What botnet characteristics can be used for distributed and collaborative network based botnet detection?

Christian Dietz
christian.dietz[at]unibw.de

Forschungszentrum
Cyber Defence
Universität der Bundeswehr München

Agenda

- Introduction
- The botnet phenomenon
- State of the art in botnet detection
- Research approach
- Early results
- Conclusions
- Outlook

Introduction

"Cyber crime costs global economy $445 billion a year!

…

About 40 million people in the United States, roughly 15 percent of the population, has had personal information stolen by hackers, it said, while high-profile breaches affected 54 million people in Turkey, 16 million in Germany and more than 20 million in China."

Source: Reuters, London, Mon Jun 9, 2014

Introduction

Botnets:

- Provide infrastructure for various cyber criminal activities e.g. SPAM, DDoS, financial fraud, data theft, extortion

Introduction

Botnets:

- A botnet is network of malware infected hosts under the command of a botmaster.
- Command and control infrastructure (C&C): IRC, HTTP, P2P

Introduction

Botnets from the perspective of a criminal:

- *Cost for botnet setup: $350-$400*
- *Infection/spreading services, under $100 per a thousand installs*
- *Botnets and rental, Direct Denial of Service (DDoS), $535 for 5 hours a day for one week, email spam, $40 per 20,000 emails, and Web spam, $2 per thirty posts.*

Source: http://resources.infosecinstitute.com/2013-impact-cybercrime/

Introduction

Botnets from the victim perspective:

"Too little is done in many countries to prevent cybercrime.

While the majority of companies have the important security building blocks, such as firewalls and IPS, needed for their security infrastructure, ***less than half of organizations in this study have advanced protections to fight botnets*** *and APTs."*

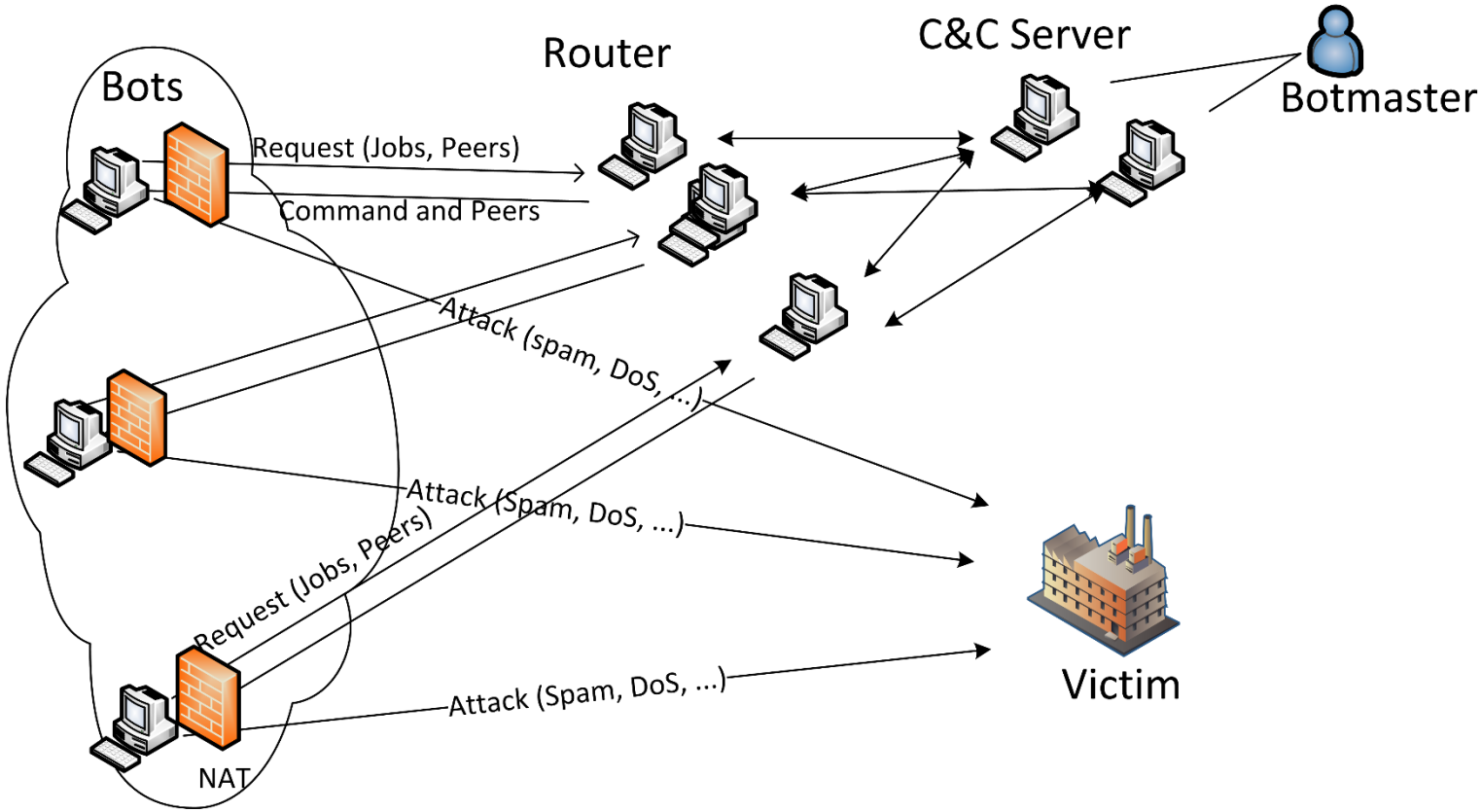
Source: Ponemon Institute, 2012, The Impact of Cybercrime on Business,
Link: http://www.ponemon.org/local/upload/file/Impact_of_Cybercrime_on_Business_FINAL.pdf

Introduction

What
botnet characteristics
can be used for
distributed and **collaborative**
network based botnet
detection?

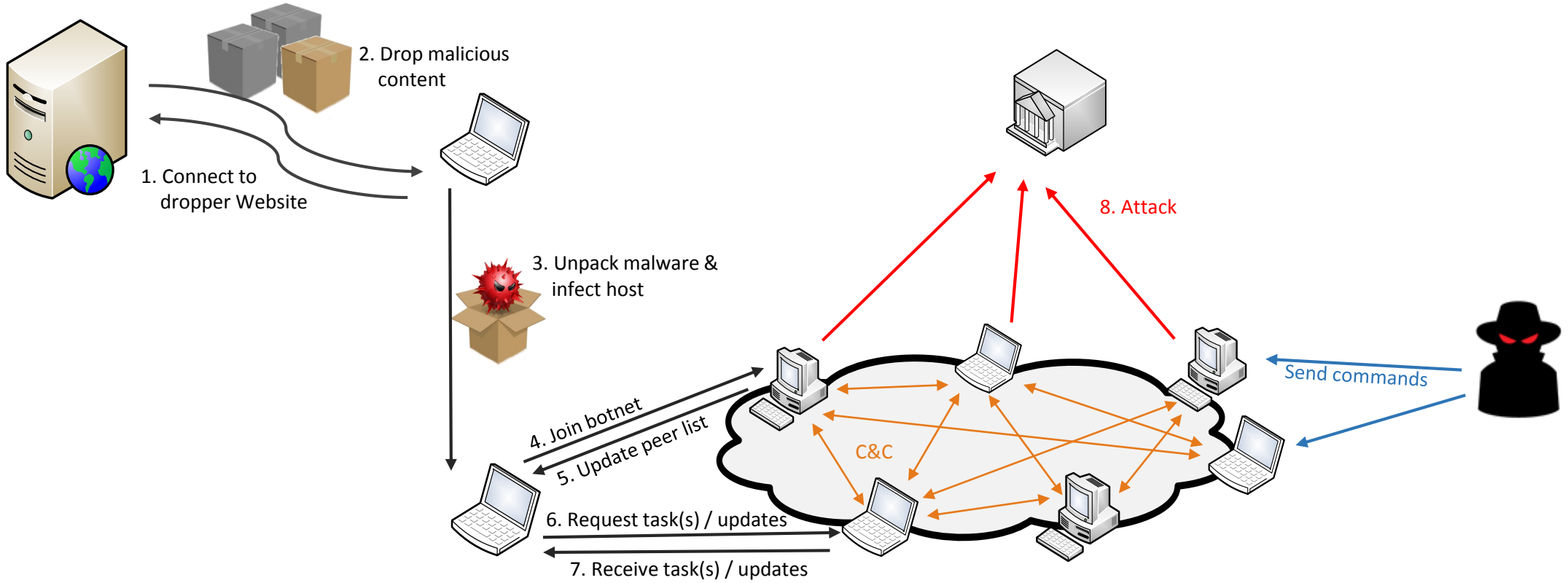
The botnet phenomenon

- Structure of botnet based attack



The botnet phenomenon

- Phases in a bot(net) life-cycle



The botnet phenomenon

Challenges:


Malware updates


P2P


Fast-Flux


Encryption


Big Data


Anonymization (TOR, …)

Icon source: *https://www.iconfinder.com/*

What
botnet characteristics
can be used for
distributed and **collaborative**
network based botnet
detection?

State of the art

- Knowledge based
- Anomaly based

State of the art

Knowledge based:

- Rules
- Signatures
- Filters
- Experts

Drawbacks:

- Only detect known threats
- Many rules necessary
- Over-fitted per definition

State of the art

Anomaly based:

- Computational Intelligence / Machine learning
- Training/Validation data
- Detect unknown attacks
- Can be automatically adapted in case of changed conditions

Drawbacks:

- False positives
- Availability of data
- Understanding and controllability of the detection process

What
botnet characteristics
can be used for
distributed and **collaborative**
network based botnet
detection?

State of the art

- Data types and formats
- Data sources

State of the art

- Data types and formats:
  - IP packets
  - Threat information
  - Log-files
  - Flows
- Data sources:
  - Sinkholes
  - Darknets
  - Real-life network environments
  - Simulations
  - Reverse engineering
  - …

State of the art in botnet detection

- Definition flow:
  - See RFC 3954: NetFlow v9

> “An IP Flow, also called Flow, is defined as a **set of IP packets** passing an observation point in the network during a certain time interval. **All packets that belong to a particular Flow have a set of common properties** derived from the data contained in the packet and from the packet treatment at the observation point.

State of the art

- Definition flow as used in this research:

  Set of records observed over a certain period of time sharing connection information plus a set of common properties derived from the data contained in the records captured at a network observation point.

State of the art

- Definition flow as used in this research:

  Set of records observed over a certain period of time sharing connection information plus a set of common properties derived from the data contained in the records captured at a network observation point.

- Examples:
  - Sinkhole trace:

    [1] "[2012-09-09 07:01:28.64365] bootstrap request from 81.214.XXX.XXX:1064"

    [2] "[2012-09-09 07:01:29.17498] bootstrap request from 81.214.XXX.XXX:1067"

    [3] "[2012-09-09 07:01:29.37554] job request from 89.29.XXX.XXX:3265 - 0c274f674d8347509234a088d359df49, v126 \"relqq26\", os info: 5.1.2600, platform 2)"

  - Netflow:

    [1] 2012-09-09 07:01:28.64365, 2012-09-09 07:01:48.44789, 1.1.1.1, 8.8.8.8, 1234, 80, 10, 984, 0, 0, .A...., UDP, ...

What
botnet characteristics
can be used for
distributed and **collaborative**
network based botnet
detection?

Research approach

- Hypothesis:
  - Botnet phenomenon -> global problem!
  - Ideal cure would be deployed on a global scale!

- Consequence:

Cooperation based on **distributed** measures and detection is the key to detect botnets and proactively stop various cyber-criminal activities!

Research approach

- What is hindering us to use the detection approaches and technology available today to achieve this goal?

Research approach

- What is hindering us to use the detection approaches and technology available today to achieve this goal?
  - Privacy concerns?
  - Big Data?
  - Availability of data?
  - Heterogenous behavior and environments!

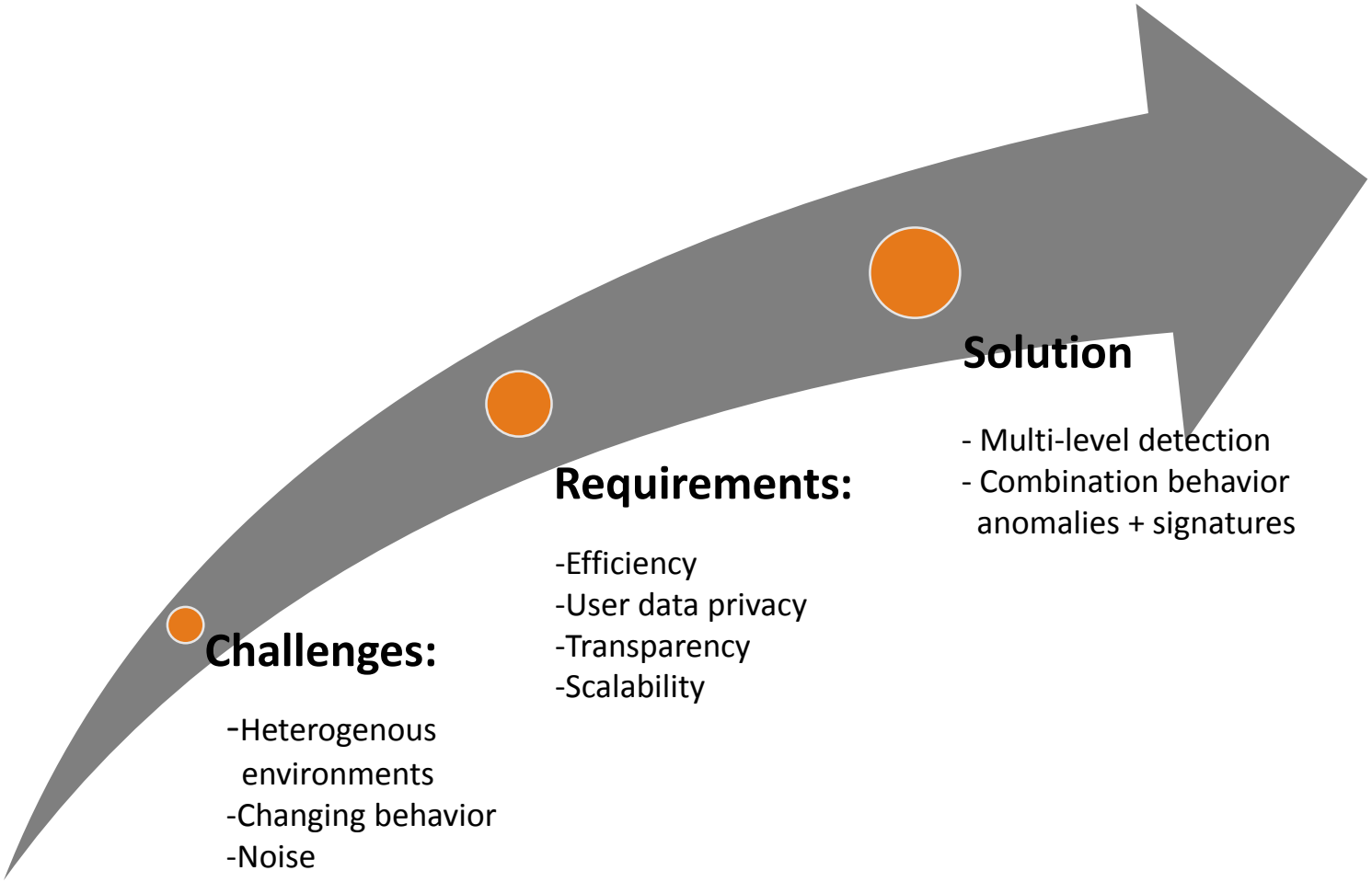
    -> Standardized formats do not guarantee standardized behavior/noise description!

    -> Exchange needs standardized formats and protocols plus normalized/standardized behavior descriptors!

Research approach

- How could we fix this?

Research approach



Challenges:

-Heterogenous environments
-Changing behavior
-Noise

Requirements:

-Efficiency
-User data privacy
-Transparency
-Scalability

Solution

- Multi-level detection
- Combination behavior anomalies + signatures

Cooperative detection:

- Noise reduction
- Behavior normalization
- Distributed measurements

Research approach

- Consequence:

  Focus on understanding the source of noise and normalization of changing heterogenous influences!

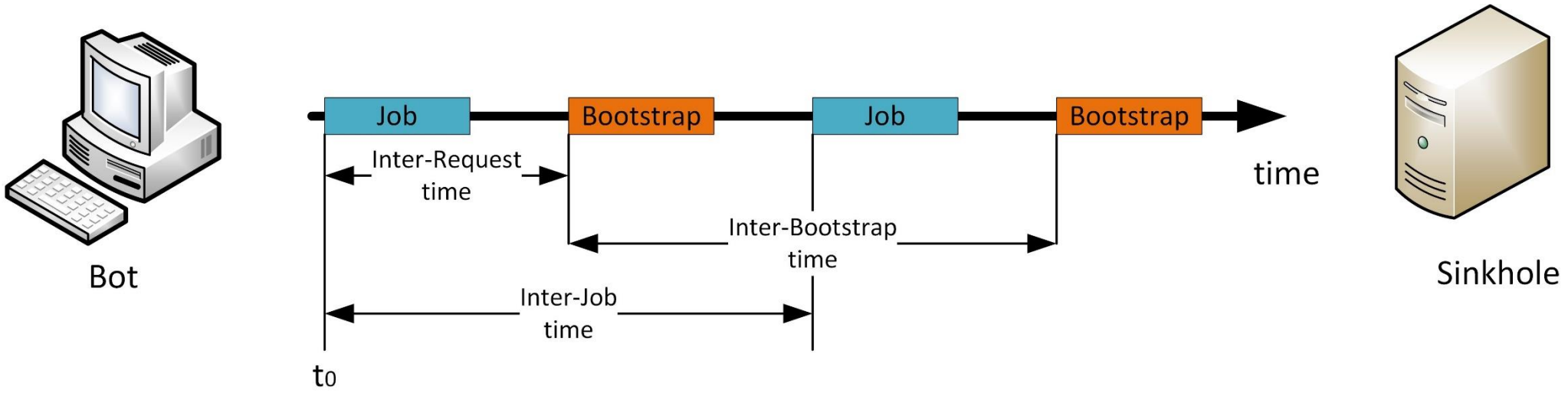
What
botnet characteristics
can be used for
distributed and **collaborative**
network based botnet
detection?

Early results

- Based on botnet traces captured from a sinkhole (Hlux/Kelihos.B case)
- Type of bot: Peer-to-peer (P2P)
- Capture period: 1 month
- Parts of botnet life-cyle observed:
  - Bootstrap-requests: Search for peers
  - Job-request: Ask for actions to perform (sending SPAM,DDoS, Bitcoin-mining)

Early results

- What we measured:

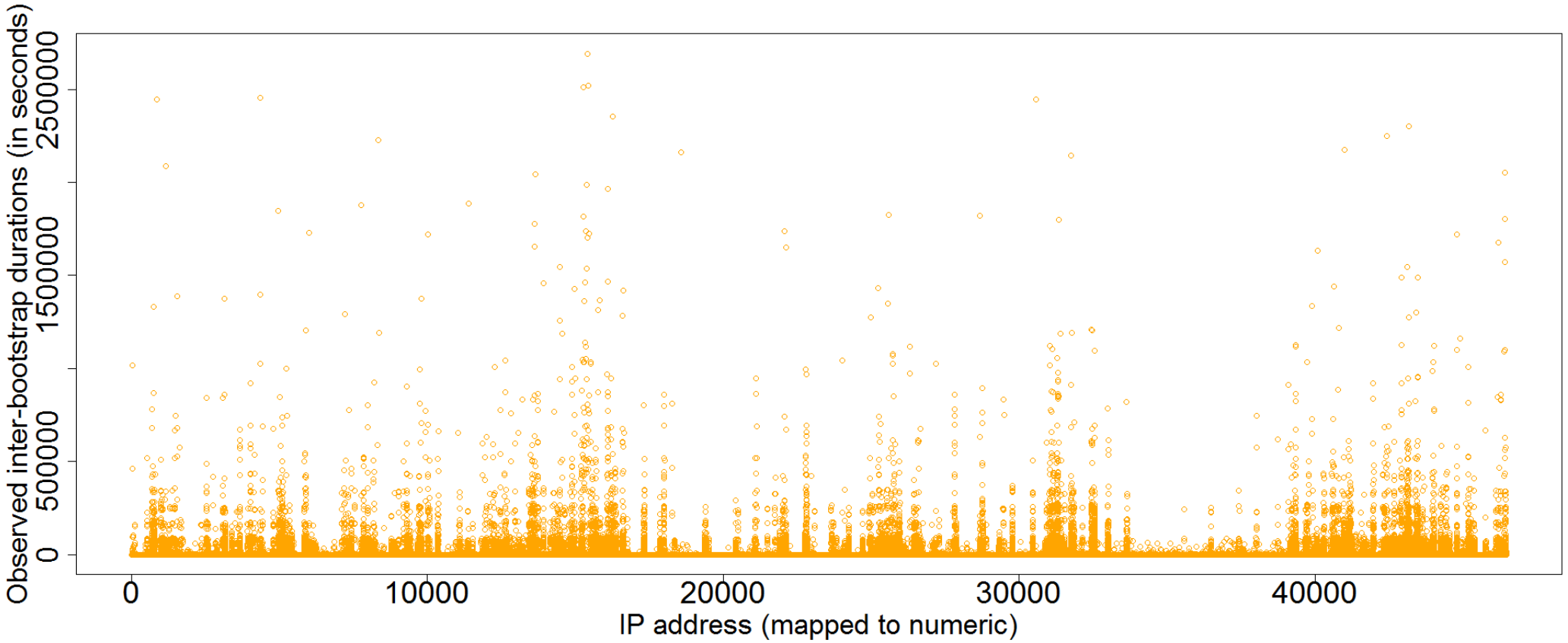


Early results

- What we observed:
  - Regular behavior
  - Difference in behavior for bootstrap and job request
  - Changing behavior characteristics with newer versions

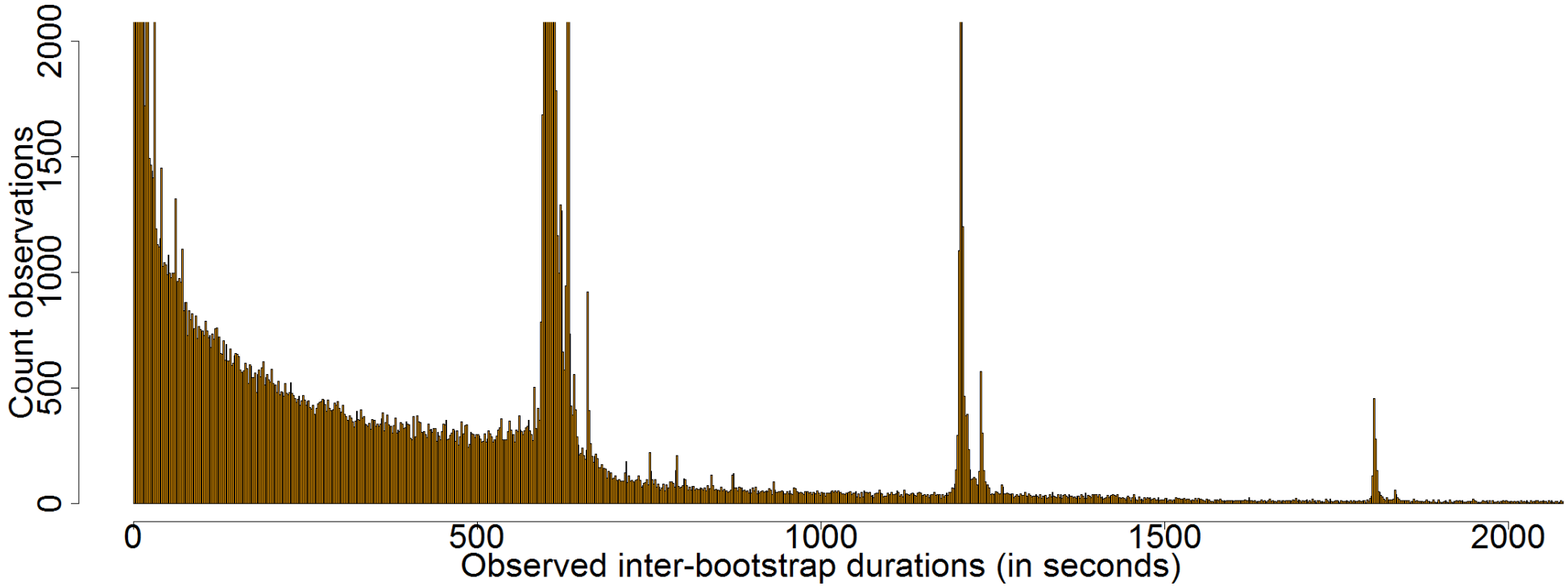
Early results

- Short inter-request times



Early results

- Regular behavior
- Difference in behavior for bootstrap and job request

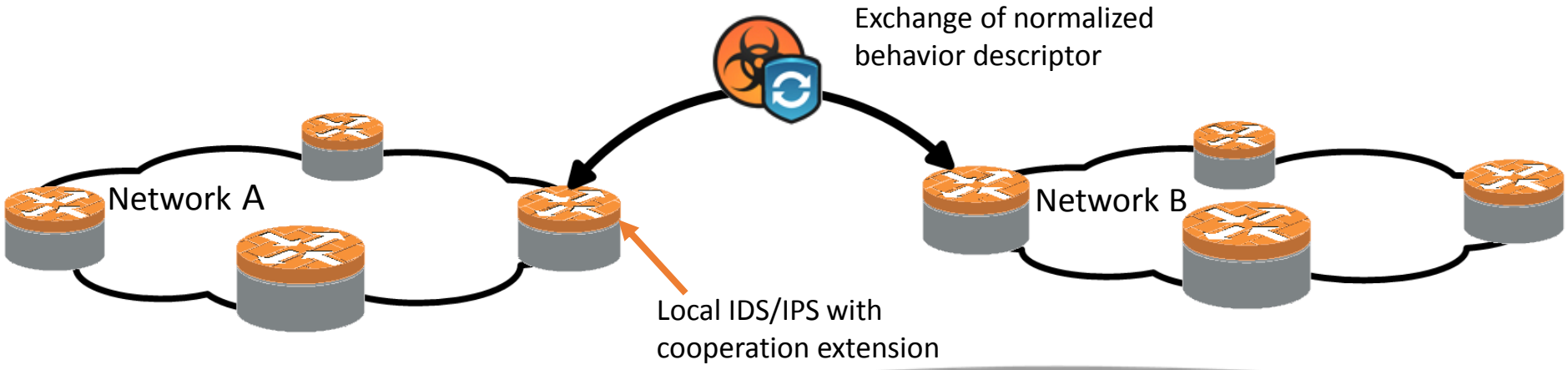


Conclusions

- Temporal patterns clearly visible
- More noise than expected
- Eliminating the noise could help to reduce false positives in detection
- Eliminating the noise would make behavior patterns exchangeable

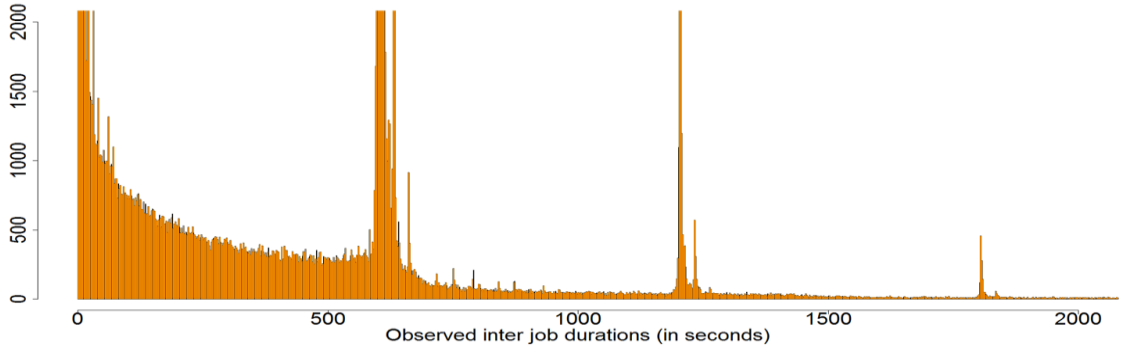
Outlook

- Exchange of normalized behavior descriptor
- Allows:
  - Efficient search for similar behavior in big data sets
  - If binarized, efficient algorithms could be used (K-nearest neighbor, Bloomfilter…)



Outlook

- Exchange of normalized (temporal) behavior descriptor
- Allows:
  - Efficient search for similar behavior in big data sets
  - If binarized, efficient algorithms could be used (K-nearest neighbor, Bloomfilter…)



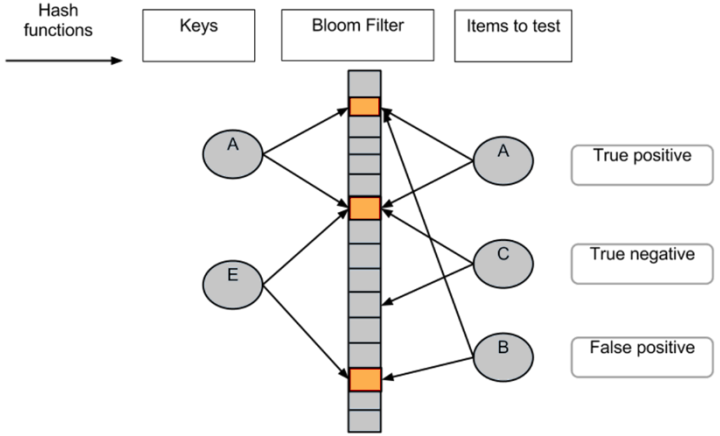
Noisy binarized (temporal) behavior signature

Normalized binarized (temporal) behavior signature

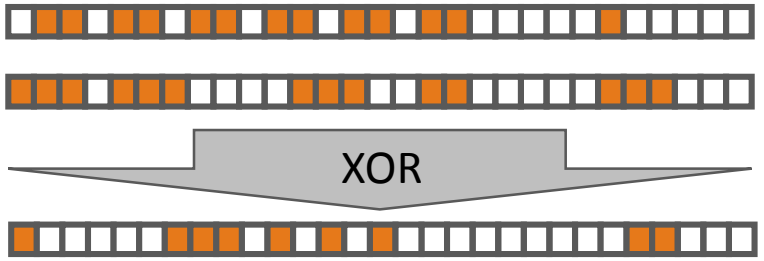
Outlook

Behavior signature database

Filter:



Matching (e.g. distance-based):



Hamming distance = 10