# Detecting Malicious Hosts Using Traffic Flows

Miguel Pupo Correia
joint work with Luís Sacramento
NavTalks, Lisboa, June 2017

## Outline

- Motivation
- Approach
- Evaluation
- Conclusion

# Outline

- **Motivation**
- Approach
- Evaluation
- Conclusion

3

# Motivation

- Scenario:
  - Large national telco/ISP connected to its own provider
  - Huge amount of traffic in/out, much is encrypted
  - Possibly new attacks / new variants


4

# Motivation

- Compromised hosts do attacks such as:
  - Distributed denial of service attacks
  - Exfiltrating confidential data
  - Sending spam
  - Mapping the network
  - Contact bot command&control centers
  - etc.

5

# Network Intrusion Detection Systems

- Traditional NIDSs:
- Knowledge-based: require signatures of attacks
  - Not good for new attacks
- Behavior-based: require clean traffic for training
  - Where to get it with our scenario?
- Most do deep packet inspection, unfeasible with too much traffic

6

## Outline

- Motivation
- **Approach**
- Evaluation
- Conclusion

7

## Our approach

- Detection framework to detect malicious hosts based on real traffic
- Not knowledge-based, to avoid need for signatures
- Not behavior-based, as no training traffic exists
- No deep packet inspection, as it is slow
- Detects hots doing new attacks or new variants

8

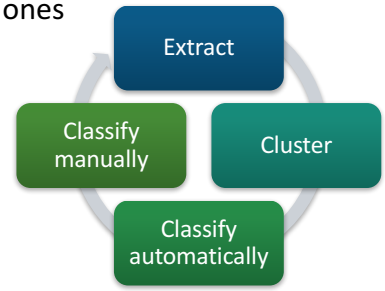
# Key ideas

- Collect traffic data summarized as network flows
- Extract data about hosts from flows
- Use unsupervised machine learning / clustering to
  - get information that humans can understand without previous knowledge about attacks
- Use supervised machine learning / classifier to automatically assign clusters to classes/categories
  - ex: web servers, mail servers, hosts sending spam, hosts doing distributed denial of service,…
- Manually label new clusters

9

# The approach

- Loop:
  - Collect flows for a period of time (e.g., 1 day)
  - Extract from the flows data about hosts with MapReduce
  - Use clustering to create groups of hosts
  - Use classifier to automatically classify hosts
  - Manually label remaining ones
  - Repeat for next period




10

# The approach

- Loop:
  - Collect flows for a period of time (e.g., 1 day)

11

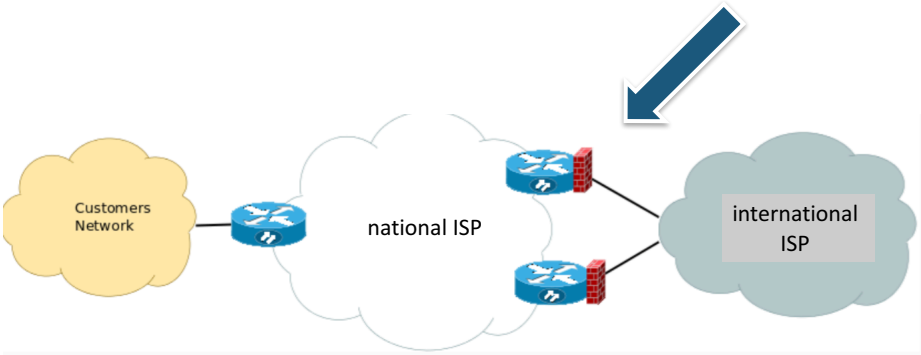
# Flows

- Flow: sequence of related packets observed during an interval of time
  - A flow is defined in terms of a subset of src IP, dest IP, protocol, src port, dest port; ex: *(*, 1.2.3.4, TCP, *, 80)*
- Netflow: monitoring approach created by Cisco
  - Idea is to capture data about network flows
  - Data: begin/end of flow timestamps, n. packets, n. bytes
  - Variants: IPFIX (standard based on Netflow 9), sFlow,…

12

# Flow collection

- Flows collected on NetFlow-enabled border routers




13

# The approach

- Loop:
  - Collect flows for a period of time (e.g., 1 day)
  - Extract from the flows data about hosts with MapReduce

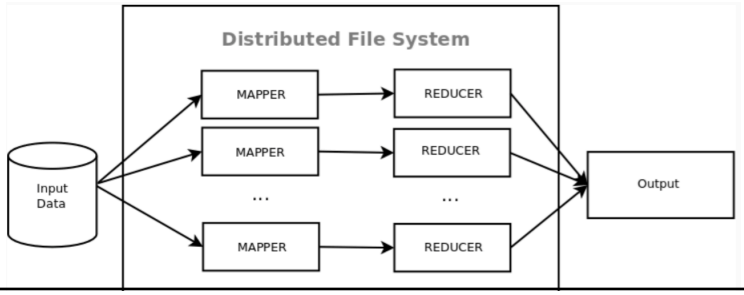
14

# Host data extraction

- Flow format:

  <Source IP, Destination IP, Source Port, Destination Port, Protocol, TCP Flags, #Bytes, #Packets, Duration>

- Use MapReduce for extracting data per host (IP)
  - aggregated by source or destination IP address




15

# Host data extraction

- Host features (data) extracted by MapReduce:

| Feature | Description |
|---|---|
| **Aggregation Key** | The IP address that will be used as an identifier, to which the below features relate to |
| **NumSIPs / NumDIPs** | The number of IP addresses contacted |
| NumSports | The number of different source ports contacted |
| **NumDport** | The number of different destination ports contacted |
| textbfNumHTTP | The number of packets to/from port 80 (HTTP) |
| **NumIRC** | The number of packets to/from ports 194 or 6667 (IRC) |
| **NumSMTP** | The number of packets to/from port 25 (SMTP) |
| **NumSSH** | The number of packets to/from port 22 (SSH) |
| **TotalNumPkts** | The total number of packets exchanged |
| PktRate | The ratio of the number of packets sent and its duration |
| ICMPRate | The ratio of ICMP packets, and total number of packets |
| SynRate | The ratio of packets with a SYN flag and the total number of packets |
| **TotalNumBytes** | The overall sum of bytes |
| **AvgPktSize** | The average packet size |
| BadSubnet | This field expresses whether the IP address belongs to a blacklisted subnet |
| MaliciousIP | This field expresses whether the IP address is blacklisted |
| OpenVaultBlacklistedIP | Same as the above, but checked from a trusted and well know threat database |
| MaliciousASN | This field shows if the IP address belongs to a blacklisted ASN |
| LocationCode | Code for the country associated with the address |

extracted from the flows directly

based on threat intelligence

# The approach

- Loop:
  - Collect flows for a period of time (e.g., 1 day)
  - Extract from the flows data about hosts with MapReduce
  - Use clustering to create groups of hosts

17

# Unsupervised ML / clustering

- Idea: group similar hosts in clusters (sets)
- Why? Humans can understand and classify a few clusters, not zillions of hosts
- How?
  - Normalize every feature into range [0,1]
  - Run clustering algorithm, e.g., K-Means, to get **k** clusters
  - **k** can be defined, e.g., with the elbow method (finds the "elbow", i.e., when adding more clusters does not improve the modelling of the data)

18

# The approach

- Loop:
  - Collect flows for a period of time (e.g., 1 day)
  - Extract from the flows data about hosts with MapReduce
  - Use clustering to create groups of hosts
  - Use classifier to automatically classify hosts
  - Manually label remaining ones

19

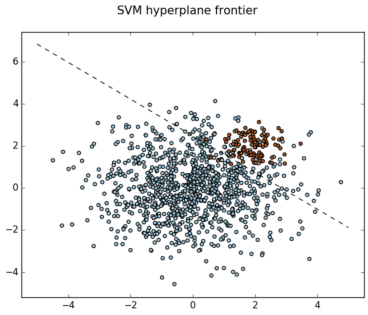
# Intrusion detection with flows

- Each cluster contains hosts with similar behavior
  - ex: web servers, mail servers, hosts sending spam, hosts doing denial of service,…
- What to do with them? (at cruise speed)
- Already seen? Use classifier to classify automatically
- Never seen?
  - Label manually, with help of the features' values
  - Focus attention on smaller clusters with odd feature distribution; often malicious
  - Retrain classifier

20

# Supervised ML / classification

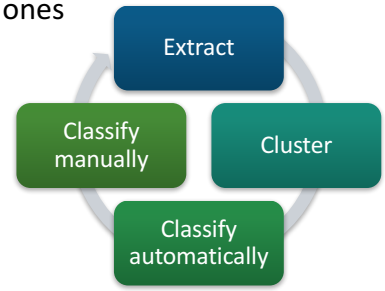
- Naïve solution: use labelled hosts to train a Binary Support Vector Machine (SVM) classifier
  - Samples/hosts classified as benign or malicious
  - Finds an hyperplane that separates samples
  - Classifies new samples (hosts)




21

# The approach

- Loop:
  - Collect flows for a period of time (e.g., 1 day)
  - Extract from the flows data about hosts with MapReduce
  - Use clustering to create groups of hosts
  - Use classifier to automatically classify hosts
  - Manually label remaining ones
  - Repeat for next period




22

# Outline

- Motivation
- Approach
- **Evaluation**
- Conclusion

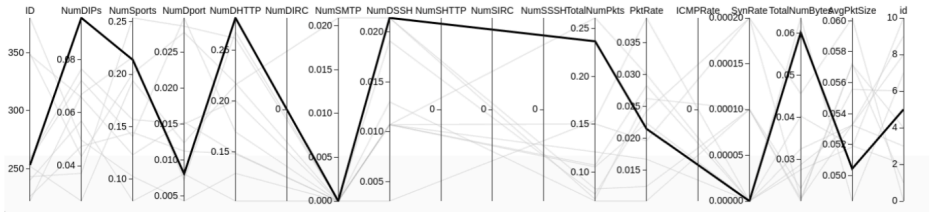
23

# Tool – interface

```
Welcome to the NIDS of the Future!

First of all, which dataset will you analyze?
 Source - 1
 Destination - 2
 $ 1

=======================================
||         SOURCE AGGREGATION         ||
||              ANALYSIS              ||
=======================================
 Please choose wich clustering technique you wish to apply
 MiniBatch - 1
 K-Means - 2
 $ 2

You are analyzing Monday, which contains 448 entries

You are now in the clustering section. What would you like to do?
 1.  See description of the dataset
 2.  Choose the optimal number of clusters
 3.  Visualize the clusters
 4.  Analyze clusters
 5.  Open web interface
 6.  Validate attacks (from database ground truth)
 7.  Plot two features
 8.  Change dataset
 9.  Change clustering type
 0.  Exit

 What will it be?
   $
```

24

# Tool – interactive visualiz. of cluster




Showing all 11 rows

| ID | NumDIPs | NumSpo... | NumDpo... | NumDHT... | NumDIRC | NumSMT... | NumDSS... | NumSHT... | NumSIRC | NumSSS... | TotalNu... | PktRate | ICMPRate | SynRate | TotalNu... | AvgPktSi... |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 222 | 0.066837... | 0.216327... | 0.007232... | 0.268031... | 0.0 | 0.0 | 0.010695... | 0.0 | 0.0 | 0.0 | 0.150463... | 0.019643... | 0.0 | 0.0002 | 0.045638... | 0.060180... |
| 226 | 0.047352... | 0.144363... | 0.011283... | 0.147856... | 0.0 | 0.0 | 0.012987... | 0.0 | 0.0 | 0.0 | 0.069767... | 0.010224... | 0.0 | 0.0001 | 0.020123... | 0.057228... |
| 240 | 0.056868... | 0.078877... | 0.008100... | 0.101520... | 0.0 | 0.0 | 0.003055... | 0.0 | 0.0 | 0.0 | 0.154001... | 0.038869... | 0.0 | 0.0 | 0.040303... | 0.051924... |
| 243 | 0.037157... | 0.253778... | 0.028641... | 0.261753... | 0.0 | 0.0 | 0.010695... | 0.0 | 0.0 | 0.0 | 0.107257... | 0.028268... | 0.0 | 0.0 | 0.028795... | 0.053265... |
| 252 | 0.026735... | 0.193790... | 0.012440... | 0.216015... | 0.0 | 0.020889... | 0.021390... | 0.0 | 0.0 | 0.0 | 0.071685... | 0.026250... | 0.0 | 0.0001 | 0.020653... | 0.057161... |
| **253** | **0.09561...** | **0.21371...** | **0.00810...** | **0.28184...** | **0.0** | **0.0** | **0.02139...** | **0.0** | **0.0** | **0.0** | **0.23679...** | **0.02151...** | **0.0** | **0.0** | **0.06019...** | **0.05043...** |
| 258 | 0.072048... | 0.140552... | 0.029798... | 0.169619... | 0.0 | 0.0 | 0.021390... | 0.0 | 0.0 | 0.0 | 0.082497... | 0.012506... | 0.0 | 0.0001 | 0.023106... | 0.055569... |
| 259 | 0.076353... | 0.156665... | 0.015044... | 0.200410... | 0.0 | 0.0 | 0.010695... | 0.0 | 0.0 | 0.0 | 0.261322... | 0.030012... | 0.0 | 0.0 | 0.063685... | 0.048351... |
| 348 | 0.065705... | 0.123078... | 0.015333... | 0.148753... | 0.0 | 0.0 | 0.010695... | 0.0 | 0.0 | 0.0 | 0.121116... | 0.016761... | 0.0 | 0.0 | 0.032508... | 0.053252... |
| 349 | 0.043501... | 0.200921... | 0.027773... | 0.217629... | 0.0 | 0.0 | 0.019098... | 0.0 | 0.0 | 0.0 | 0.077022... | 0.024301... | 0.0 | 0.0002 | 0.020687... | 0.053289... |

25

# Evaluation

- Two parts:
- Synthetic dataset (ISCX)
  - Designed for IDSs
  - Flows are labelled
  - Allows validating the approach
- Real dataset collected at the telco
  - No ground truth

26

# ISCX dataset evaluation

| Clusters | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|
| # Entries | 1 | 5 | 1 | 1 | 5 | 5 | 16 | 8 | 3 | 1 |
| Features | | | | | | | | | | |
| 1 | - | - | - | - | - | - | - | - | - | - |
| 2 | - | 0.125 ; 0.023 | - | 1.0 | 0.243 ; 0.064 | - | - | 0.200 ; 0.016 | 0.277 ; 0.076 | 0.115 |
| 3 | - | - | - | - | - | - | - | - | - | 0.227 |
| 4 | - | 0.131 ; 0.023 | - | - | 0.418 ; 0.56 | - | - | 0.213 ; 0.031 | - | - |
| 5 | - | - | - | - | - | - | - | - | - | - |
| 6 | - | - | - | 0.325 | - | - | - | - | - | - |
| 7 | - | - | 1.0 | - | - | - | - | - | - | - |
| 8 | - | - | - | - | - | - | - | - | - | - |
| 9 | - | - | - | - | - | - | - | - | - | - |
| 10 | - | - | - | - | - | - | - | - | - | - |
| 11 | - | - | - | - | - | - | - | - | - | - |
| 12 | - | - | - | - | - | - | - | - | - | - |
| 13 | - | - | - | - | - | 0.305 | - | - | - | - |
| 14 | - | - | - | - | - | - | - | - | - | - |
| 15 | - | - | - | - | - | - | - | - | - | - |
| 16 | - | - | - | - | - | - | - | - | 0.128 ; 0.005 | 0.184 |

Avg ; StdDev

Low values

Saturday clusters

- Brute-Force SSH attack found during this day (cluster 3)
  - Maximum for SSH connections (and high, not seen in table)

27

# Telco dataset evaluation

| Cluster # | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Hosts | 1605 | 51773 | 6485 | 13305 | 529 | 1730 | 1729 | 21507 | 8523 | 8522 | 1498 | 4686 | 10 | 5653 |
| Features | | | | | | | | | | | | | | |
| 1 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 3 | - | - | - | - | - | - | - | - | - | - | - | - | 0.368 ; 0.138 | - |
| 4 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 5 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 6 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 7 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 8 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 9 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 10 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 11 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 12 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 13 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 14 | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 15 | - | - | - | - | - | - | - | - | - | - | - | - | 0.61 ; 0.208 | - |
| 16 | - | - | - | - | - | | - | - | - | - | - | - | - | - |
| 17 | - | - | - | - | - | | - | - | | - | - | - | - | - |

Cluster data with source aggregation key (i.e., aggregated by IP inside the ISP) cluster data – 1st part

28

# Telco dataset evaluation

| Cluster # | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| # Hosts | 1 | 824 | 5 | 4606 | 1864 | 1676 | 12 | 107 | 13 | 2233 | 2233 | 8091 | 10 | 13897 | 16843 | 35 |
| Features | | | | | | | | | | | | | | | | |
| 1 | 1.0 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 2 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 3 | 1.0 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 4 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 5 | 0.667 | - | - | - | - | - | - | - | 0.384 ; 0.184 | - | - | - | - | - | - | - |
| 6 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | 0.237 ; 0.158 |
| 7 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 8 | 1.0 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 9 | - | - | - | - | - | - | 0.542 ; 0.18 | - | - | - | - | - | - | - | - | - |
| 10 | - | - | 0.626 ; 0.21 | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 11 | 1.0 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 12 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 13 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 14 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 15 | 0.843 | - | - | - | - | - | - | - | - | - | - | - | - | - | - | - |
| 16 | - | - | - | - | - | | 0.261 ; 0.06 | - | - | - | - | - | - | - | - | - |
| 17 | - | - | - | - | - | | - | - | | - | - | - | - | - | - | - |

Cluster data with source aggregation key – 2nd part

29

# Source aggregation key – cluster 15

i.e., aggregated by IP inside the ISP

| Cluster # | 15 |
|---|---|
| # Hosts | 1 |
| Features | |
| 1 | 1.0 |
| 2 | - |
| 3 | 1.0 |
| 4 | - |
| 5 | 0.667 |
| 6 | - |
| 7 | - |
| 8 | 0.843 |

| Cluster # | 15 |
|---|---|
| # Hosts | 1 |
| Features | |
| 9 | - |
| 10 | - |
| 11 | 1.0 |
| 12 | - |
| 13 | - |
| 14 | - |
| 15 | 0.843 |
| 16 | - |

- Spammer or denial of service (?)
  - High connectivity to various users, many ports, receiving communication on IRC port, communication through HTTP, high number of packets sent, high number of bytes

30

# Source aggregation - Cluster 21

| Cluster # | 21 |
|---|---|
| # Hosts | 12 |
| Features | |
| 1 | - |
| 2 | - |
| 3 | - |
| 4 | - |
| 5 | - |
| 6 | - |
| 7 | - |
| 8 | - |

| Cluster # | 21 |
|---|---|
| # Hosts | 12 |
| Features | |
| 9 | 0.541 ; 0.181 |
| 10 | - |
| 11 | - |
| 12 | - |
| 13 | - |
| 14 | - |
| 15 | - |
| 16 | 0.261 ; 0.026 |

- Bot communicating with C&C server
  - High IRC communication + high average packet size
  - Confirmed by accessing the IP of the C&C server

31

# Telco dataset evaluation summary

| Cluster # | Aggregation Key | Highlighted Features | Type of Attack |
|---|---|---|---|
| 15 | Source | 1, 3, 5, 8, 11, 15 | Spam / DoS |
| 16 | Destination | 1, 3, 6 | DoS |
| 17 | Source | 10 | Brute-Force SSH |
| 20 | Destination | 1, 2, 15 | Network Scan |
| 21 | Source | 9, 16 | Botnet Communication |
| 22 | Destination | 1, 3, 8, 15 | Web Application Probing |
| 27 | Source | 1, 2, 5, 8, 11, 15 | DDoS IRC Botnet |
| 29 | Destination | 1, 2, 4, 11, 15 | DDoS Botnet |

32

## Outline

- Motivation
- Approach
- Evaluation
- **Conclusion**

33

## Conclusion

- Network Intrusion Detection for identifying malicious hosts using flows
- ...without having to say how entities misbehave
- Use clustering (unsupervised ML) to reduce the size of the problem and
- a classifier (supervised ML) to automatize classification
- Keep humans in the loop; mandatory w/evolving threats
- Detects attacks involving many packets, not low traffic attacks like buffer overflows or SQL injection

34

U LISBOA | UNIVERSIDADE DE LISBOA

# Thanks! Questions?

This work was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UID/CEC/50021/2013 (INESC-ID)

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA CIÊNCIA, TECNOLOGIA E ENSINO SUPERIOR

TÉCNICO LISBOA

inesc id lisboa