Metodi e tecniche per la rilevazione e il contrasto di botnet in reti universitarie

Andrea Balboni

andreabalboni@gmail.com

6° Borsisti Day

24/03/2015

Roma – Consortium GARR

Borse di Studio Orio Carlini Consortium GARR

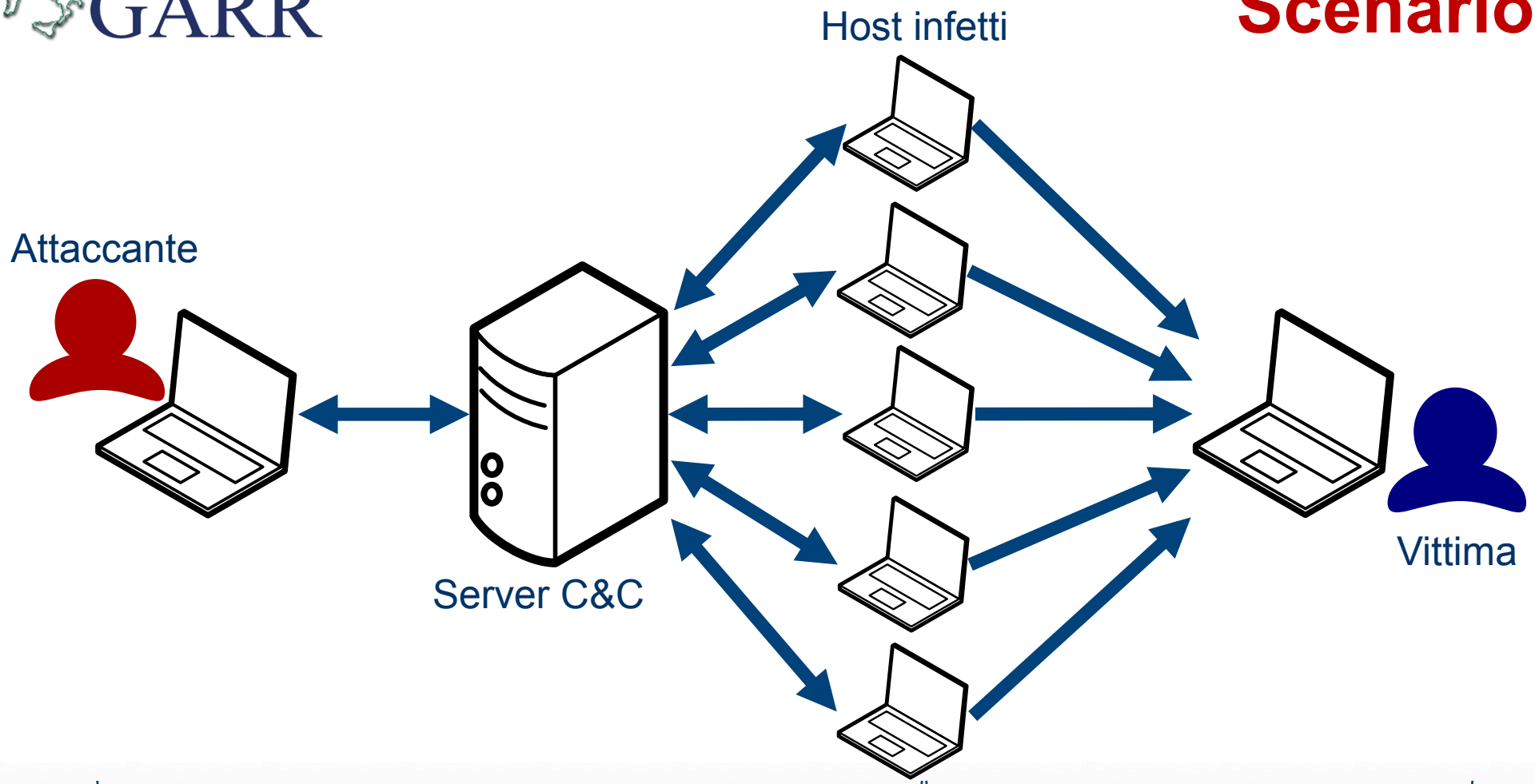
Obiettivi del progetto

Studio tecniche e metodi innovativi per l'individuazione ed il contrasto di botnet

- Acquisizione e integrazione dati da sorgenti multiple ed eterogenee
- Focalizzazione: correlazione e analisi dei **log DNS** per l'individuazione di **bot(net)**, e server di **Comando e Controllo** (C&C)

Borse di Studio Orio Carlini
Consortium GARR

Scenario



Borse di Studio Orio Carlini
Consortium GARR

Due fasi del progetto

1. Acquisizione dati

2. Analisi

Borse di Studio Orio Carlini Consortium GARR

1 - Acquisizione dati

Sorgente di informazioni derivanti da:

- **log DNS**
- allarmi forniti dal **CERT GARR**
- allarmi NIDS
- log HTTP
- netflow

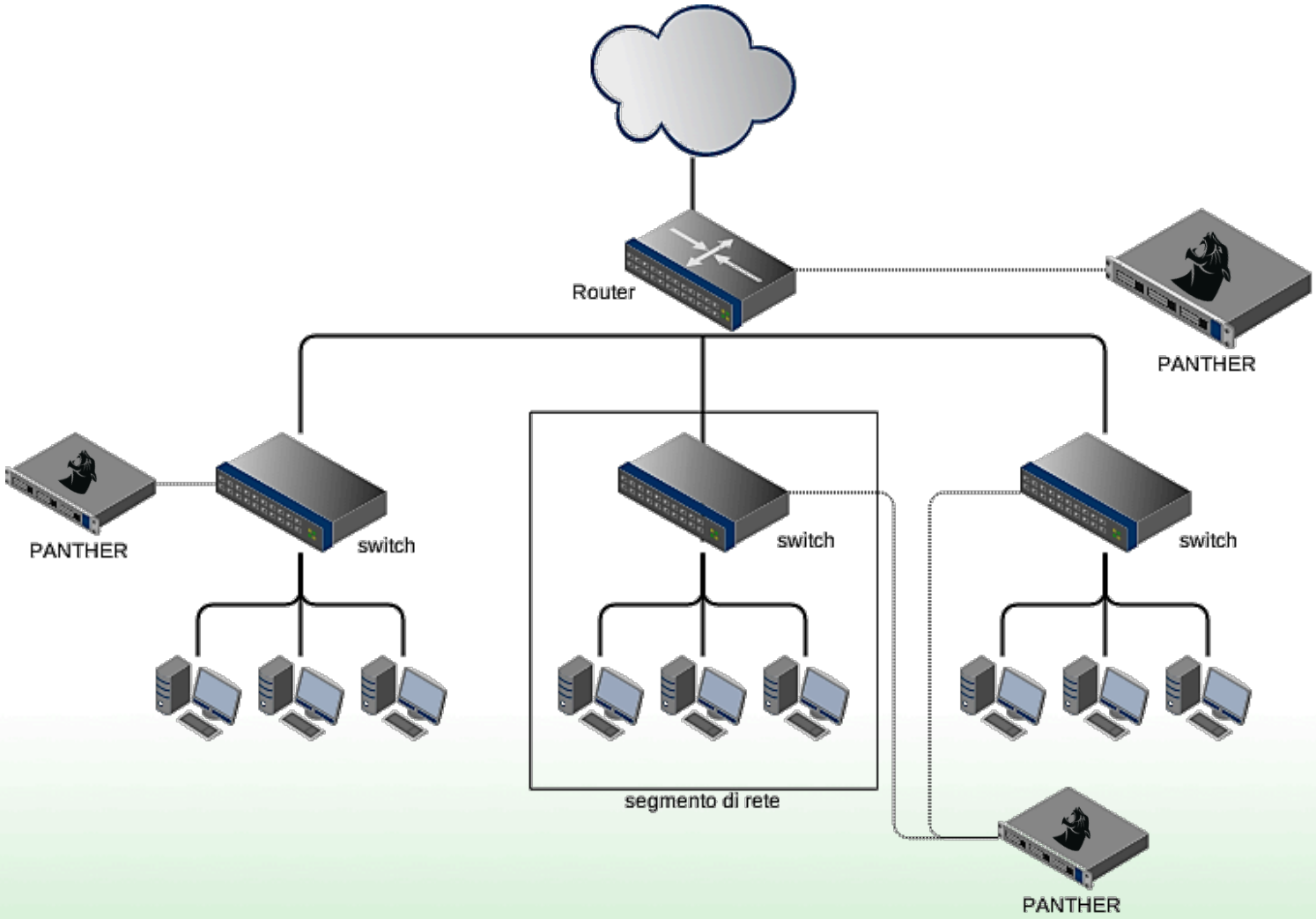
➔ *Privacy by design delle architetture*

Parallel Analyzer of Network Traffic for High-speed EnviRonments

PANTHER è un sistema per l'analisi in tempo reale di tutto il traffico in ingresso e in uscita anche su reti a larga banda.

- Ottimizzato per **realizzazione di architetture distribuite** per l'analisi del traffico e la rilevazione degli allarmi in **reti complesse e segmentate.**

PANTHER


Router
PANTHER
PANTHER
switch
switch
switch
segmento di rete
PANTHER

Borse di Studio Orio Carlini Consortium GARR

2 - Analisi: focalizzata su DNS

- DNS coinvolto in diversi ambiti nelle botnet
  - **Comunicazioni** tra bot e server di comando e controllo
  - Attività di perlustrazione del bot-master per verificare l'eventuale blacklisting DNS dei **nomi di dominio associati ai bot o server di controllo**
- Rilevazione botnet **indipendentemente dal protocollo** che utilizzano (HTTP, IRC, …)

Borse di Studio Orio Carlini Consortium GARR

Domain Generation Algorithms

DGA ⇒ i bot contattano i server C&C attraverso nomi di dominio **auto-generati**.

Esempi

- `razzrwsbzum.org`
- `jbfygzlcjak.info`

Obiettivo

Classificazione automatica dei nomi di dominio ⇒ {*sospetti, non sospetti*}

Analisi DNS

Anomaly Detection

- Frequenza di variazione dell'associazione domain name ⇔ indirizzo IP
  - identificazione di botnet fast-flux
- Cadenza temporale delle richieste di risoluzione di determinati nomi di dominio
  - identificazione di bot che contattano periodicamente il bot-master

Borse di Studio Orio Carlini Consortium GARR

Passive DNS

Possibile **integrazione** del motore di analisi con sistemi **Passive DNS**.

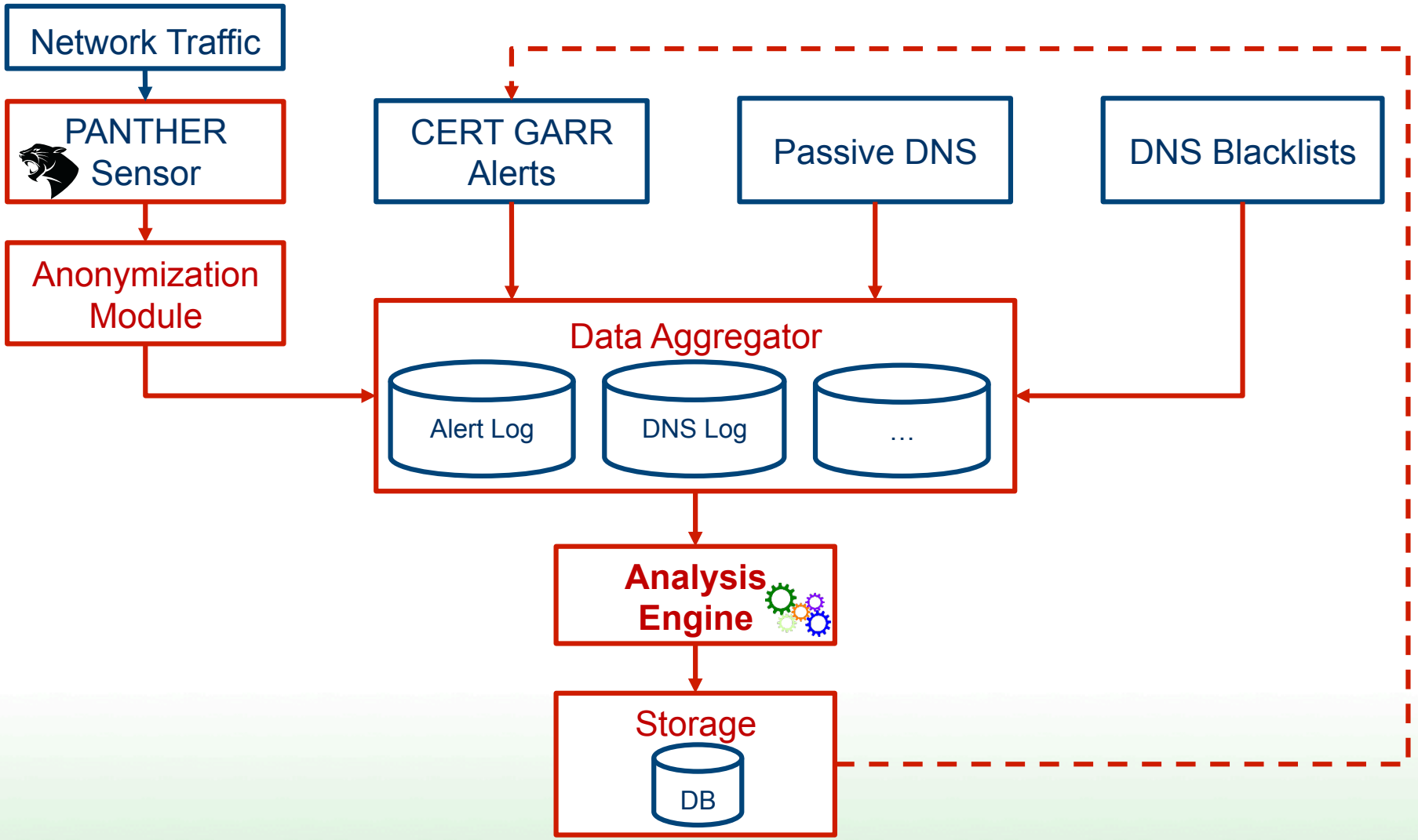
Informazioni ricavabili

- storico corrispondenze nome dominio ⇔ IP
- nomi di dominio gestiti da un NameServer
- nomi di dominio che puntano ad una sottorete IP
- sottodomini di uno specifico dominio

Attività svolta nel primo mese

Borse di Studio Orio Carlini
Consortium GARR

Progetto architettura



Borse di Studio Orio Carlini Consortium GARR

Log anonimizzati

```
Query TX 58db [**] 229.143.222.192.in-addr.arpa [**]
PTR [**] 172.21.140.92:35459 -> 72.52.71.2:53

Query TX d91d [**] ns2.electronicbox.com [**] A [**]
172.21.140.92:35459 -> 96.127.255.8:53

Response TX 58db [**] 143.222.192.in-addr.arpa [**]
NS [**] TTL 20864 [**] ns2.electronicbox.com [**]
72.52.71.2:53 -> 172.21.140.92:35459

Response TX 58db [**] 143.222.192.in-addr.arpa [**]
NS [**] TTL 20864 [**] ns1.electronicbox.net [**]
72.52.71.2:53 -> 172.21.140.92:35459
```

Estrapolazione transazioni DNS

```
Transaction NAME="ns2.electronicbox.com" TYPE=A
CLIENT=172.21.140.92 SERVER=96.127.255.8

Transaction NAME="229.143.222.192.in-addr.arpa"
TYPE=PTR CLIENT=172.21.140.92 SERVER=72.52.71.2
     <Response TYPE=NS NAME="143.222.192.in-
addr.arpa" TTL=20864 VALUE="ns2.electronicbox.com">
     <Response TYPE=NS NAME="143.222.192.in-
addr.arpa" TTL=20864 VALUE="ns1.electronicbox.net">
```

Euristiche per rilevare DGA / 1

Calcolo **score** per nomi di dominio di **primo** e **secondo livello**.

Metriche:

- lunghezza del nome
- elevato numero di consonanti
- elevato/ridotto numero di vocali

Euristiche per rilevare DGA / 2

- coefficiente di ***bhattacharyya*** utilizzando la distribuzione di probabilità delle lettere nella lingua inglese e quella nei nomi di dominio

- identificazione di parole da **dizionario inglese** ed **italiano** all'interno dei nomi di dominio di primo e secondo livello

Borse di Studio Orio Carlini Consortium GARR

Estrazione automatica DGA

- Realizzato strumento per l'estrazione dei domini con nome auto-generato dai log
- Caso di studio: log DNS relativo 10 ore di traffico
  - 1.299.594 transazioni
  - 239.940 nomi di dominio
  - **88 DGA** rilevati di cui 5 falsi positivi

Piano di lavoro a breve termine

Maggio 2015

- Realizzazione di un **classificatore** per identificare DGA dato un nome di dominio
  - **Training set**: DGA ricavati da euristiche + nomi di dominio sicuramente validi (es. Alexa top100)

Settembre 2015

- Integrazione con un sistema **Passive DNS**
  - Arricchimento delle informazioni contenute nei log