Dynamic profiles for malware communication

Joao Marques, Mick Cox

MSc System & Network Engineering
University of Amsterdam

Monday 6 February, 2017

Outline

Introduction

Some context

Hosting organization

Organization:

- Company: Deloitte Amsterdam
- Department: Cyber Risk Services
- Unit: Red team

Supervisor:

- Cedric van Bockhaven (OS3 alumnus)

Notable other:

- Joey Dreijer (OS3 alumnus)

Research Question

The goal

Is it possible to construct a dynamic network profile between a Command & Control server and the beacon, which is undetectable by state-of-the-art detection frameworks?

Intrusion Detection

A brief taxonomy

Intrusion Detection

Definition

Intrusion Detection & Prevention Systems in short:

- Collect data from the network or host
- Validated by a detection engine
- Reports if it suspects an intrusion
- Acts (isolates, shuts down) if it supports prevention

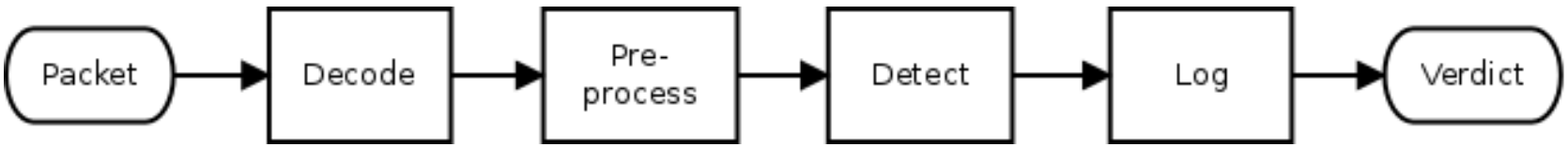


Figure 1: Simplified Snort 2 Architecture

Intrusion Detection

Network, DNS and Host-based

Network Based IDS (NIDS)

- Data collection from network (packets)
- Sensors in the network to validation engine
- Few sensors can capture all traffic
- Open source systems include Snort, Suricata and Bro

Host Based IDS (HIDS)

- Data collection from host systems (system metrics, usage)
- Agent on the host to validation engine
- Every agent needs agent to cover the network
- Open source systems include OSSEC, Tripwire

Others proposed types include DNS based, Storage based, Wireless, Hybrid, and more.

Intrusion Detection

Methods

Signature based IDS

- Based on predefined rules (malicious usage)
- Mostly pattern matching
- Generally unable to detect 0-days
- High true positive and false negative

Anomaly based IDS

- Based on training set (normal behavior)
- Mostly machine learning
- Detects deviations from normal behavior (anomalies)
- High false positives and true negative

Signature or anomaly based detection exists across the location (Host/Network)

Intrusion Detection

Validation engine

Rule Header

- Rule Actions (Alert, log, pass, activate, dynamic . . .)
- Protocols (TCP, UDP, ICMP, . . .)
- IP address / Port and direction
- Activate and dynamic rules

Rule Options

- General (msg, classification, . . .)
- Payload (content, length, depth, distance , . . .)
- Non Payload (fragoffset, ttl, flags, . . .)
- Post-Detection (logto, react, replace, . . .)

Dynamic modules and preprocessors

Intrusion Detection

Example rule

An example for matching content:

```
alert tcp any any -> any 139 (content:"|5c
00|P|00|I|00|P|00|E|00 5c|";)
```

Botnets & Advanced Persistent Threats

A brief taxonomy

Botnets & Advanced Persistent Threats

Botnets in short:

- A botnet is a network of infected computers, called bots
- Bots communicate with a Command & Control server, mostly over:
- Communication is common over HTTP(S), IRC or P2P systems
- Communication system on the bot is called a beacon

Advanced Persistent Threats in short:

- Targeted attack by a determined attacker
- Government or organizational funding
- Often utilizing botnets

Botnets

Architecture

Different architectures between C&C's to bots exist:

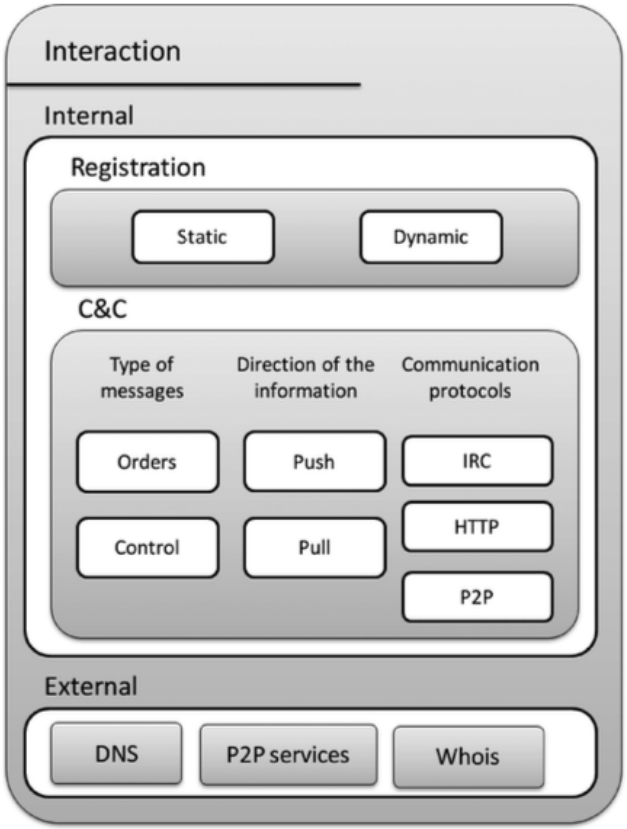
- Centralized: fast convergence, single point of failure
- Decentralized: resilient but slow convergence
- Hybrid: best of both

Botnets

The whole process

In summarization: most attacks do follow the following process.

1. Malware is distributed, often over multiple channels
2. Host gets infected by exploiting a vulnerability and downloading the malware as a result
   - Downloads the main executable/script
   - Main script downloads necessary libraries
3. Reports to C&C
4. Communicates frequent keepalive to C&C
5. Execution of commands
6. Self replicates (optional)
7. Self destructs (optional)



Botnets

Detection techniques

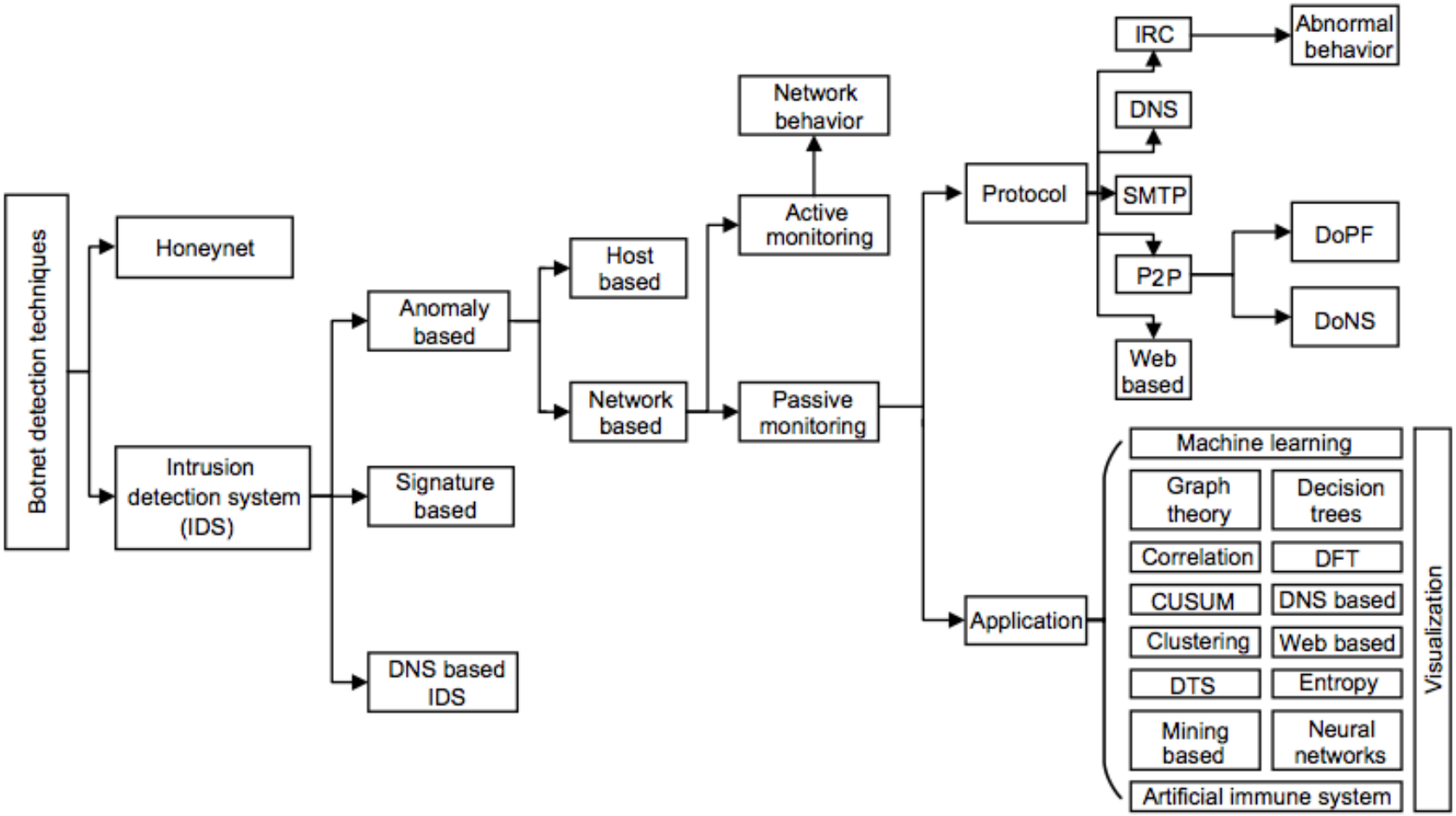


Figure 2: Botnet detection techniques

Botnets

Hiding mechanisms

Some of the reported hiding mechanisms include:[1]

- Multi-hopping (Usage of multiple proxies / gateways)
- Network traffic encryption
- Binary obfuscation
- Code polymorphism
- Fast flux networks (Quickly change DNS)
- E-mail spoofing (for spam)

[1] Survey and taxonomy of botnet res. thr. life-cycle, Rodríguez-Gómez et al. (2013)

Research Outline

Initial plan & the pivot

Initial plan

Start with exploiting signature based detection.

1. In dept research of signatures & signature based IDS
2. Find a weakness in the Snort 3 engine
3. Does it hold up against anomaly based techniques

Setup & Experiment

VMware EXSi server at reims.studlab.os3.nl contains a virtual test environment as seen in the figure bellow:

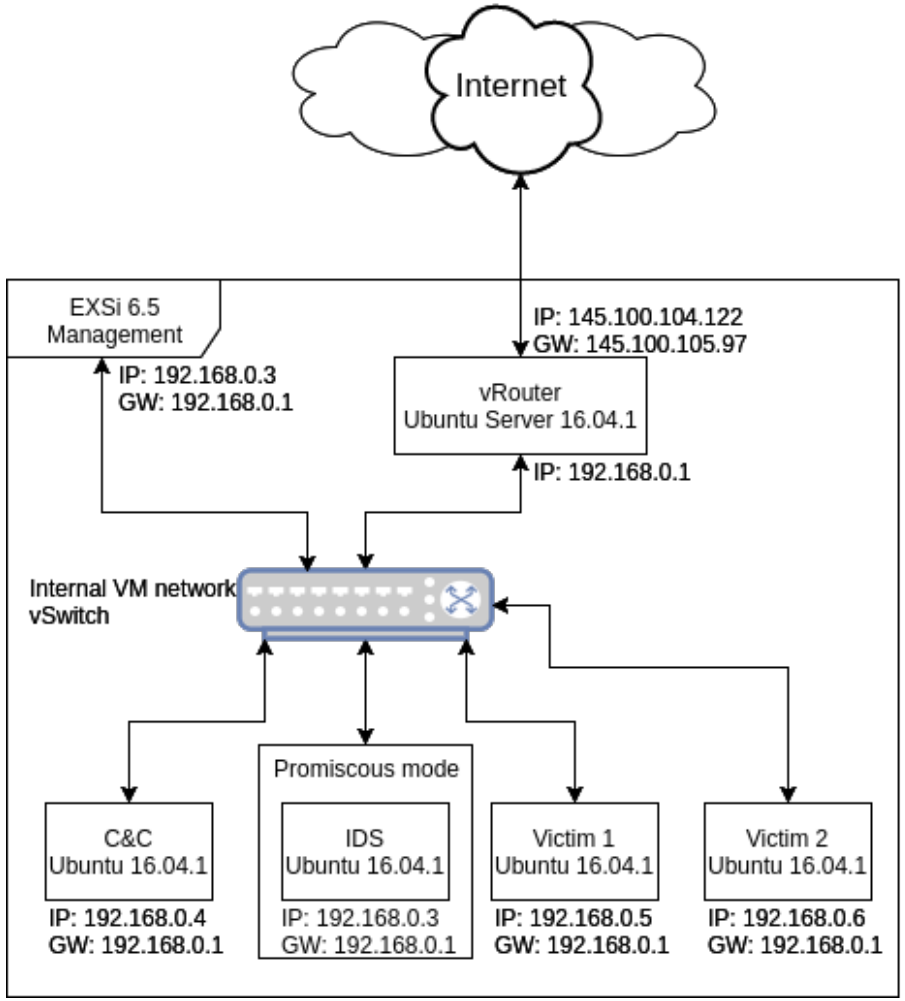


Figure 3: Test environment

Some considerations

Leading to the pivot

- Signature are by definition deterministic
  - No existing signatures for new malwares, evasion by default
  - Due to modular design, shortcomings can be patched
- Anomalies are by definition not normal
  - Normal behavior is defined by a representative data training set.
    - Training set context dependent & difficult to collect
  - If normal exists $\Longrightarrow$ not normal exists, for every area.
    - Mostly theoretical frameworks described in literature
    - Mostly machine learning ...

Intelligent Malware

A proposed framework

Intelligent Malware

The concept

Malware that can make an educated guess prior to starting communication with the C&C, to avoid using anomalous methods of communication that could end up in the detection of the infection.

Intelligent Malware

The objective

The objective of this degree of "intelligence" is to:

- Hide in plain sight
- Frustrate signature making
- Frustrate anomaly detection

System Overview

Assumptions

Basic assumptions:

- Network is monitored by an IDS or an IPS
- Network traffic is being filtered
- There is no HIDS in infected systems
- At least one of the types of communication is being used in the infected host.

System Overview

Malware Operation Method

A vulnerability in the victim is exploited and the payload executed. The malware gets downloaded and executed. From this point on the malware:

- Sniffs all DNS and SSH traffic of the victim for a limited amount of time
- Checks if any of the SSH connections initiated in the host
- Checks for specific domain lookups in the DNS traffic

System Overview

Malware Operation Method

- Once the sniffing operations is done (timeout) it does an "intelligent" analysis of the acquired information
- Downloads the module it requires to run that type of communication
- Starts communicating with the C&C that is listening on all types of communication
- if no suitable method of communication is found it deletes itself in a secure manner to prevent/hinder the creation of signatures

System status

Functionality Implemented

- Host Identification - Implemented
- Network Sniffing - Implemented
- Decision Making - Partially Implemented
- Modular Communication - Not Implemented
- Self Deletion - Not Implemented

Possible Countermeasures

Against a smart and dynamic malware

Security

Usability trade-off

Enforcing heavy restrictions on users

- Anomaly detection on the initial download
- Restricting even very known and popular services. like Dropbox
- The usage of network services and applications outside of the normal patterns
- Less freedom implies stricter patterns, anomalies will be easier to

Host Based Intrusion Detection

Defined yet again

In essence, what are HIDS?

- Agent on the client
- Central logging server to which it reports

Host Based Intrusion Detection

Anomaly based System Profiling

Checks for metrics and performance indicators

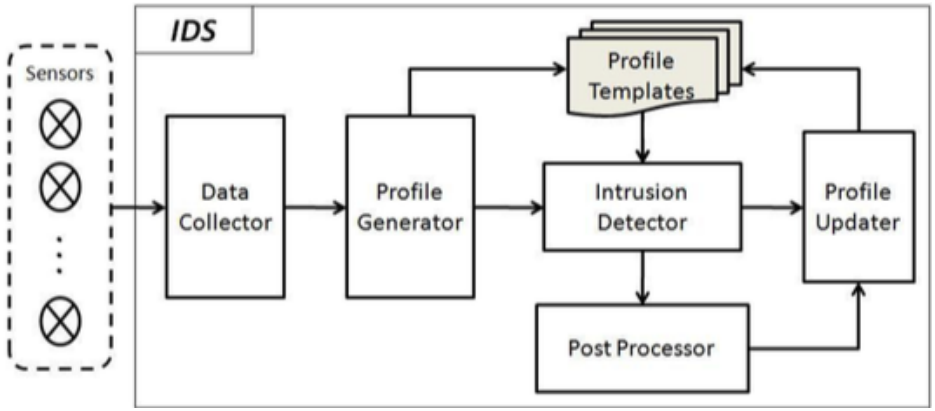
- Workload
- Traffic
- Logs
- Can be configured to send to the server anything...

Host Based Intrusion Detection

Anomaly based User Profiling

Anomaly based user profiling can be done on the basis of:[1]

- Psychometrics (intelligence, decisions)
- Behavioral biometrics
- Examples: System and network usage, keystroke analysis, commands, lexical and syntactic features.



Frequent or continuous (re)training of the training set is required, risking an attacker can over time manipulate the profile away from the genuine user.

[1] User Profiling in Intrusion Detection, Peng et al. (2016)

Host Based Intrusion Detection

Would it work?

The good:

- A vast amount of information that can be used to detect out of ordinary operation
- Enables a well managed network to separate any anomalous device from the network

The bad:

- Creates a lot of logs!
- Possibly a high amount of false positives
- impossible to scale

The ugly:

- An immense Administrative burden

Host Based Intrusion Detection

Log analysis

Actually already other systems, but for convenience listed here.
Difficult to disable logging. Some options do exist:

- Security information and event management (SIEM)
- System iNtrusion Analysis and Reporting Environment (SNARE)

Correlation

Combining the previous

Correlates multiple information sources: HIDS, NIST, signature and anomaly, both in log or metrics. Cross reference them to reduce false positives.

Correlation

Would it work?

The good:

- Cross referencing anomaly based detection, removing false positives
- Working towards an omniscient system

The bad:

- Creates a lot of logs!
- Probably a lot of false positives
- Development can be very complex

The ugly:

- Big brother becomes a bit bigger

Discussion, Conclusion & Future Work

Dealing with heuristics

Discussion & Conclusion

Bridging the gap

Intrusion Detection: looking for a needle in the haystack, involves heuristics

Furthermore, evasion against signature based systems is by default and anomaly is not yet that effective due to large rate of false positives.

In order to uncover some of the advanced communication methods such as advanced covert channels and side channel attacks, misusing current applications and protocols to hide in plain sight, developing such tools is needed.

With our proposed system we hope to contribute to the advancement of such research.

Future work

For better understanding

In regards to the proposed framework

- Advanced reconnaissance features can be build into the framework to make its decision more reliable and therefore evasive
- More advanced modular beacons are needed for wider usage
- In reaction, the defensive side (blue team) can then make an effort in actual

(Maybe OS3 students can do their RP2 on advanced and stealthy covert channels and side channel attacks.)