EURECOM
Sophia Antipolis

The quest for multi-headed worms

Van-Hau PHAM, Marc DACIER,
Guillaume URVOY-KELLER,
Taoufik EN-NAJJARY

Outline

- **Introduction**
- Method and Implementation
  - Experimental Environment
  - Approach
  - Results
- Conclusion

Definition of **multi-headed worms**

- ***Combining several known exploits***
- ***Only one exploit used to attack against a new target***
- ***Less efficient to propagate but more stealthy***

  - ***Example: Welchia***

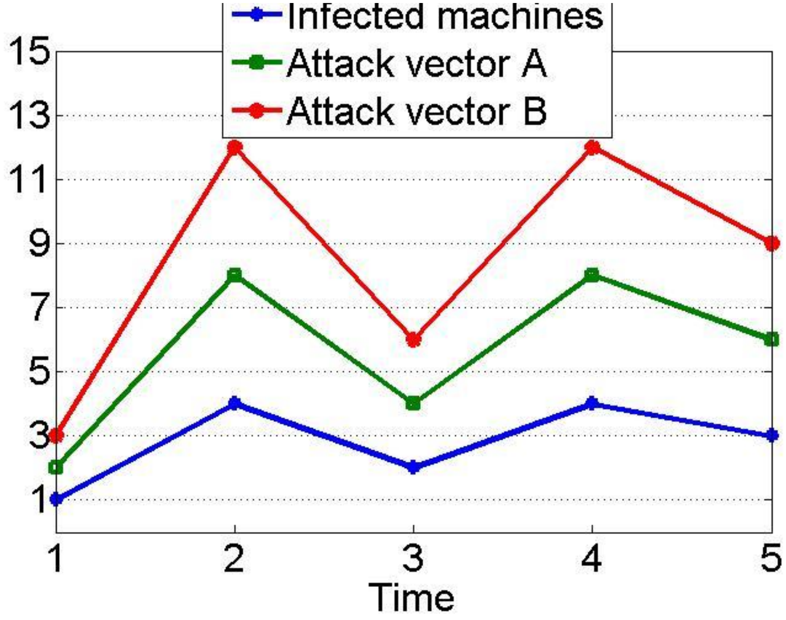
  (Pouget, Fabien;Urvoy-Keller, Guillaume;Dacier, Marc "Time signatures to detect multi-headed stealthy attack tools" 18th Annual FIRST Conference, June 25-30, 2006, Baltimore, USA)

HOW TO DETECT **multi-headed worms ?**

- **Multi-headed worms leave correlated attack traces**

  Example: a multi-headed worm carries two attack vectors A and B

  - At each time-step, an infected machine makes 5 attacks
  - In 2/5 times, using attack vector A, in 3/5 times, using attack vector B



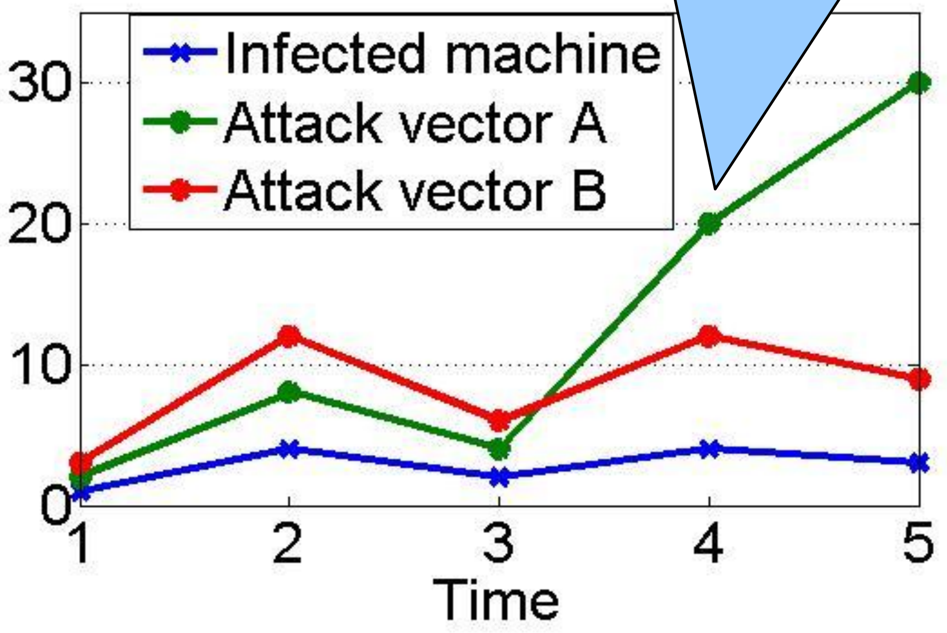
# of attacks of A and B always vary together, and they are a function of # of infected machines

- **Correlation of attack traces is a sign of multi-headed worms**

Shortcomings when applying to a large dataset

- **Too many attack traces**
- **Sliding windows vs. whole period: to deal with**
  - the overlapping between different activities
  - the incompleteness of observation

Another tool reuses attack vector A



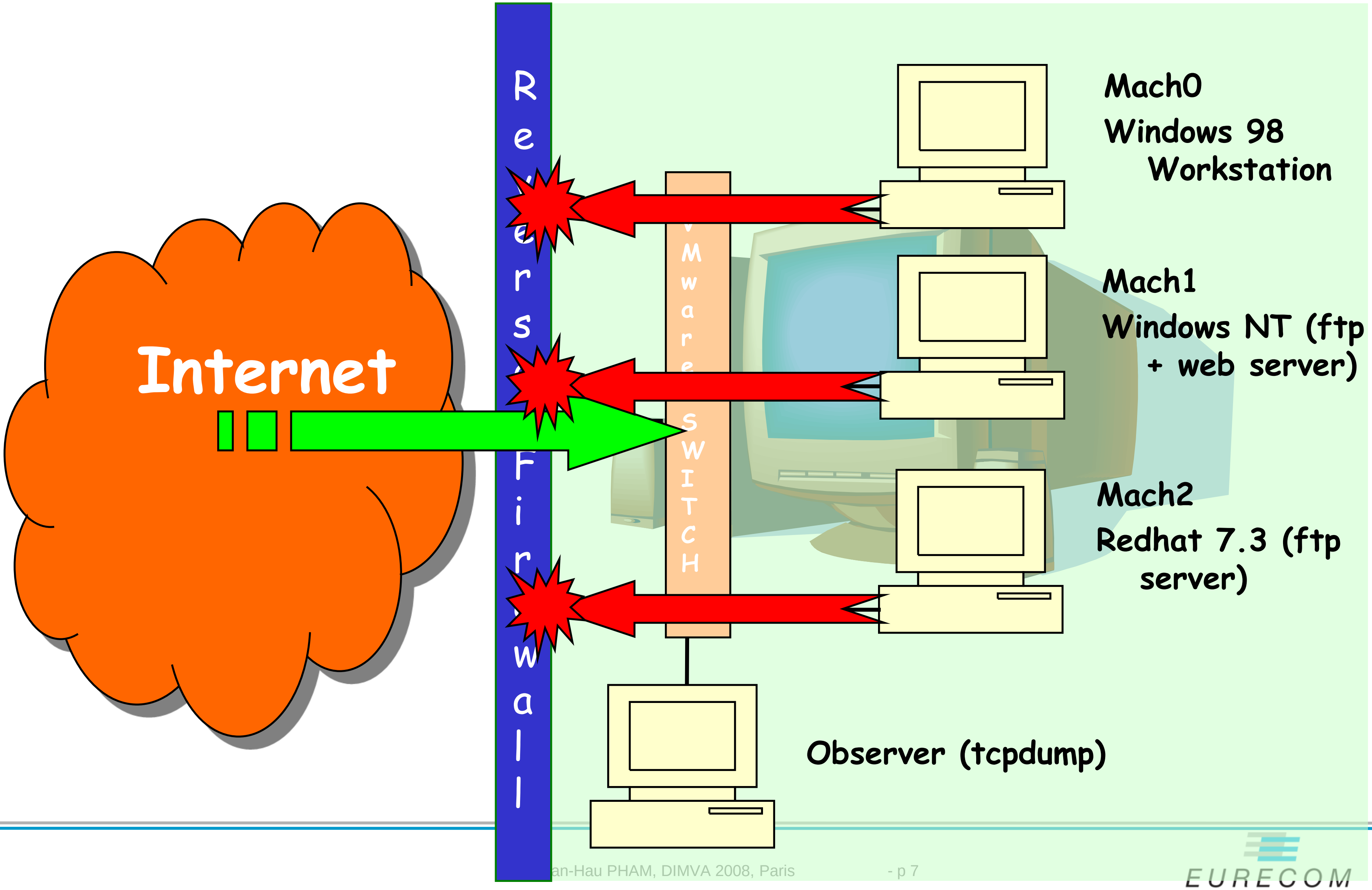
☞ At time = 4, another tool reuses attack vector A

☞ *Correlation period of trace A and B is from 1 to 3*

Outline

- Introduction
- **Method and implementation**
  - **Experimental Environment**
  - Approach
  - Results
- Conclusion

Platform


Internet
Reverse Firewall
VMware SWITCH
Mach0
Windows 98 Workstation
Mach1
Windows NT (ftp + web server)
Mach2
Redhat 7.3 (ftp server)
Observer (tcpdump)

Leurré.com: 50 platforms, 30 countries, 5 continents

Terminology

- **Cluster:** attacking sources leaving similar traces on our platforms
  - Traces: list of ports (ex 445 tcp , 139 TCP), amount of packets, attack duration,...
- **Cluster time series:** amount of sources, on a daily basis, associated to a given cluster on a given platform
- **Platform time series:** sum of all cluster time series associated to a given platform

Dataset description

- **15 months of data**
- **28 platforms**
  - ❑With the uptime rate higher than 90%
- **15 countries**

59,000 cluster time series,

a huge amount of data!!!

Outline

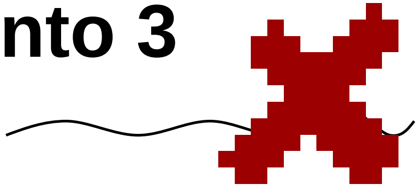
- Introduction
- Method and Implementation
  - Experimental Environment
  - Approach
  - Results
- Conclusion

Our approach

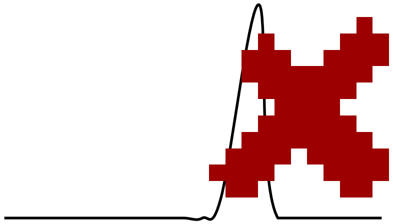
- **Preprocessing technique**
  - Reduce the number of clusters
- **Correlated groups of platform time series**
  - Instead of correlations between clusters
- **Root cause extraction**
  - Relate clusters time series to platform time series

Preprocessing technique

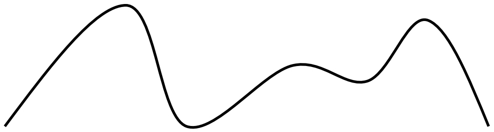
- **Cluster time series can be classified into 3 families:**
  - Stable time series:
    excluded since correlation is meaningless
  - Peaked time series:
    trivial cases, leave for future work
  - Strongly varying time series:
    strongly active attack tools, kept for our analysis



Stable time series



Peaked time series



Strong varying time series

We are left with 1% of the initial amount of time series

Groups of correlated platform time series (1)

Technique

- **We filter out the stable and peak time series to build platform time series**

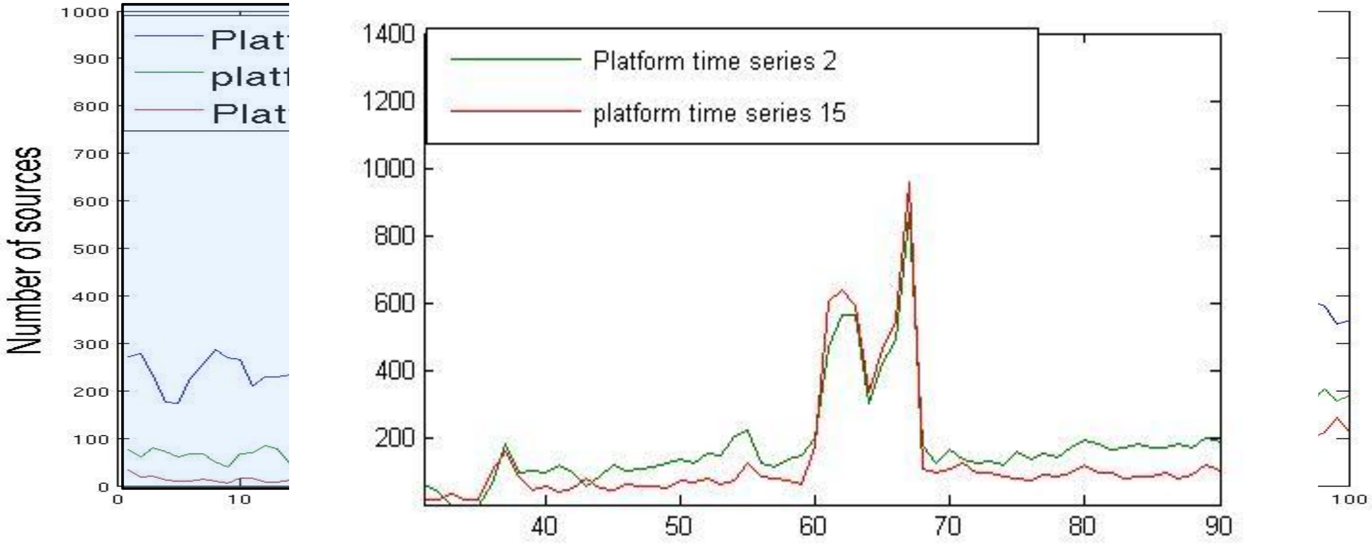
- **We use the sliding window to identify all periods where there exist groups of correlated platform time series**

→ ~ 28^2 *(450-30) instead of ~ 59000^2 *(450-30) operations to compute the correlation

Groups of correlated platform time series (2)

Example

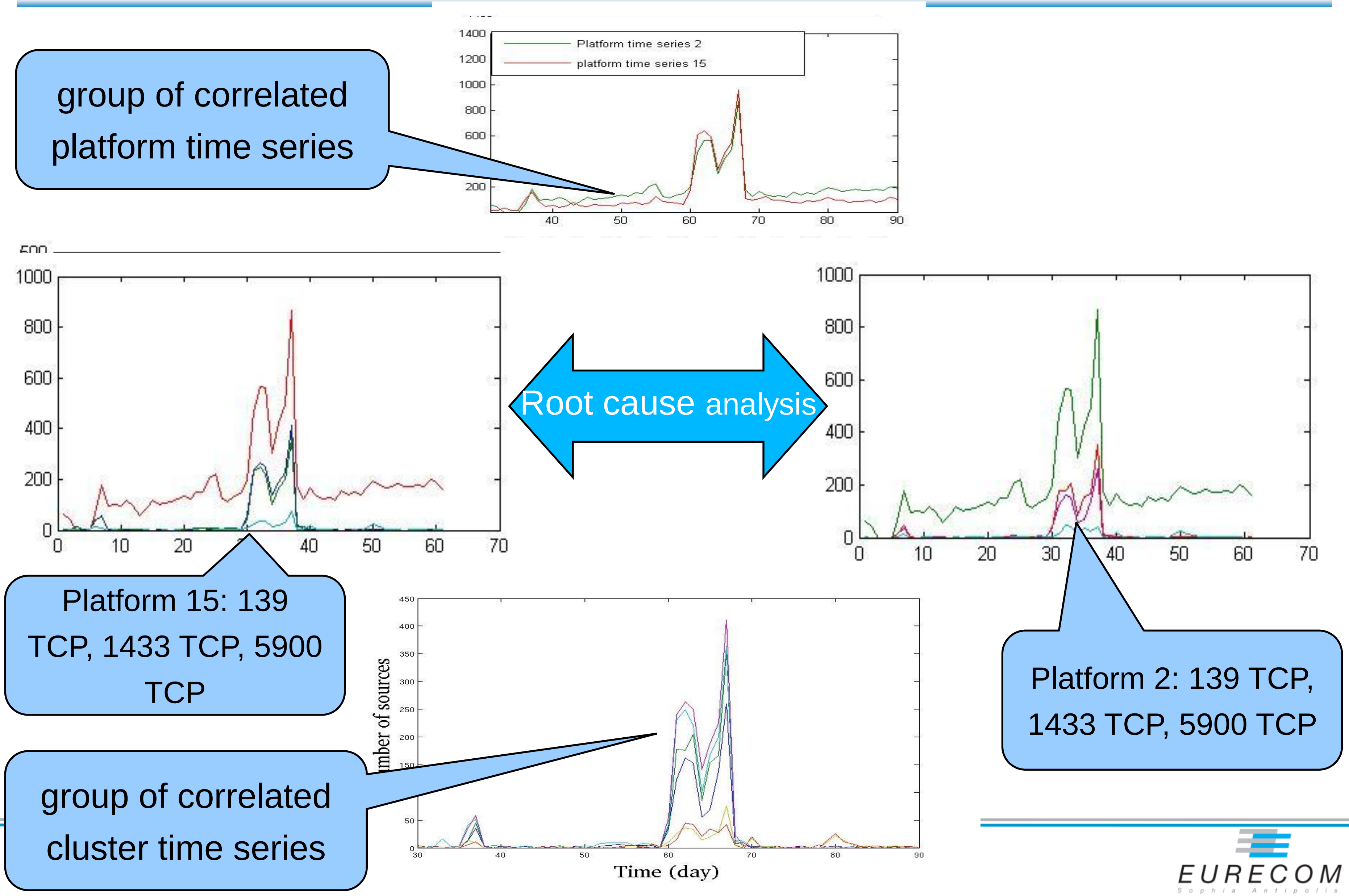
- **platform time series:** 1, 2, 15
- **Period:** from day 1 to day 100
- **Result:** correlation of platform time series 2 and 15 from day 30 to day 90



Root cause extraction (1)

- The root causes are clusters that explain the correlation of the groups of correlated platform time series
- In each correlated period, to identify them, we look for the clusters that are similar to the platform time series, platform by platform

Root cause extraction (2)



group of correlated platform time series

Platform 15: 139 TCP, 1433 TCP, 5900 TCP

Platform 2: 139 TCP, 1433 TCP, 5900 TCP

group of correlated cluster time series

Outline

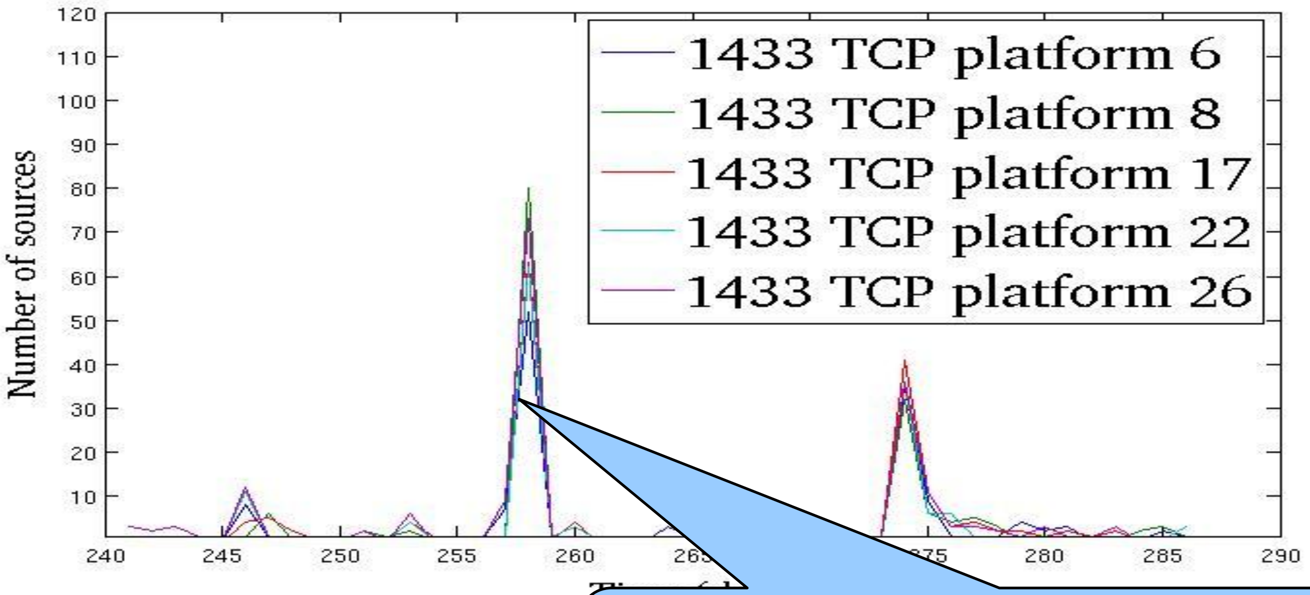
- Introduction
- Method and Implementation
  - Experimental Environment
  - Approach
  - **Results**
- Conclusion

Results

- **We found out 28 correlated groups involving 130 cluster time series, which can be classified into:**
  - Non multi-headed worms groups (21 groups)
    - Single root cause groups (10 groups)
    - Multiple root causes groups (11 groups)
  - Multi-headed worms (7 groups)

Single root cause (10 groups)

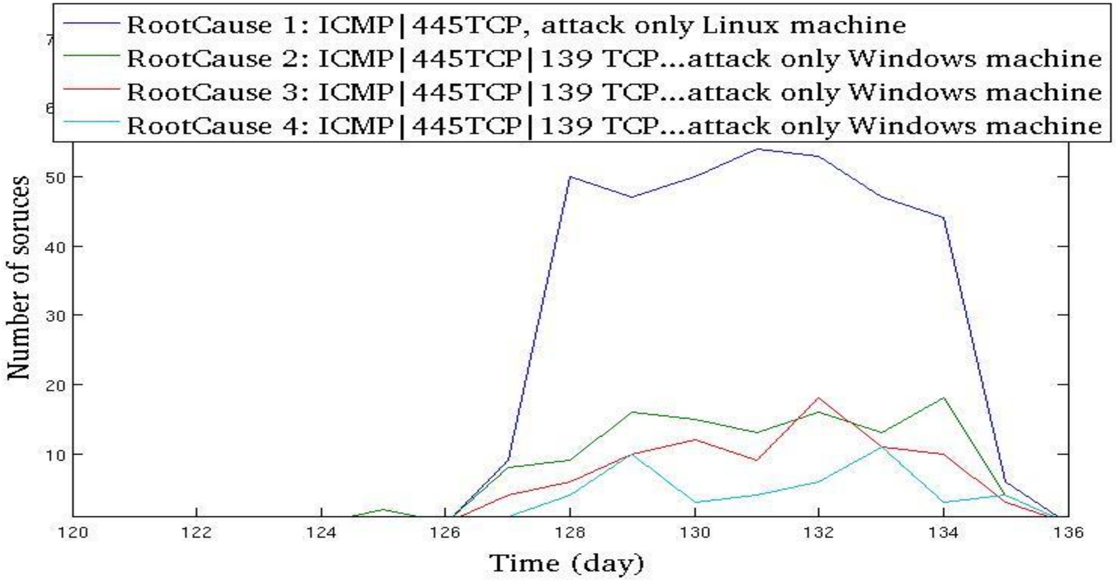
- **They correspond to phenomena where a single, and always the same, cluster is the root cause of the correlation of platform time series**



In most cases, sources of the attacks on different platforms are not the same

Multiple root causes groups

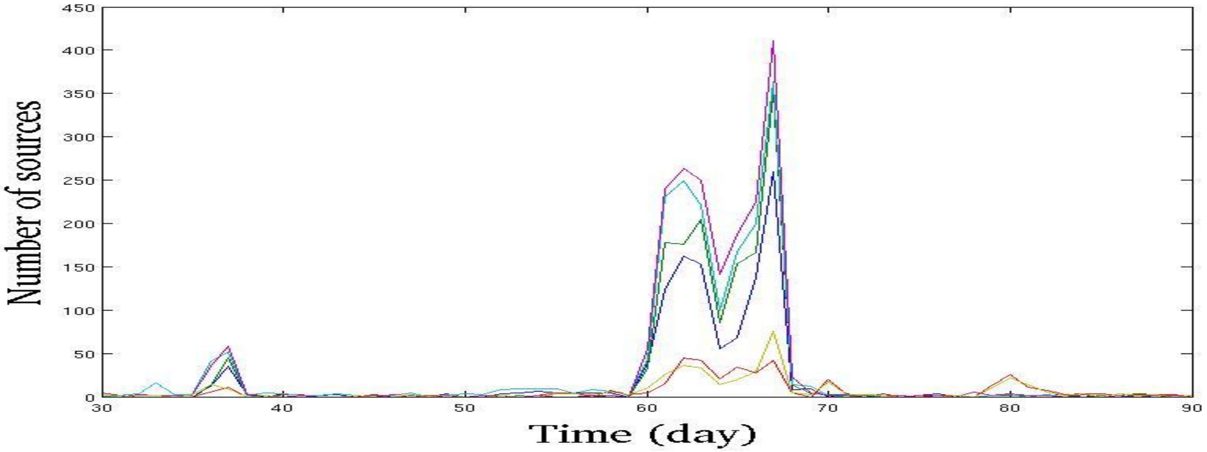
- **Non deterministic attack tools**
  - Attack the same list of ports but in different orders,…
  - Leave different traces → different clusters
- **Fingerprinting worms leave different attack traces on different operating systems**



Example of a multi-headed worm

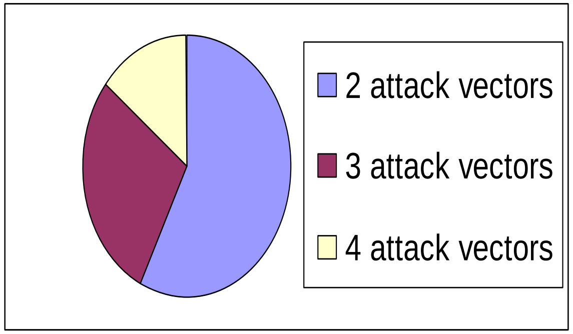
- Multi-headed worms carry many attack vectors, but they use only one of them to attack a given target.

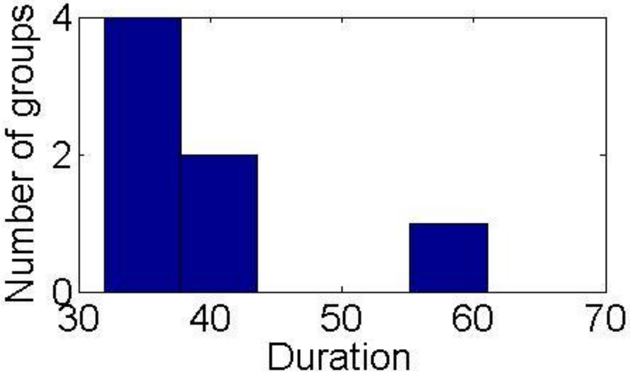
- Example: A multi-headed worm, observed on two platforms 2 and 15, has three attack vectors to attack 139 TCP, 1433 TCP, and 5900 TCP

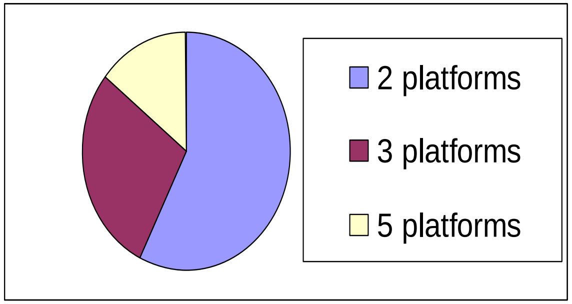


Some characteristics of multi-headed worms

- Around 60 % of multi-headed worms have 2 attack vectors
- 80 % of cases, the duration of appearance is from 30 to 40 days
- 60 % of them have been seen only on 2 platforms







Conclusion

- The approach based on platform time series works and it returns not only multi-headed worms, but other interesting phenomena.
- There are not so many multi-headed worms existing in the wild, and they have the locality property, and appear only in a short period of time

Future works

- Testing the brute-force approach on a limited amount of platforms to detect all possible correlation.
- Applying the method recursively
- Studying the peaked time series family