored cyber threat group
  - active since 2009
  - responsible for many disruptive attacks (e.g., the wiper attack against Sony Pictures Entertainment)
- Wannacry samples use:
  - DGA
  - host/port scan

[1] WSJ.com. *It's Official: North Korea Is Behind WannaCry*. https://www.wsj.com/articles/its-official-north-korea-is-behind-wannacry-1513642537.
[2] MITRE.org. *Wannacry*. https://attack.mitre.org/software/S0366/.

The infamous Wannacry ransomware

- Exploitation of MS implementation of SMB protocol through Eternal Blue[3]
- Wannacry attempts to contact kill-switch domains to establish if it has to encrypt disk(s)
  - kill-switch domains are of the form www[.]iuqerfsodp9ifjaposdfjhgosurijfaewrwergwea[.]com[4]
  - Infected machines querying kill-switch domains are more active in the network with respect to healthy machines[5]
- Infected systems not able to contact kill-switch domains starts a reconnaissance phase and concurrently encrypt data

[3] MITRE.org. *CVE-2017-0144*. https://cve.mitre.org/cgi-bin/cvename.cgi?name=cve-2017-0144.
[4] FireEye.com. *Wannacry Ransomware Campaign: Threat Details and Risk Management*. https://www.fireeye.com/blog/products-and-services/2017/05/wannacry-ransomware-campaign.html.
[5] Cisco.com. *The Hours of WannaCry*. https://umbrella.cisco.com/blog/the-hours-of-wannacry.

The infamous Wannacry ransomware

- Reconnaissance phase consists in
  - scanning random IPs for open TCP 445 ports[6]
    - about 25 IP addresses per second

  to rapidly spread the infection
- Execution of Eternal Blue exploit to infect machines with open TCP 445 ports
- Reconnaissance and exploitation phase are run in parallel

[6] FireEye.com. *Wannacry Ransomware Campaign: Threat Details and Risk Management*. https://www.fireeye.com/blog/products-and-services/2017/05/wannacry-ransomware-campaign.html.

Detect DGAs with ML

- Kill-switch domains seem to be obtained from “someone smashed a few keys on the upper rows of the keyboard”[7]

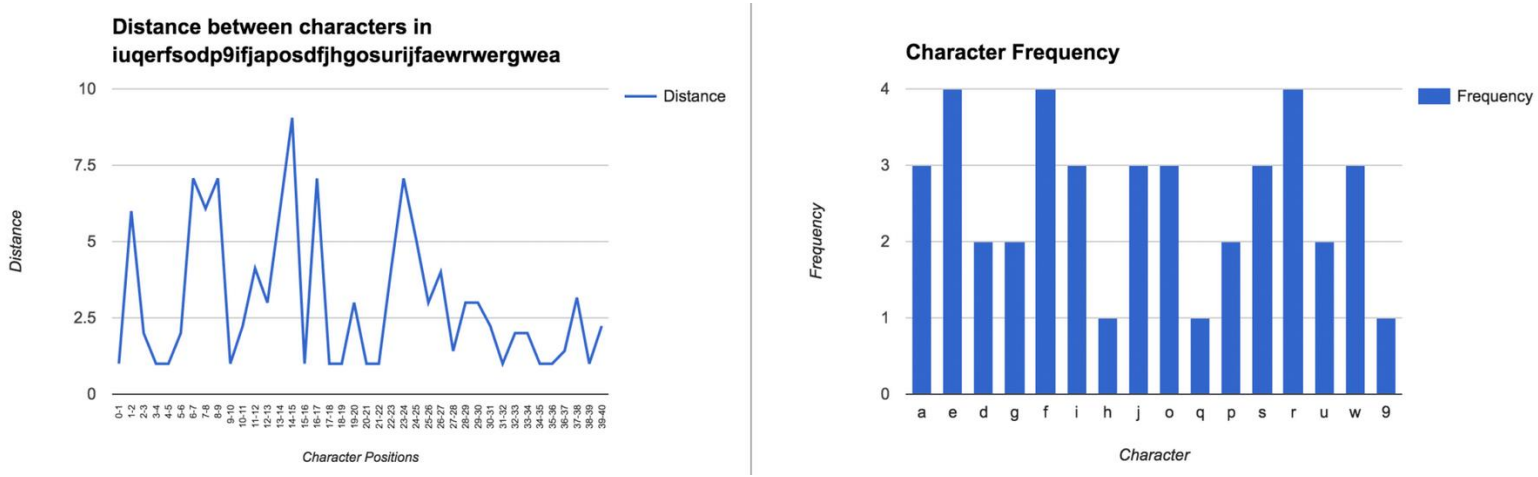


Figure: Character distance and frequency of the kill-switch domain “iuqerfsodp9ifjaposdfjhgosurijfaewewergwea”

[7] Cisco.com. *The Hours of WannaCry*. https://umbrella.cisco.com/blog/the-hours-of-wannacry.

Detect DGAs with ML

- However, kill-switch domains look like as algorithmically generated, also known as Domain Generated Algorithm

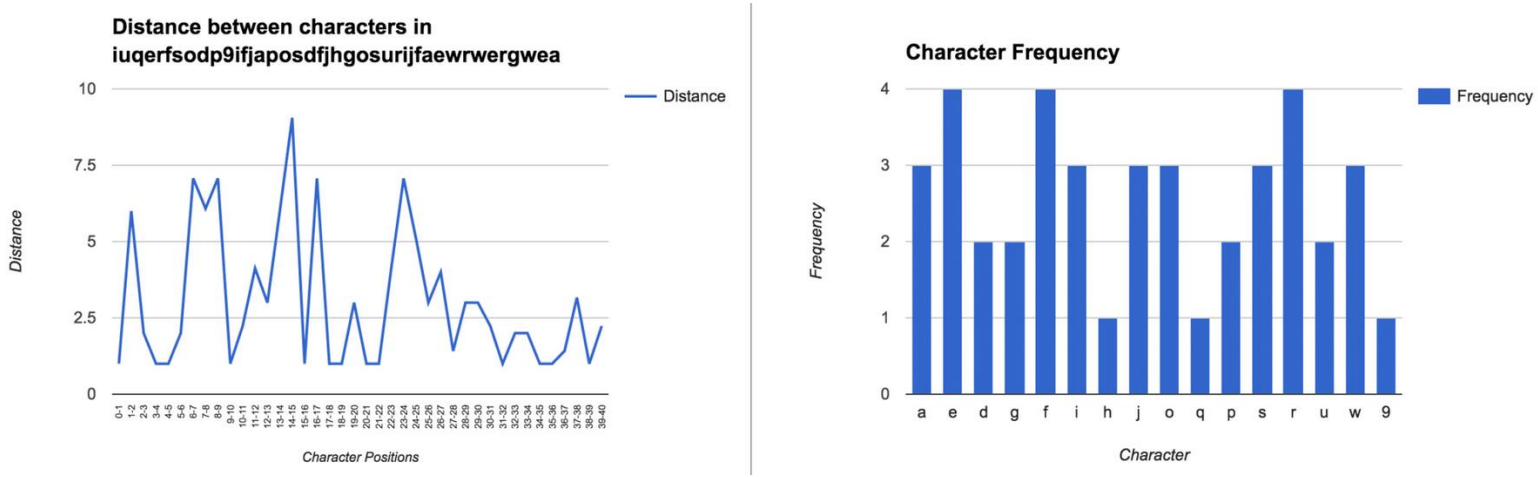
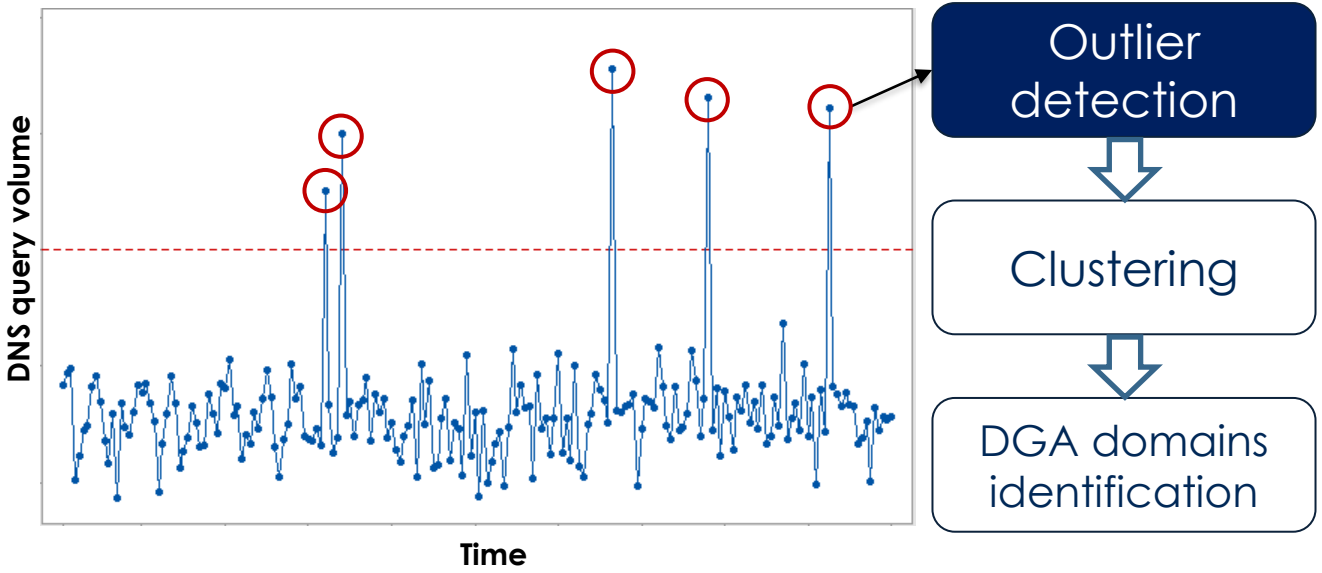


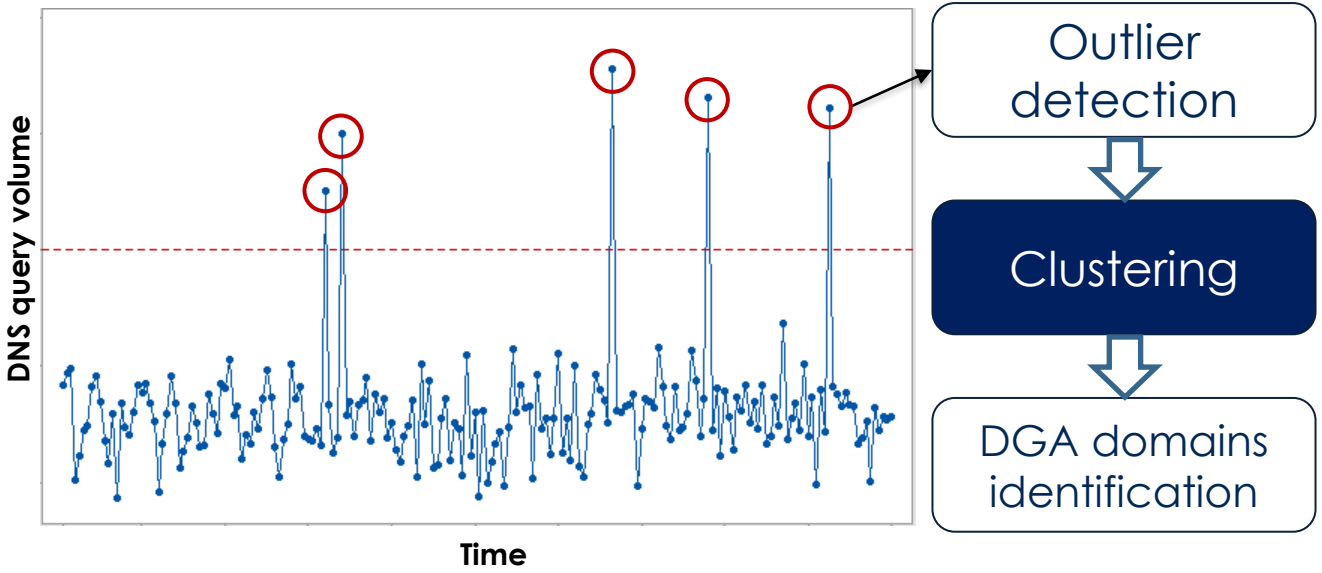
Figure: Character distance and frequency of the kill-switch domain "iuqerfsodp9ifjaposdfjhgosurijfaewewergwea"

Detect DGAs with ML



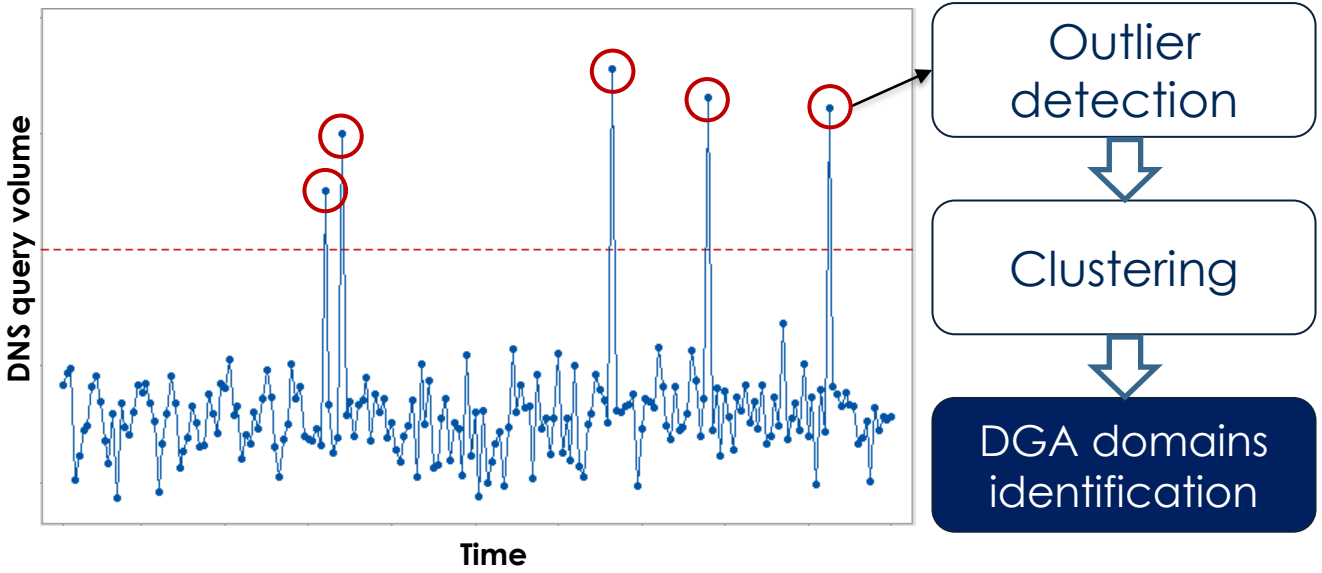
- DNS query peaks in a specific time window are outliers that deviate from the usual query volume of a machine
- Outliers can be determined on the basis of the combination of statistical indicators which model deviations from machine usual DNS queries

Detect DGAs with ML



- Characterizing features of domains involved in DNS query peaks are extracted
- A clustering algorithm can be applied on extracted features to group similar domains
- Resulting clusters represent legitimate domains and possible different DGAs

Detect DGAs with ML



- Both cluster and lexical analyses establish the presence of possible algorithmically generated domains
- Resolutions of these domains increases the criticality level

Detect host/port scans with ML

- Host and port scannings can be identified by
  - analyzing flags of connections between machines that represent
    - TCP and UDP pings, TCP ACK pings, and TCP SYN pings in case of host discovery
    - TCP and UDP scans and TCP SYN/NULL/FIN scans in case of port scans
  - collecting features about the number of destination machines (and ports) reached in a specific time interval
  - outliers in collected features describe likely scanning activities

The Astaroth malware family

- Astaroth[8] samples are trojans and steal information about
  - e-mails
  - e-commerce
  - banking accounts
- Attacks companies in Europe and Latin America since late 2017
- In July 2021, samples have been distributed in a Malspam campaign through mails containing an infected link

[8] MITRE.org. *Astaroth*. https://attack.mitre.org/software/S0373/.

The Astaroth malware family

- The infected link provides access to a compressed Powershell script performing different infection stages:
  - creation of a support file specifying a C&C domain
  - HTTP connection to C&C to download an XML file containing
    - Javascript source code executed to download malicious DLLs
    - a compiler for guaranteeing persistence
- Downloaded DLLs extract sensitive information, later exfiltrated to either predetermined or algorithmically generated malicious domains
- Communications with the C&C are established through DGA and are both encrypted and in clear

Detect covert channels over DNS with ML

- A covert channel is a technique to communicate, transfer or exfiltrate data using different protocols, after having exploited machines in a network[9]

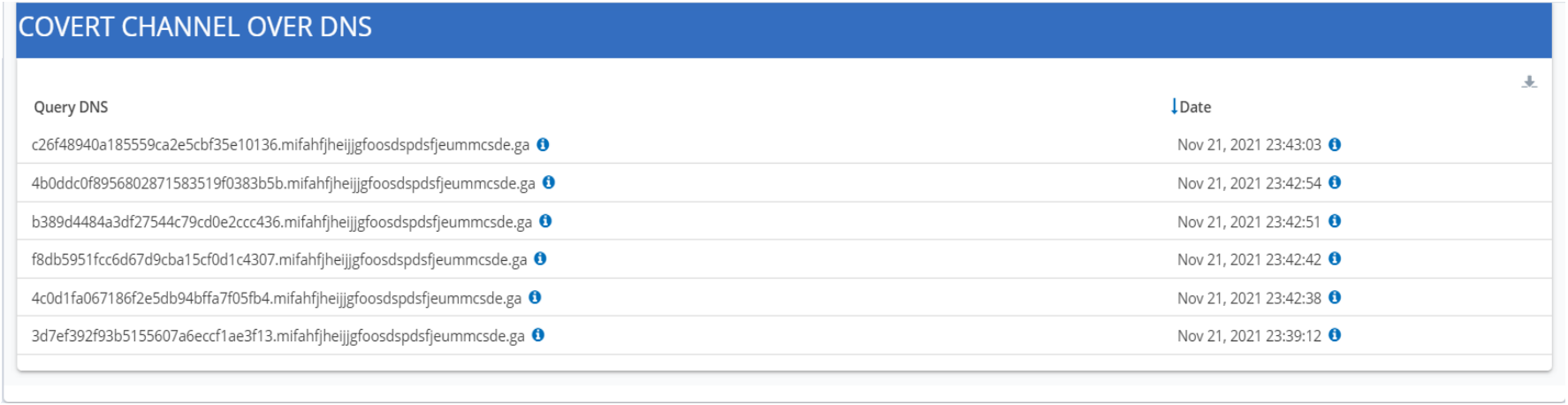
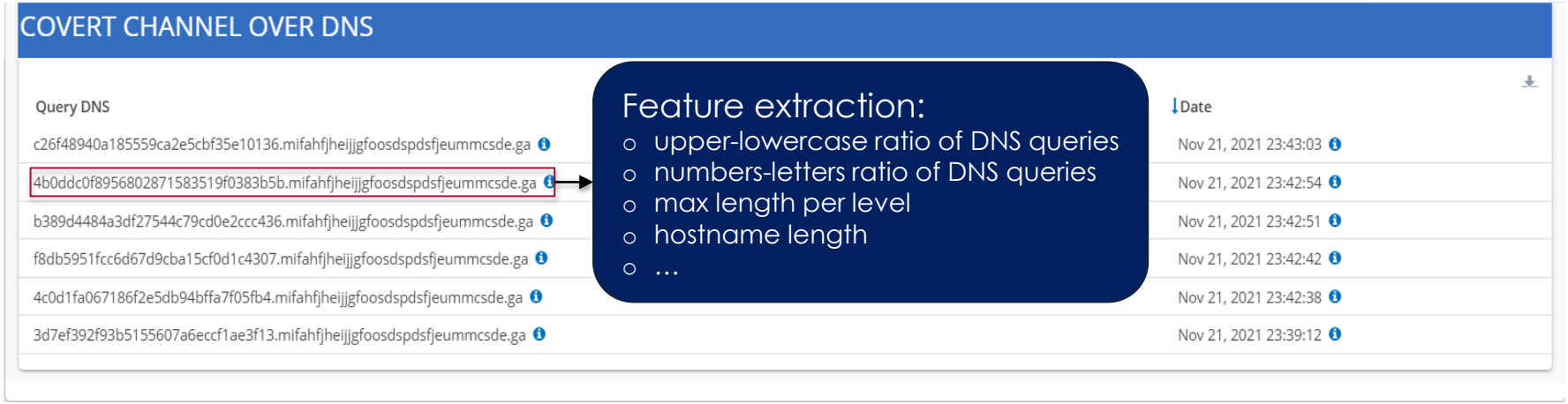


Figure: Examples of covert channel over DNS used by a sample belonging to the Astaroth malware family

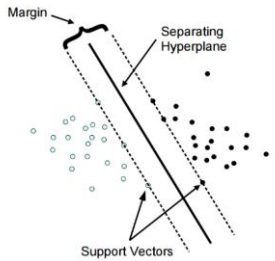
[9] Saeli S., Bisio F., Lombardo P., and Massa D. (2020). *DNS Covert Channel Detection via Behavioral Analysis: a Machine Learning Approach*. MALWARE conference, 2020.

Detect covert channels over DNS with ML



Detect covert channels over DNS with ML

- A one-class SVM classifier is trained with the features previously discussed, periodically extracted from legit DNS queries
- Features extracted from real-time DNS queries are given in input to the SVM to test if they are legitimate or anomalous
- Suspicious DNS queries are then aggregated with other anomaly indicators to classify them as malicious, such as:
  - high number of unique requests/hostnames to/per domain
  - high query entropy
  - high distance of monograms and bigrams distributions of the query from dictionary ones



Support Vector Machine

Detect anomalies in encrypted communications with ML

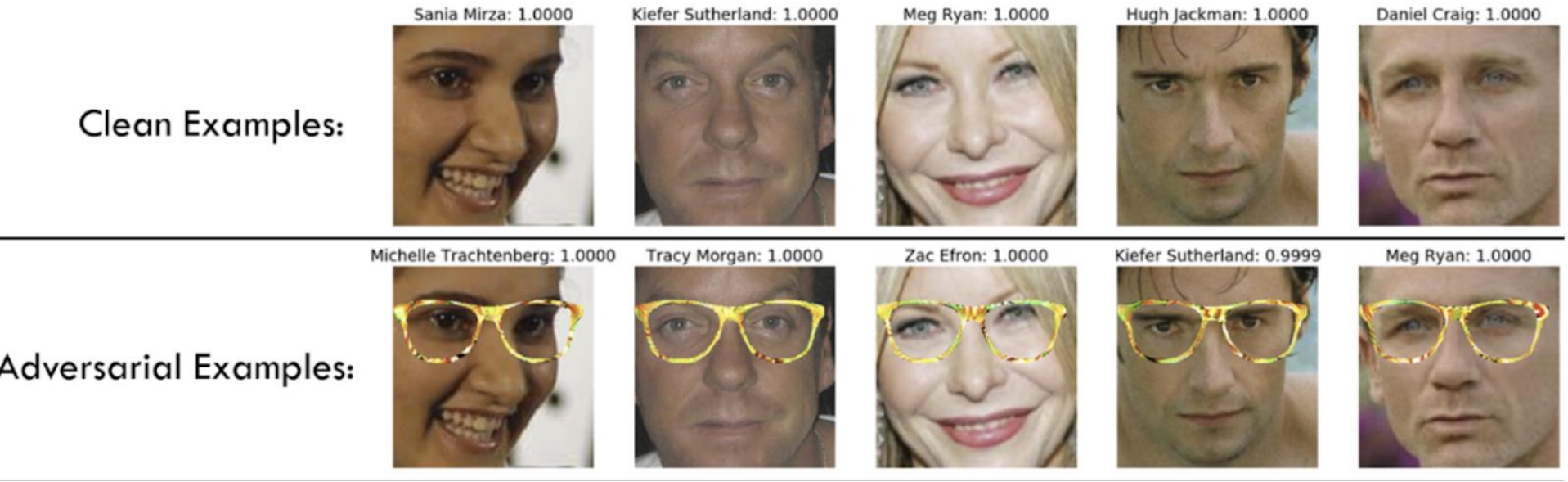
- Suspicious encrypted communications can be detected by analyzing and extracting features from exchanged secure protocol (e.g., HTTPS) messages
- The extraction phase processes SSL/TLS metadata and data contained in the fields of X.509 certificates to detect anomalies in a SSL/TLS handshake
- Suspicious connections are identified by combining unsupervised ML technique with JA3 hashes
- A JA3 hash is a client fingerprint of a SSL/TLS flow


```
{
  "version" : "TLSv12",
  "server_name" : "teams.microsoft.com",
  "curve" : "secp384r1",
  "subject" : "CN=teams.microsoft.com",
  "issuer" : "CN=Microsoft RSA TLS CA 01,
              O=Microsoft Corporation,C=US",
  "server_cert_chain" : [
   {
    "md5" : "28211f1f8a50966b518ec39d3546d57d",
    "sha1" : "4a263f1f39dd526901987ecdb09e2d1297e2bc51",
    "x509" : {
        "version" : 3,
        "key_type" : "rsa",
        "key_alg" : "rsaEncryption",
        "key_length" : 2048,
        "sig_alg" : "sha256WithRSAEncryption",
        "not_valid_before" : 1606847889.0,
        "not_valid_after" : 1638383889.0,
        "subject" : "CN=teams.microsoft.com",
        "issuer" : "CN=Microsoft RSA TLS CA 01,
                    O=Microsoft Corporation,C=US",
    }
   }
  ],
  "ja3" : "7f805430de1e7d98b1de033adb58cf46",
  "ja3s" : "0f14538e1c9070becdad7739c67d6363",
  "cipher" : "TLS_ECDHE_RSA_WITH_AES_256_GCM_SHA384",
  "machineDest" : "TEAMS.MICROSOFT.COM"
}
```

Detect anomalies in encrypted communications with ML

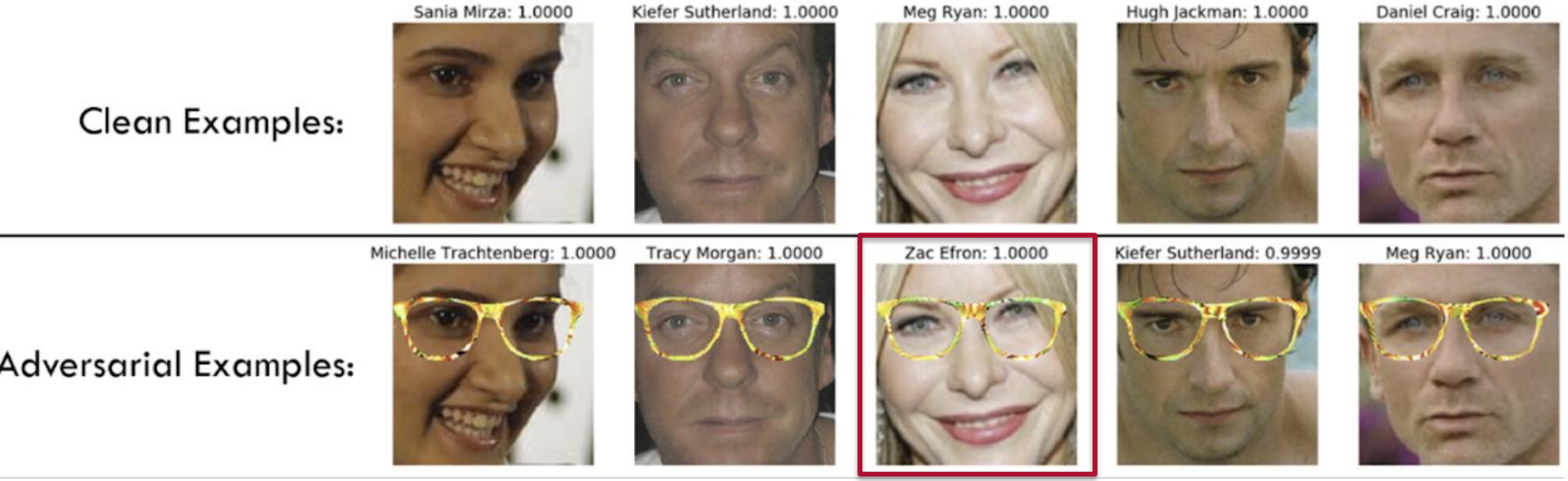
- Suspicious encrypted communications can be detected by analyzing and extracting features from exchanged secure protocol (e.g., HTTPS) messages
- The extraction phase processes SSL/TLS metadata and data contained in the fields of X.509 certificates to detect anomalies in a SSL/TLS handshake
- Suspicious connections are identified by combining unsupervised ML technique with JA3 hashes
- A JA3 hash is a client fingerprint of a SSL/TLS flow

Offensive AI



Source: Inovex.de. *Robustifying Machine Perception for Image Recognition Systems: Defense Against the Dark Arts*. 2019. https://www.inovex.de/de/blog/machine-perception-face-recognition/

Offensive AI



Source: Inovex.de. *Robustifying Machine Perception for Image Recognition Systems: Defense Against the Dark Arts*. 2019. https://www.inovex.de/de/blog/machine-perception-face-recognition/

Thanks for your attention

Contacts:

daniele.ucci@aizoongroup.com

Follow us on Twitter: @aizoongroup