NETWORK-BASED ANOMALY DETECTION
DNS NINJA & ABC TOOL
Irina Chiscop
Advanced attack detection by anomaly detection on system log data, network traffic & flow data, using data analytics (incl. AI and deep learning); enable detecting attacks in encrypted traffic.
FOCUS ON INTERNAL NETWORK PATTERNS IN COMBINATION WITH ANOMALY DETECTION

- Advanced targeted attacks typically consist of attack phases as shown in the lifecycle visualization.

- Most commercial monitoring and detection products focus on inbound network traffic & rule-based detection.

- Blind spot: analyzing internal network patterns in combination with anomaly detection.
ASSUMPTION: INTRUSION ALREADY TOOK PLACE

Cyber threats

IDS/IPS

Internal traffic

#2 ABC tool

#1 DNS NINJA

#3 Risky external connections

Company network

DNS traffic

SSL traffic

...
WHY DNS?

- Malicious software often tries to contact the outside world.
- DNS is commonly used for finding external servers ➔
- Infections and attacks within a company network may be detected better / quicker if we take DNS queries into account.
  - Examples: botnets, command and control traffic, ransomware, key loggers, DNS tunneling, …
  - Wannacry
    - luqerfsodp9ifjaposdfjhgosurijfaewrwergwea.com
    - Ifferfsodp9ifjaposdfjhgosurijfaewrwergwea.com
OUR SOLUTION: DNS NINJA
EXTRA LINE OF DEFENSE, SPECIALIZED IN DNS

DNS queries also provides useful context information during security investigations: what was going on on a workstation?
DNS NINJA OVERVIEW

Real-time filtering & classification of DNS queries

Real-time detection of “pseudo random” host and domain names (DGA) & puny code confusable

Alerts with “anomaly-scores” to ArcSight SIEM for correlation with other information/events

Web-based user interface for inspection and analysis of detected anomalies

Detecting deviations in statistical properties of DNS traffic of individual hosts
First the DNS queries of each machine are divided over 21 different (sub)classes: internal, whitelisted, reverse lookup, NXDOMAIN, DGA, MX, SOA, etc.

Detector #1 detects computers of which the DNS traffic statistics deviate too much of normal random human behavior for one or more of the 21 classes.

Each dot in the graph represents the DNS queries of a computer over a number of hours.
HOW THE ANOMALY DETECTION WORKS (EX. 2/3)

- Detector #2 detects unlikely domain names (based on length, frequencies of letter combinations, and dictionaries)

<table>
<thead>
<tr>
<th>Time</th>
<th>Type</th>
<th>Domain</th>
<th>Anomaly Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>08:47:15</td>
<td>A</td>
<td>rgtryhbgddtyh.biz</td>
<td>422 A</td>
</tr>
<tr>
<td>08:49:53</td>
<td>A</td>
<td>wertdghbyrukl.ch</td>
<td>420 A</td>
</tr>
<tr>
<td>10:06:40</td>
<td>A</td>
<td>wertdghbyrukl.ch</td>
<td>418 A</td>
</tr>
<tr>
<td>11:35:19</td>
<td>A</td>
<td>wertdghbyrukl.ch</td>
<td>415 A</td>
</tr>
</tbody>
</table>

Example domain names and anomaly scores

Real malware
Detector #3 detects computers of which the DNS queries show a too regular pattern for one or more of the 21 classes.

- Periodic (class: DGA) – Real malware
- Highly periodic (class: NXDOMAIN)
WEB-BASED USER INTERFACE
ASSUMPTION: INTRUSION ALREADY TOOK PLACE

Cyber threats

IDS/IPS

Internal traffic

#2 ABC tool

Company network

#1 DNS NINJA

#3 Risky external connections

DNS traffic

SSL traffic

Internal traffic
INTERNAL NETWORK PATTERNS CAUSED BY TARGETED ATTACKS

Figure VII.3: Examples of how internal network structures may change. Hosts in the secure zone do not have external Internet access, hosts outside the secure zone have. Red hosts indicate infected hosts, red arrows indicate observable, changed behaviour caused by the corresponding APT characteristic.
IDEA: LOOK AT CLUSTERS OF HOSTS

› Individual host behave relatively unpredictable

› Groups of hosts often behave similarly

› Determine normal behaviour per cluster
  › Observe when individual host deviates

› Group hosts that often communicate with each other using community clustering
  › More robust to minor changes
  › More efficient
PROCESS STEPS OF ANOMALY BASED CLUSTERING

1 Data selection
- Internal DNS logs
- Internal NetFlow
- File access logs

2 Data parsing
- Towards generic input
- Remove corrupt lines
- Filtering

3 Clustering & Network visualization
- Non-anomalous training data
- Potentially anomalous test data
- Visualization
  - PDF
  - Interactive

4 Clustering & Cluster modelling
- Generic data format
- Non-anomalous training data
- Model the normal internal network patterns of clustered hosts

5 Anomaly detection
- Inter-cluster communication models
- Compare test data to the models
- Anomaly list with three anomalous pattern types

6 Anomaly inspection
- Why is a host identified as anomalous
- Context of a host

Data describes internal network patterns of hosts
- Deviating patterns may indicate attack

Deviating patterns may indicate attack

Towards generic input
- Remove corrupt lines
- Filtering

Compare test data to the models
- Anomaly list with three anomalous pattern types

Why is a host identified as anomalous
- Context of a host
CLUSTERING & CLUSTER MODELLING
MODELLING THE INTER-CLUSTER COMMUNICATION BEHAVIOUR

After community clustering, each host is within a cluster

Cluster modelling:
- Each cluster has its own *inter-cluster communication model* based on a statistical analysis
  - Communications within a community cluster are considered normal
  - Communications between different community clusters are modelled

Example of a simplified inter-cluster communication model

Network-based Anomaly Detection

26 August 2019
CLUSTERING & CLUSTER MODELLING
MODELLING THE INTER-CLUSTER COMMUNICATION BEHAVIOUR

Example of a simplified inter-cluster communication model

Cluster A
- H1
- 100 connections per day

Cluster B
- H2
- H3

Cluster C
- H4
- 20 connections per day

Example of anomalous host behaviour

Cluster A
- H1

Cluster B
- H2
- H3

Cluster C
- H4
- 70 connections per day

Anomaly!

Cluster C
- H4
- 80 connections per day

Network-based Anomaly Detection
Visualizations of the cluster communication model and the behaviour of the anomalous host indicate why the host is identified as anomalous.
CONCLUSIONS AND TAKEAWAYS

- **Internal network traffic:**
  - Valuable source of information for targeted attack detection

- **Behavioural based clustering techniques:**
  - Improve anomaly detection algorithms
THANK YOU FOR YOUR ATTENTION

Contact:
Irina Chiscop, Scientist Innovator
irina.chiscop@tno.nl