01-07
Clustering: exemple d’application (BotMiner)

OPTIONNEL

NOUS ÉCLAIRONS.
VOUS BRILLEZ.

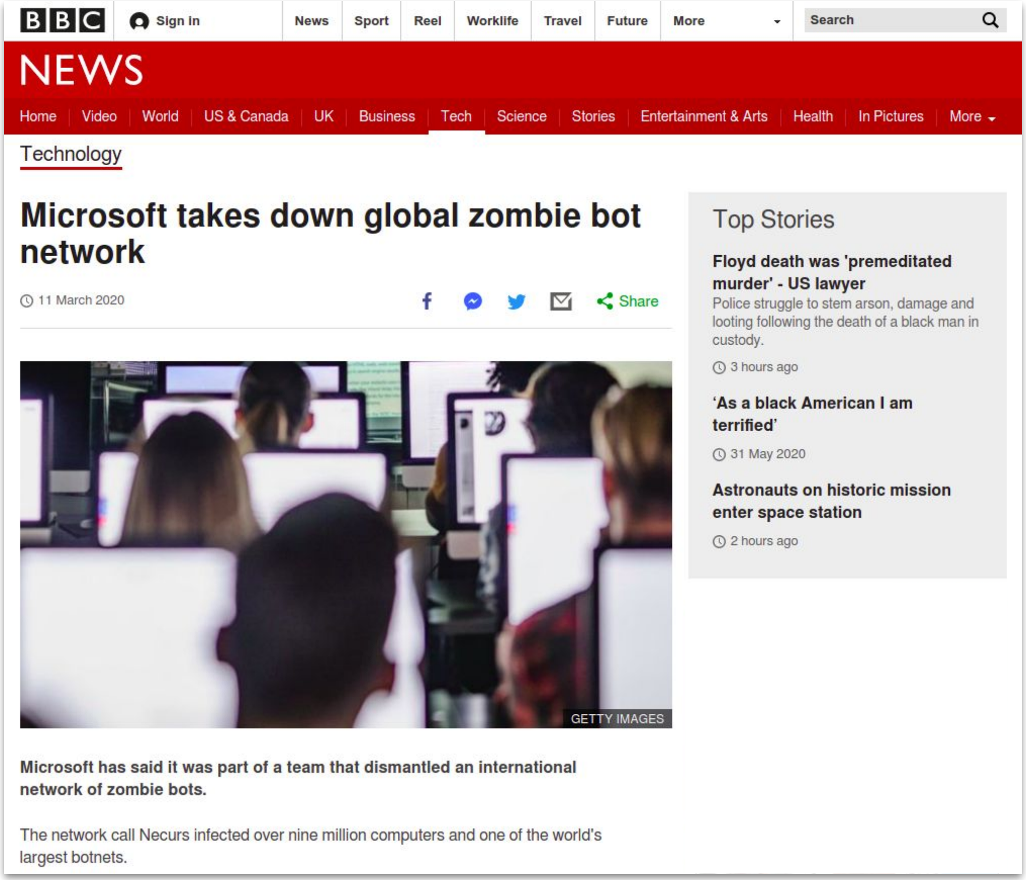
FORMATION CONTINUE
ET SERVICES AUX ENTREPRISES

Cégep de Sainte-Foy

Spécialisation technique en intelligence artificielle
Algorithmes d’apprentissage non supervisé — 420-A58-SF — M. Swawola, M.Sc.

1

Un peu d’actualités …



Technology

Microsoft takes down global zombie bot network

11 March 2020

Microsoft has said it was part of a team that dismantled an international network of zombie bots.

The network call Necurs infected over nine million computers and one of the world's largest botnets.

Top Stories

Floyd death was 'premeditated murder' - US lawyer

Police struggle to stem arson, damage and looting following the death of a black man in custody.

3 hours ago

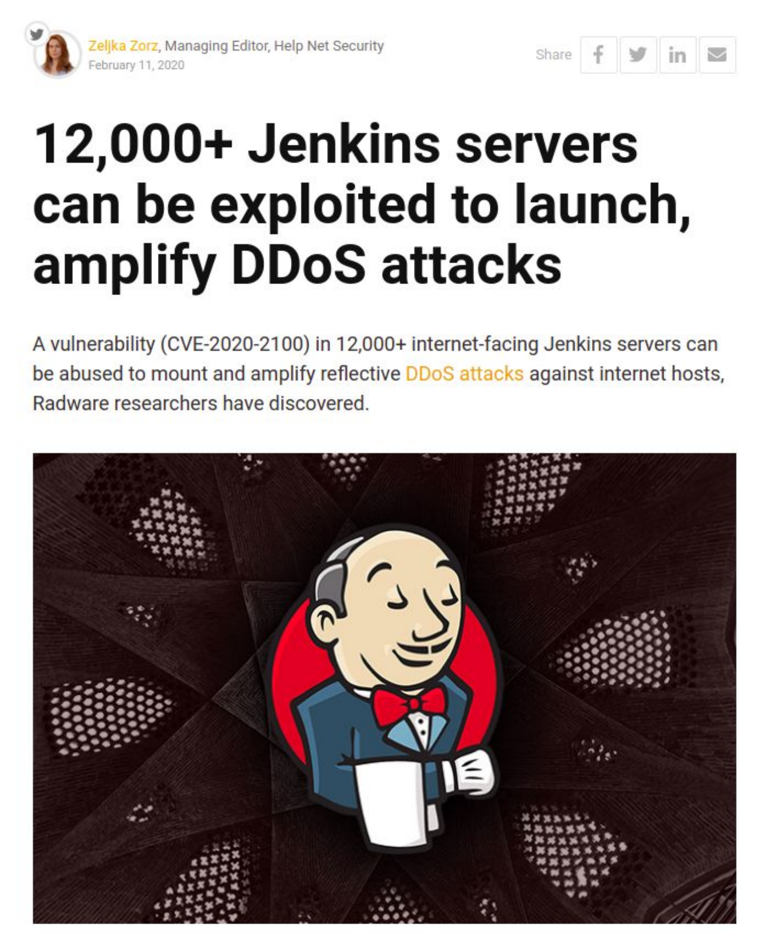
'As a black American I am terrified'

31 May 2020

Astronauts on historic mission enter space station

2 hours ago

https://www.bbc.com/news/technology-51828781

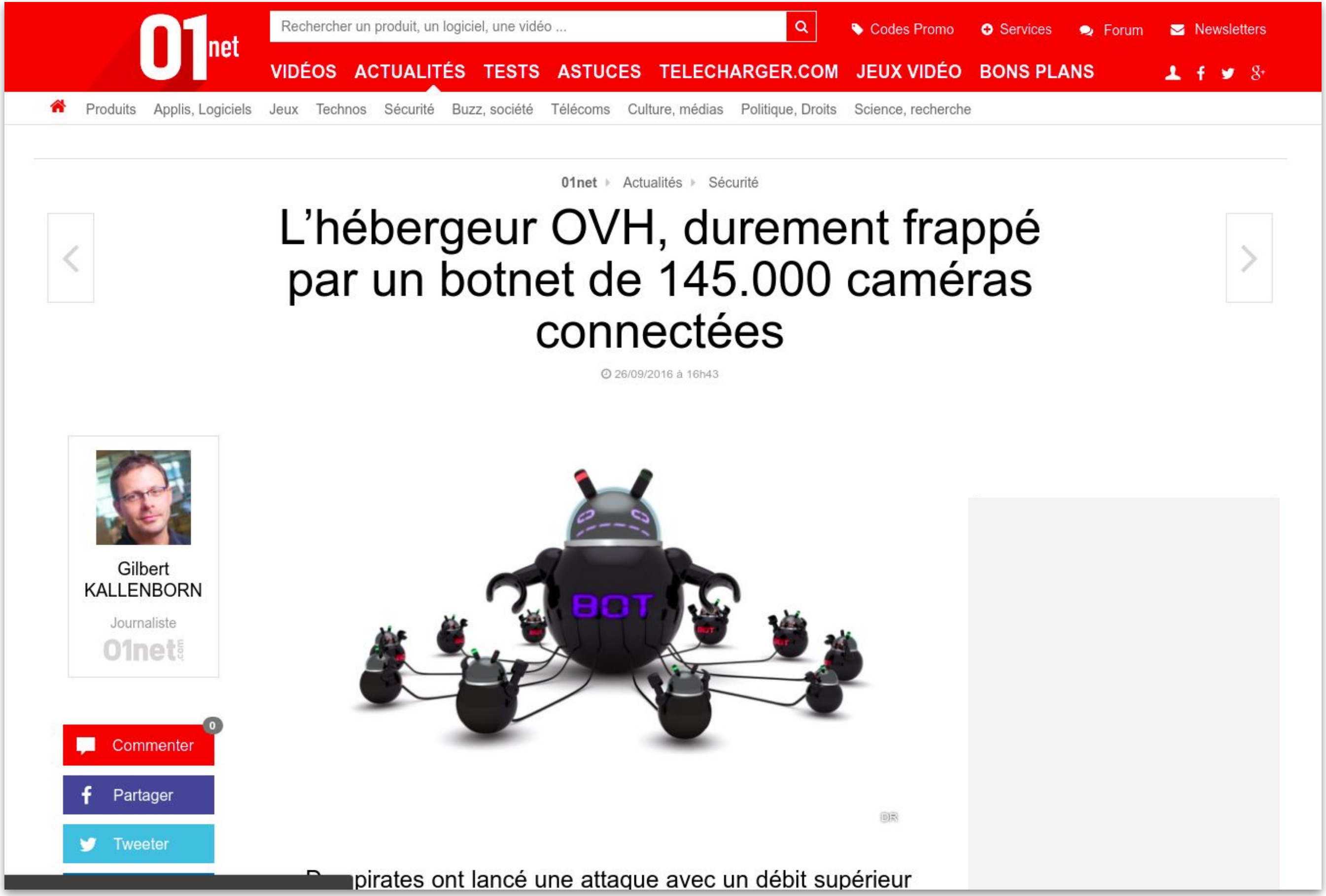


Zeljka Zorz, Managing Editor, Help Net Security
February 11, 2020

12,000+ Jenkins servers can be exploited to launch, amplify DDoS attacks

A vulnerability (CVE-2020-2100) in 12,000+ internet-facing Jenkins servers can be abused to mount and amplify reflective DDoS attacks against internet hosts, Radware researchers have discovered.

https://www.helpnetsecurity.com/2020/02/11/cve-2020-2100/



01net › Actualités › Sécurité

L'hébergeur OVH, durement frappé par un botnet de 145.000 caméras connectées

26/09/2016 à 16h43

Gilbert KALLENBORN

Journaliste

01net.com

D pirates ont lancé une attaque avec un débit supérieur

1 Tbit/s

https://www.01net.com/actualites/l-hebergeur-ovh-durement-frappe-par-un-botnet-de-145-000-cameras-connectees-1041185.html

L’Intelligence Artificielle peut elle aider à contrer ces attaques ?

Rechercher un produit, un logiciel, une vidéo ...

Codes Promo · Services · Forum · Newsletters

VIDÉOS · ACTUALITÉS · TESTS · ASTUCES · TELECHARGER.COM · JEUX VIDÉO · BONS PLANS

Produits · Applis, Logiciels · Jeux · Technos · Sécurité · Buzz, société · Télécoms · Culture, médias · Politique, Droits · Science, recherche

01net › Actualités › Sécurité

L’hébergeur OVH, durement frappé par un botnet de 145.000 caméras connectées

26/09/2016 à 16h43

Gilbert KALLENBORN

Journaliste 01net

Commenter

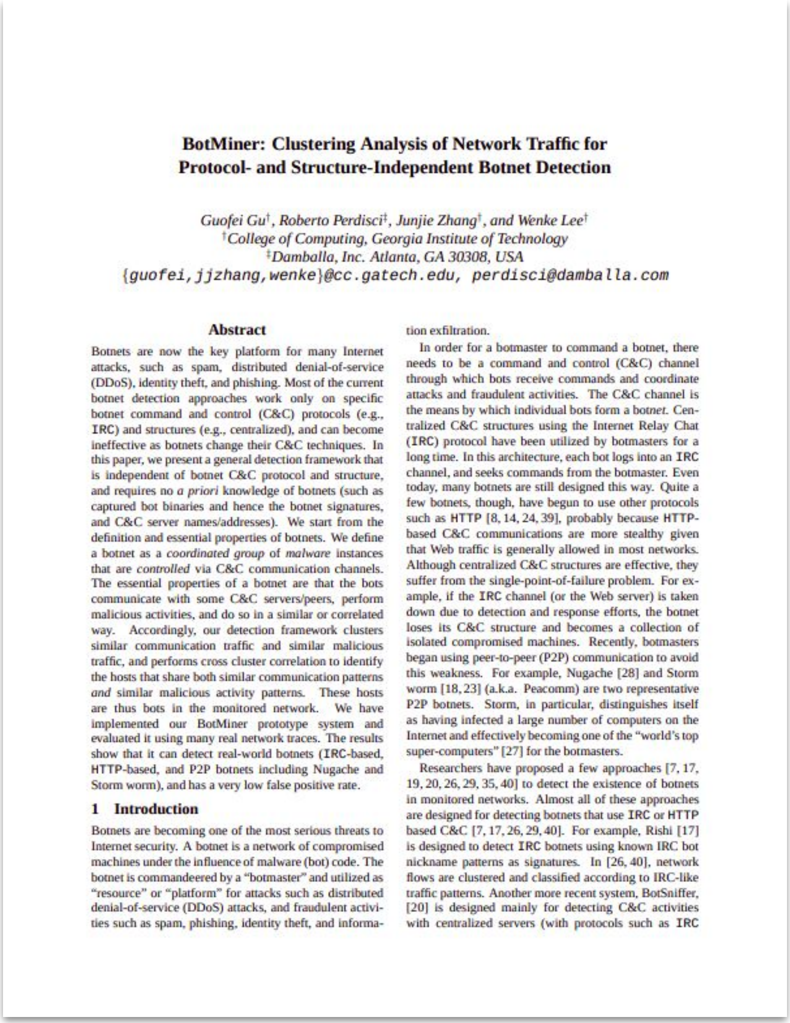
Partager

Tweeter

1 Tbit/s

pirates … débit supérieur

https://www.01net.com/actualites/l-hebergeur-ovh-durement-frappe-par-un-botnet-de-145-000-cameras-connectees-1041185.html



BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection

Guofei Gu[†]*, Roberto Perdisci*[‡]*, Junjie Zhang*[†]*, and Wenke Lee*[†]
[†]*College of Computing, Georgia Institute of Technology*
[‡]*Damballa, Inc. Atlanta, GA 30308, USA*
{*guofei,jjzhang,wenke*}*@cc.gatech.edu, perdisci@damballa.com*

Abstract

Botnets are now the key platform for many Internet attacks, such as spam, distributed denial-of-service (DDoS), identity theft, and phishing. Most of the current botnet detection approaches work only on specific botnet command and control (C&C) protocols (e.g., IRC) and structures (e.g., centralized), and can become ineffective as botnets change their C&C techniques. In this paper, we present a general detection framework that is independent of botnet C&C protocol and structure, and requires no *a priori* knowledge of botnets (such as captured bot binaries and hence the botnet signatures, and C&C server names/addresses). We start from the definition and essential properties of botnets. We define a botnet as a *coordinated group* of *malware* instances that are *controlled* via C&C communication channels. The essential properties of a botnet are that the bots communicate with some C&C servers/peers, perform malicious activities, and do so in a similar or correlated way. Accordingly, our detection framework clusters similar communication traffic and similar malicious traffic, and performs cross cluster correlation to identify the hosts that share both similar communication patterns *and* similar malicious activity patterns. These hosts are thus bots in the monitored network. We have implemented our BotMiner prototype system and evaluated it using many real network traces. The results show that it can detect real-world botnets (IRC-based, HTTP-based, and P2P botnets including Nugache and Storm worm), and has a very low false positive rate.

1 Introduction

Botnets are becoming one of the most serious threats to Internet security. A botnet is a network of compromised machines under the influence of malware (bot) code. The botnet is commandeered by a "botmaster" and utilized as "resource" or "platform" for attacks such as distributed denial-of-service (DDoS) attacks, and fraudulent activities such as spam, phishing, identity theft, and information exfiltration.

In order for a botmaster to command a botnet, there needs to be a command and control (C&C) channel through which bots receive commands and coordinate attacks and fraudulent activities. The C&C channel is the means by which individual bots form a botnet. Centralized C&C structures using the Internet Relay Chat (IRC) protocol have been utilized by botmasters for a long time. In this architecture, each bot logs into an IRC channel, and seeks commands from the botmaster. Even today, many botnets are still designed this way. Quite a few botnets, though, have begun to use other protocols such as HTTP [8, 14, 24, 39], probably because HTTP-based C&C communications are more stealthy given that Web traffic is generally allowed in most networks. Although centralized C&C structures are effective, they suffer from the single-point-of-failure problem. For example, if the IRC channel (or the Web server) is taken down due to detection and response efforts, the botnet loses its C&C structure and becomes a collection of isolated compromised machines. Recently, botmasters began using peer-to-peer (P2P) communication to avoid this weakness. For example, Nugache [28] and Storm worm [18, 23] (a.k.a. Peacomm) are two representative P2P botnets. Storm, in particular, distinguishes itself as having infected a large number of computers on the Internet and effectively becoming one of the "world's top super-computers" [27] for the botmasters.

Researchers have proposed a few approaches [7, 17, 19, 20, 26, 29, 35, 40] to detect the existence of botnets in monitored networks. Almost all of these approaches are designed for detecting botnets that use IRC or HTTP based C&C [7, 17, 26, 29, 40]. For example, Rishi [17] is designed to detect IRC botnets using known IRC bot nickname patterns as signatures. In [26, 40], network flows are clustered and classified according to IRC-like traffic patterns. Another more recent system, BotSniffer, [20] is designed mainly for detecting C&C activities with centralized servers (with protocols such as IRC

BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection

Avertissement: il ne s’agit pas d’un cours de cybersécurité !

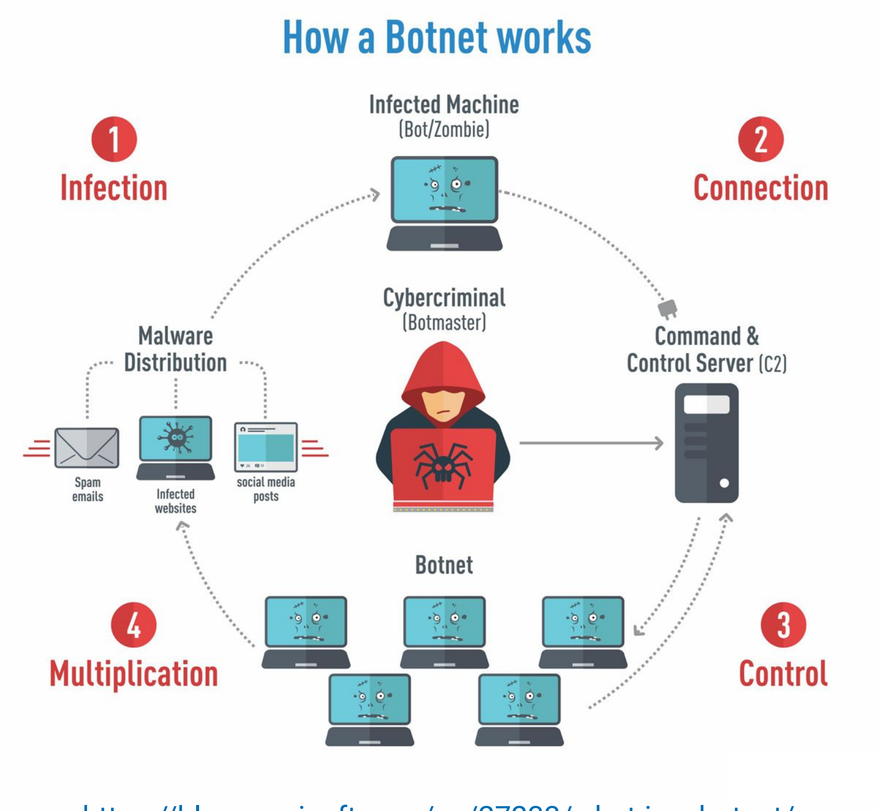
2

Qu’est-ce qu’un Botnet ?

Botnet

- **Botnet** = groupe d’ordinateurs ou appareils compromis (présence d’un malware) et contrôlés à distance par des hackers
- Permet d’effectuer des activités malveillantes comme
  - **Attaques DDoS**
  - Spam
  - Phishing
  - Vol d’identité
- Basé sur les protocoles: IRC, HTTP, ...
- Centralisés et P2P

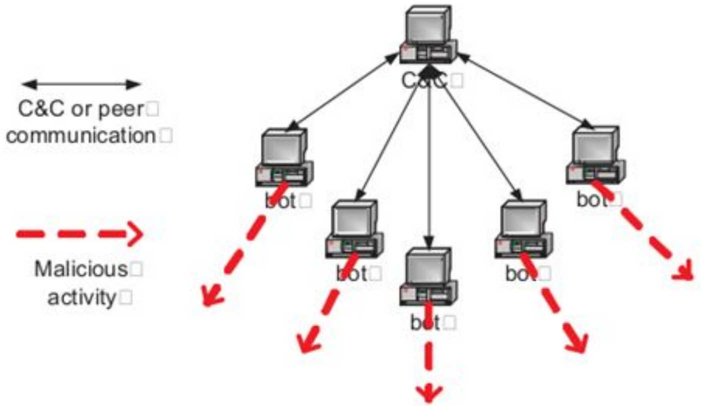
Botnet



https://blog.emsisoft.com/en/27233/what-is-a-botnet/

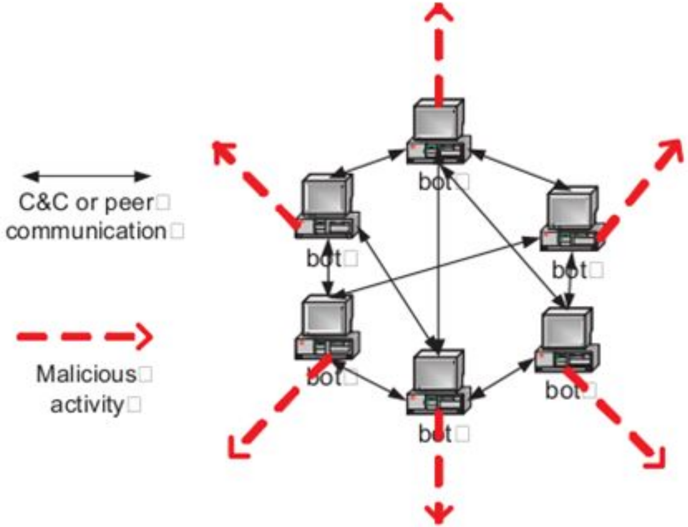
Botnets centralisés

- **Un Botmaster** envoi des commandes à un serveur C2 désigné
- Les Bots (ou zombies) requièrent des commandes du serveur C2



Botnets P2P

- **Pas de serveur C2**
- **Un Botmaster** envoi des commandes à tous les bots
- Les bots partagent les commandes reçues avec les bots voisins



3 Outils de détection

Outils de détection

Rishi

- **Détecte les botnets basés sur IRC**
- **Surveille le traffic**
  - Pseudos / nicknames suspects
  - Serveurs suspects
  - Ports non communs

BotSniffer

- Détection d’anomalies basé sur les réseaux
  - Tous les bots à l’intérieur d’un botnet partagent des patterns de traffic similaires
  - Détecte les botnets **IRC** et **HTTP**
  - Ne détecte pas les botnets P2P

Outils de détection

BotHunter

- **IRC et HTTP**
- **Permet de détecter les botnets P2P**
- **Se base sur un modèle de cycle d’infection**
  - Ne fonctionne plus si le cycle d'infection change !

4

BotMiner

Objectif de BotMiner

- Detect groups of compromised machines that are part of a botnet
- **Independent** of C&C communication structure and content
- Minimal false positives
- Resource efficient detection
- Based on **unsupervised learning** techniques !

Architecture

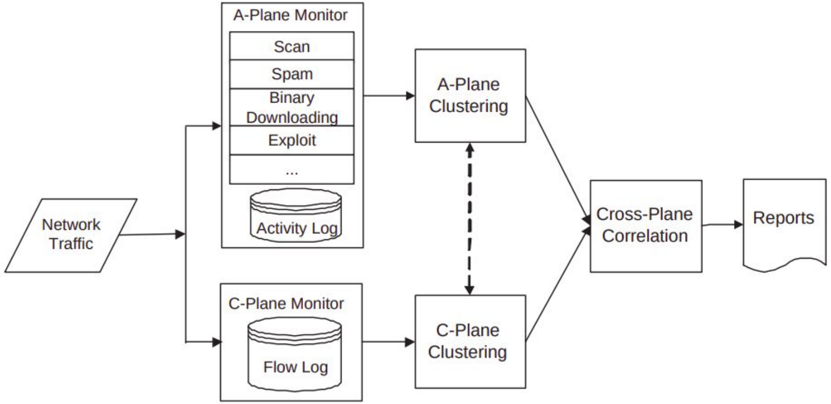
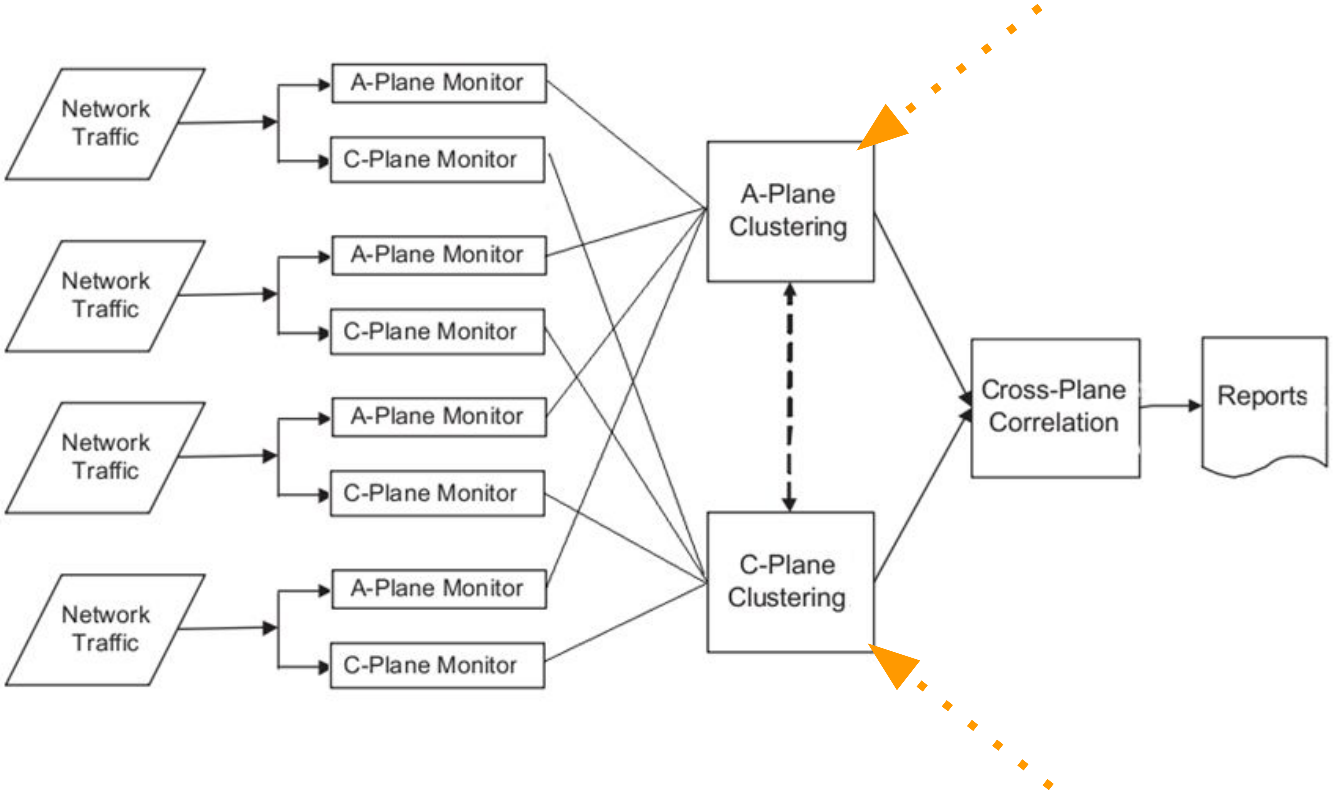


Figure 2: Architecture overview of our BotMiner detection framework.

Architecture



C-plane and A-plane Monitor

C-plane (communications)

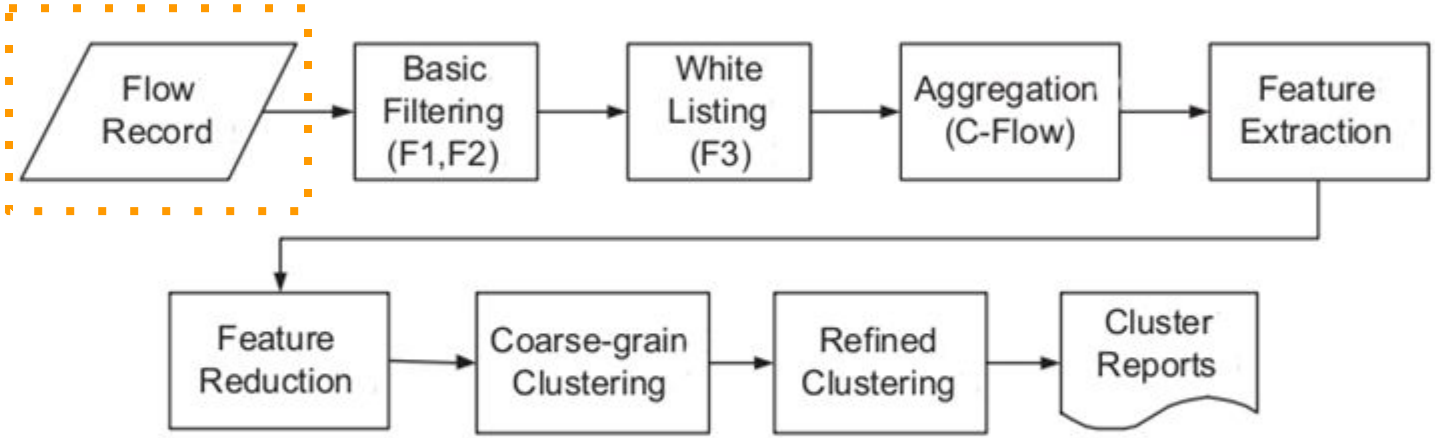
- Who is talking to whom?
  - TCP and UDP traffic flows
  - Time, duration
  - Source, destination
  - Packet count, bytes transferred
  - Manageable log size less than 1GB per day for 300 Mbps network

A-plane (activities)

- Who is doing what?
  - Detects malicious activities
  - Scanning / binary downloading
  - Spamming / exploit attempts
  - Snort with custom plugins

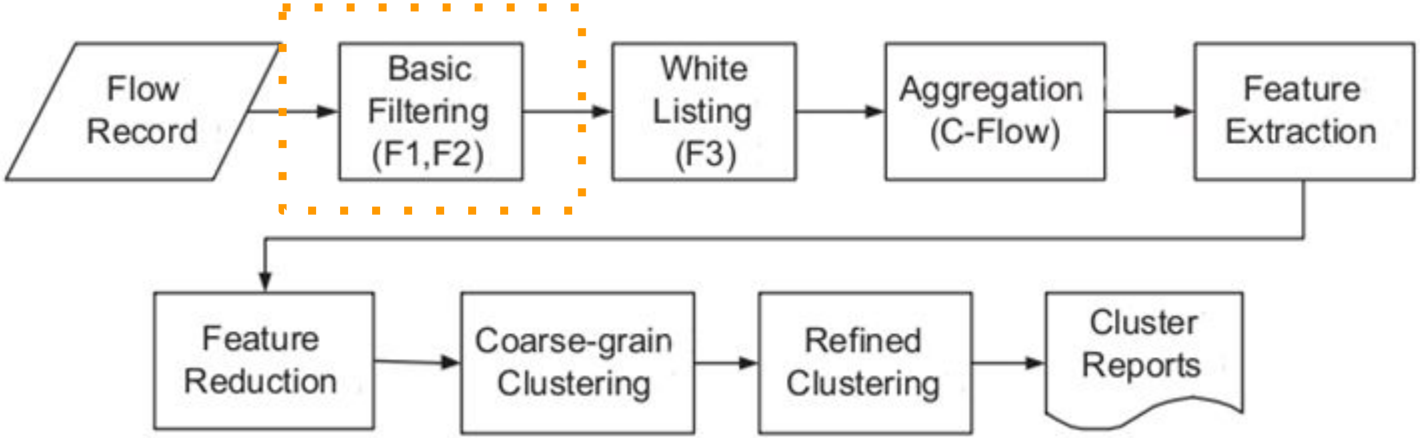
5 C-plane clustering

C-plane clustering



Which machines have similar communication patterns?
C-plane monitor logs → cluster reports

C-plane clustering

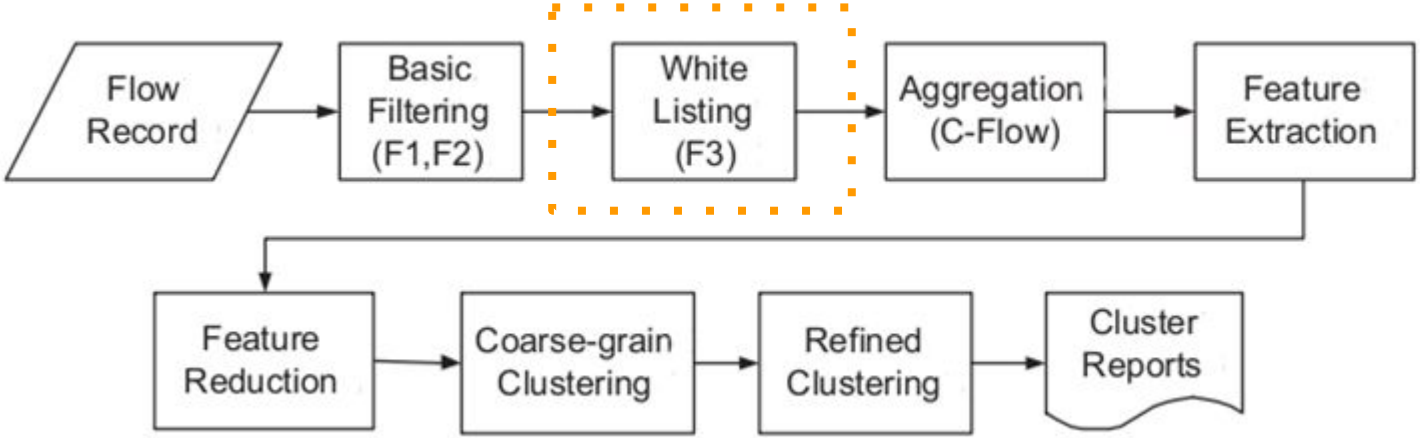


Basic Filtering

Remove internal flows

Remove one way flows

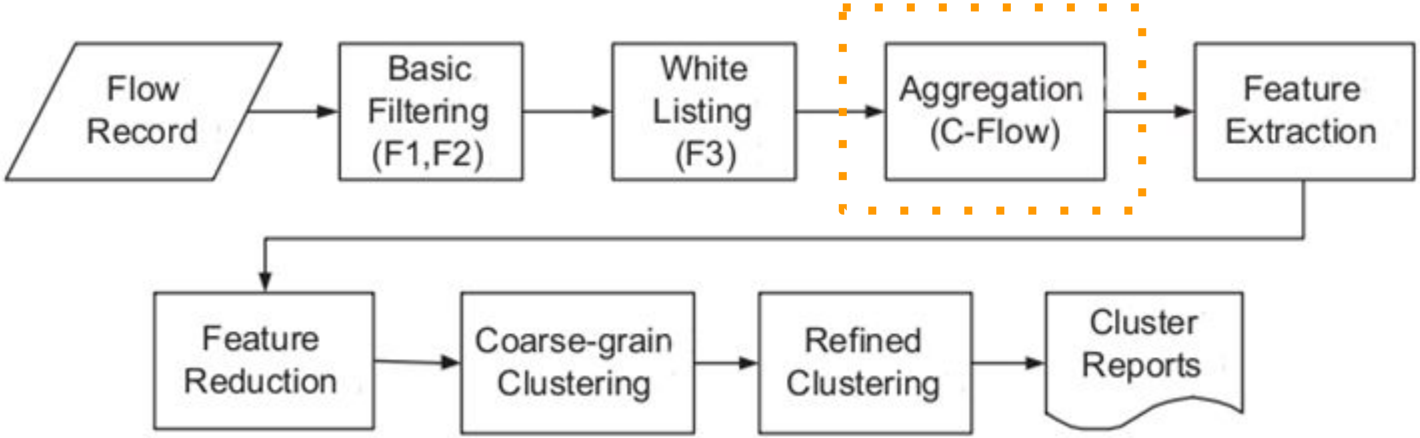
C-plane clustering



White Listing

Remove flows to popular destination (Google, Yahoo, etc.)

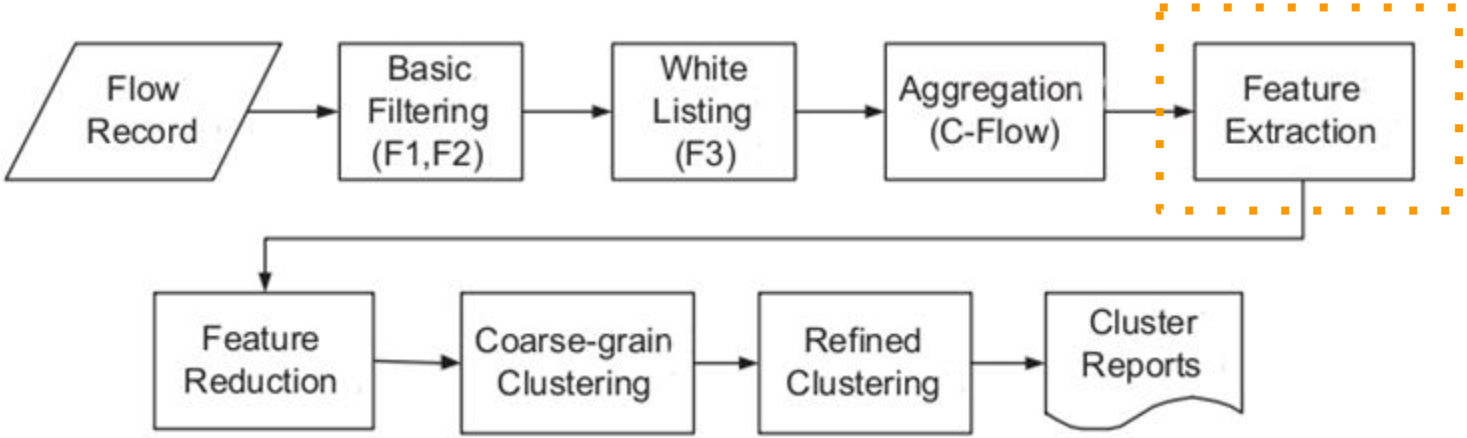
C-plane clustering



Aggregation

C-flow: all traffic flows over a period of time that share the same source, destination, and protocol

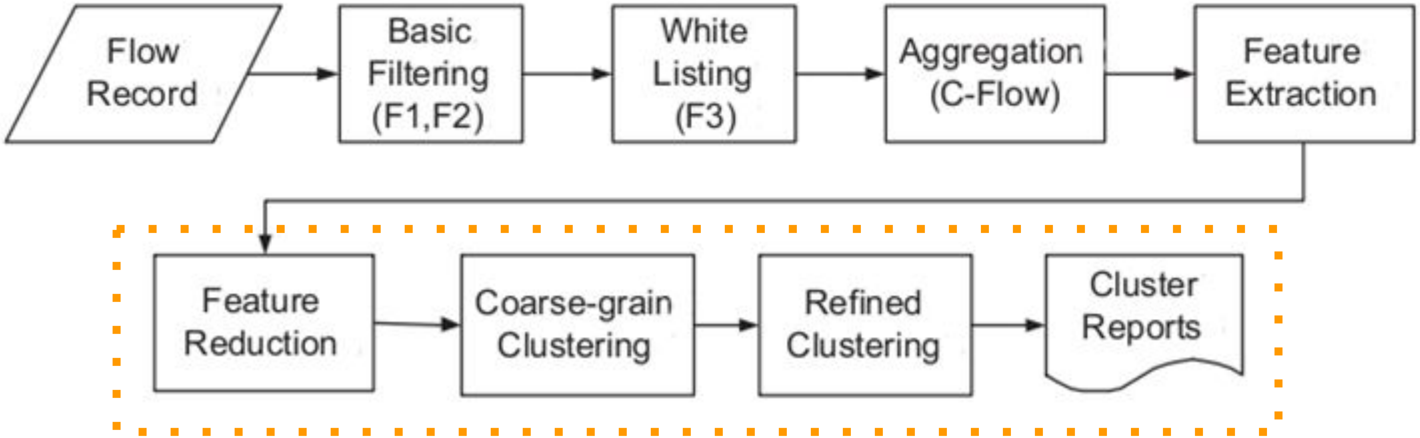
C-plane clustering



Feature Extraction

flows per hour (bytes per packet)
packets per flow (bytes per second)

C-plane clustering



Two-step Clustering

Coarse-grain and Refined clustering

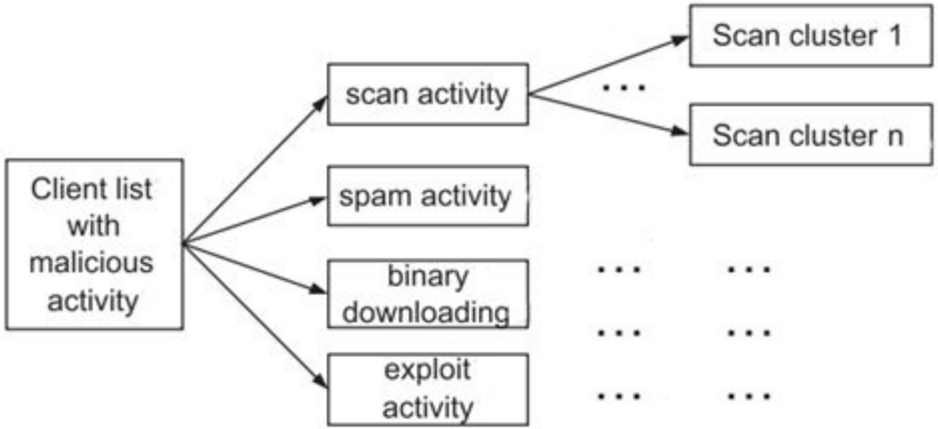
X-means clustering algorithm

6

A-plane clustering

A-plane clustering

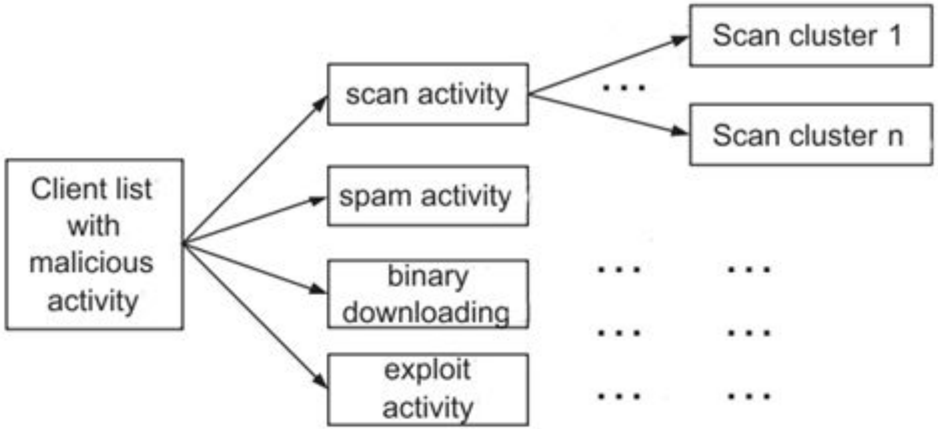
Which machines have similar activity patterns?
A-plane monitor logs
→ cluster reports



A-plane clustering

Activity Type Clustering

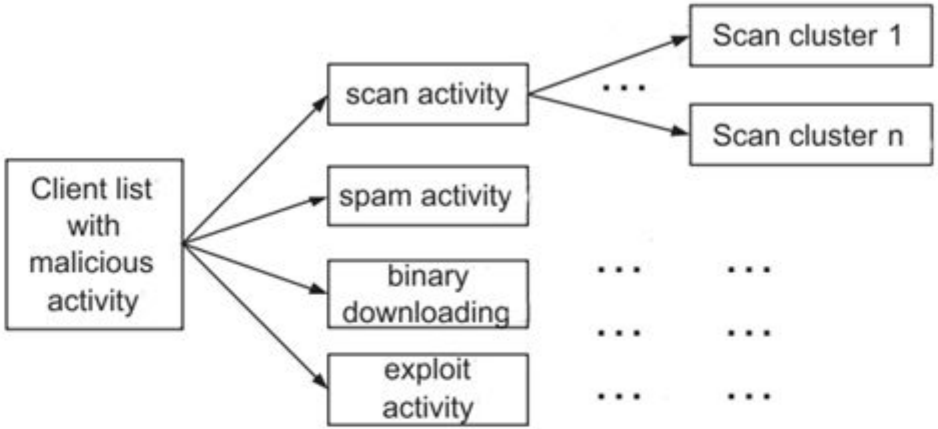
- Scan
- Spam
- Binary download
- Exploit



A-plane clustering

Activity Feature Clustering

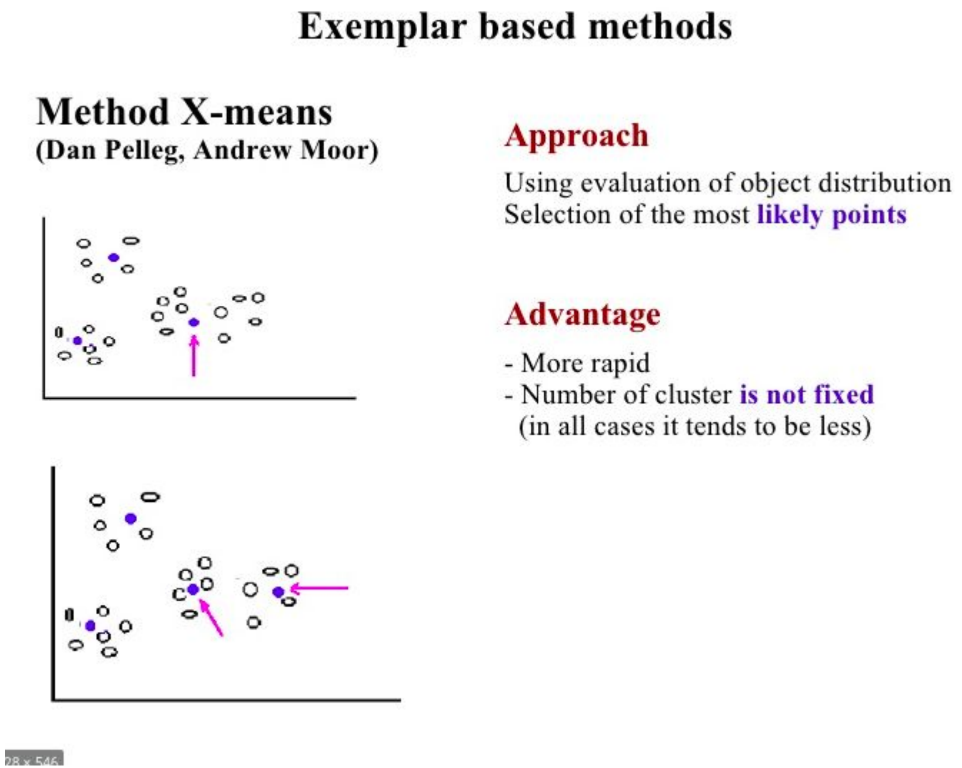
- Target subnet
- Similar binary
- Spam content
- Exploit type



7

Algorithme X-means

X-means clustering (1/3)



Exemplar based methods

Method X-means
(Dan Pelleg, Andrew Moor)

Approach

Using evaluation of object distribution
Selection of the most **likely points**

Advantage

- More rapid
- Number of cluster **is not fixed**
  (in all cases it tends to be less)

28 x 546

X-means clustering (2/3)

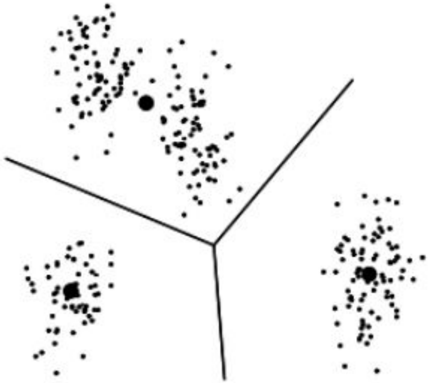


Figure 1. The result of running K-means with three centroids.

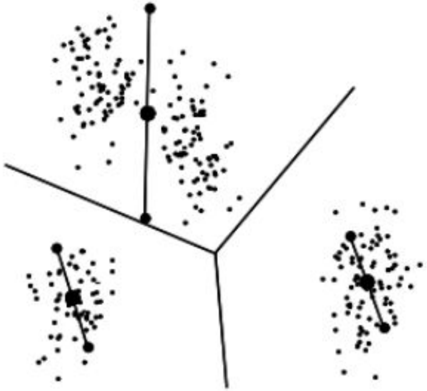


Figure 2. Each original centroid splits into two children.

X-means clustering (3/3)

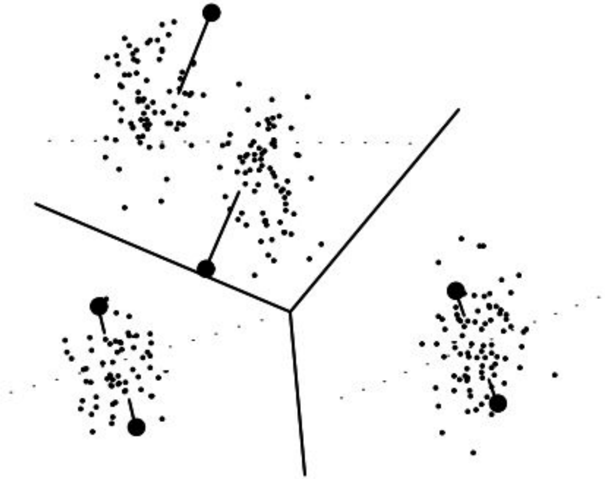


Figure 3: The first step of parallel local 2-means. The line coming out of each centroid shows where it moves to.

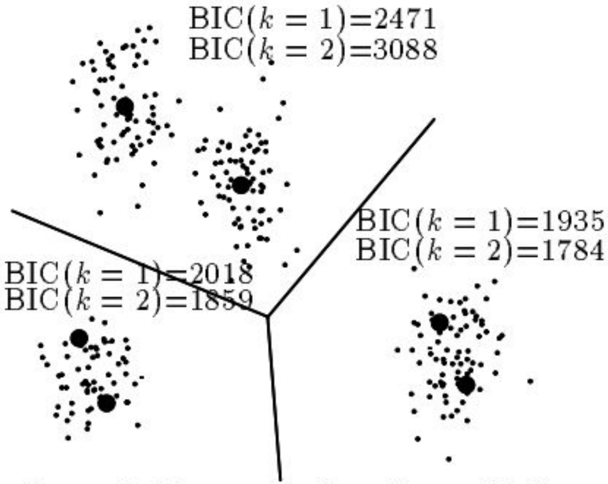


Figure 4: The result after all parallel 2-means have terminated.

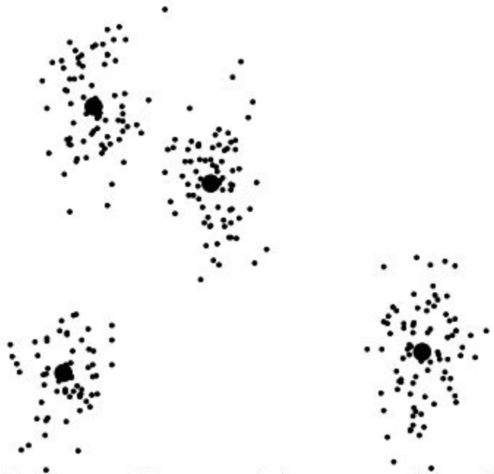


Figure 5: The surviving centroids after all the local model scoring tests.

8

Cross-plane correlation

Cross-plane correlation

Which machines are in a botnet?
Botnet score
Number of clusters
Score of other hosts in cluster
Activity weighting
Which bots are in the same botnet?

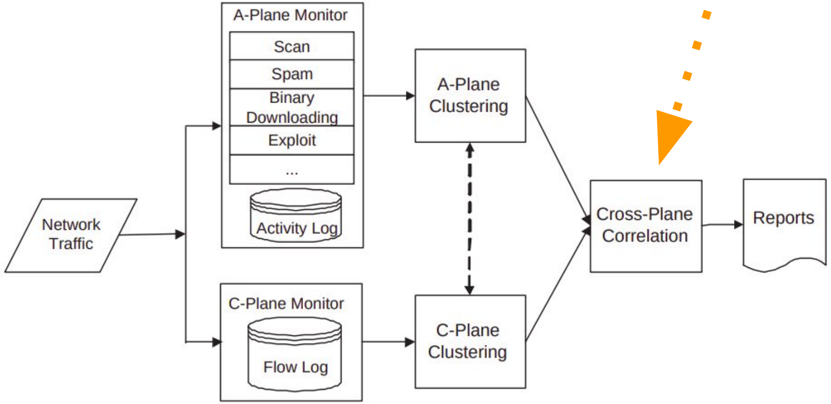
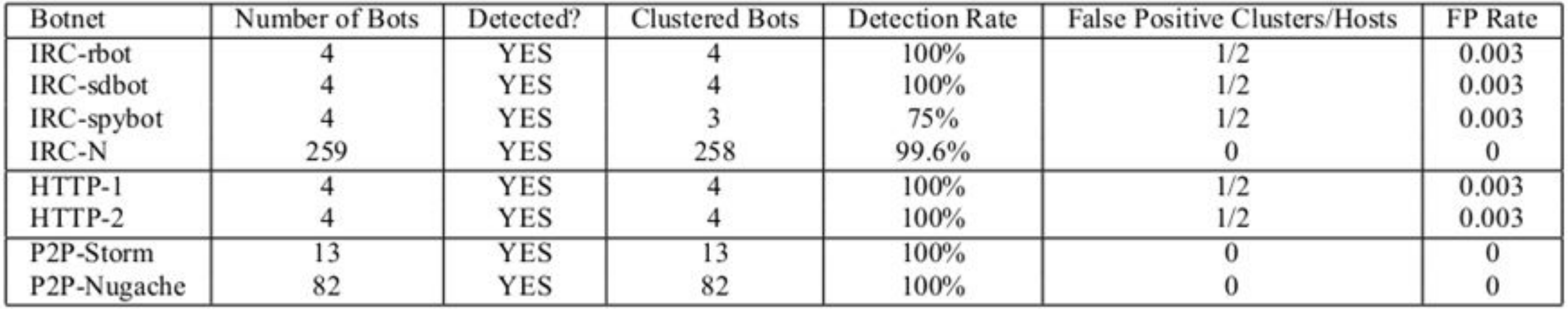


Figure 2: Architecture overview of our BotMiner detection framework.

9

Résultats

Résultats



Botnet	Number of Bots	Detected?	Clustered Bots	Detection Rate	False Positive Clusters/Hosts	FP Rate
IRC-rbot	4	YES	4	100%	1/2	0.003
IRC-sdbot	4	YES	4	100%	1/2	0.003
IRC-spybot	4	YES	3	75%	1/2	0.003
IRC-N	259	YES	258	99.6%	0	0
HTTP-1	4	YES	4	100%	1/2	0.003
HTTP-2	4	YES	4	100%	1/2	0.003
P2P-Storm	13	YES	13	100%	0	0
P2P-Nugache	82	YES	82	100%	0	0

All botnets detected
99.6% bot detection
0.3% false positive rate

Limitations

- randomization and mimicry
- C-plane cluster evasion
- Individual or group commands
- A-plane cluster evasion
- Delay bot tasks
- Cross-plane analysis evasion

10 Lectures et références

Références

[1] BotMiner: Clustering Analysis of Network Traffic for Protocol- and Structure-Independent Botnet Detection

[2] X-means: Extending K-means with Efficient Estimation of the Number of Clusters