Toward a Source Detection of Botclouds: a PCA-based Approach

Badis HAMMI
Guillaume DOYEN
Rida KHATOUN

Autonomous Network Environment (ERA) team
Troyes University of Technology (UTT)
CNRS UMR 6281 ICD
Contrôle Autonome et Sécurité dans le Cloud Computing (CASCC) research project

IFIP International Conference on Autonomous Infrastructure, Management and Security
AIMS 2014

Plan

1. Introduction
   - Context
   - Research problem
   - Our contribution
2. Related work
   - Host Based IDS
   - Collaborative IDS
   - Source based IDS
3. Toward a source based approach
   - Our previous work
   - Detection approach
4. Numerical results
5. Conclusion and Future work

Plan

Context

Cloud computing

- Cloud computing is rapidly gaining ground
- Cloud computing market of $ 40.7 billion in 2010 will grow to more than $ 240 billion in 2020 [R.Stephan ET AL., 2010]

Cloud Services

- Infrastructure as a Service
- Platform as a Service
- Software as a Service

Cloud benefits

- Fast deployment
- Cost reduction
- Pay-per-use billing
- Massive scalability

Research problem

Malicious use of cloud computing

- Very dynamic and widely distributed attacks
- Attacker anonymity could be guaranteed

Botnets represent the greatest beneficiaries of this conversion into an attack support

- Setup on demand, at large scale
- Don't require a long dissemination phase
- Attack as a Service

Research problem : Botclouds

[P. Hayati ET AL., 2012]

- Botclouds' setup up on 5 famous CSPs
- Realization of many attacks (DDoS, shellcode, malware traffic, malformed traffic)
- Duration of 21 days (48 hours non stop for DDoS attacks)

[C. Kassidy ET AL., 2011]

- Set up of a botcloud on Amazon EC2
- Realization of DDoS attacks *(flooding and click fraud)*

Our contribution

Observation

- Successful realization of attacks
- No reactions or countermeasures from CSPs

Goal

- Development of a detection mechanism against malicious activity leveraged by botclouds

Originality

- Avoiding side effect damages $\rightarrow$ Source-based detection
- Detection of malicious activity $\rightarrow$ Consideration of system metrics
- Scalability support $\rightarrow$ Autonomous collaborative distributed detection system

Plan

Host Based IDS

Unsupervised Behavior Learning [J. D. Daniel et al., 2012]

- Host based IDS (CPU, MEM, TX, RX)
- Hypervisor level implementation
- Leverages Self Organizing Map (SOM)
- Looks at early deviations from the normal system behavior

Limits

- Can not build a global view of distributed attacks
- Not effecient in detecting fast-spreading attacks such as DDoS

Collaborative IDS

Firecol [J. François et al., 2012] A collaborative IDS

- Detects attacks as close as possible from the source
- Relies on multiple IPSs forming overlay networks of protection rings around subscribed customers

[J. Li et al., 2007] A hierarchical collaborative IDS

- Participating hosts are clustered into cooperating regions
- Using Markov model to aggregate alerts on hosts within a region
- Using sequential hypothesis tests to correlate findings across the regions

Limits

- Implementation close to, or even at target location
- Unavoidable side-effect damages

Source based detection

DWARD [J. Mirkovic ET AL., 2005] a DDoS defense mechanism

- Autonomously detects and stops attacks
- Monitoring of two-way traffic flows
- Comparison with normal flow model

Limits

- Large number of independent network administrative domains that must deploy it to be efficient

Plan

Overall context

- Public CSP providing IaaS (Such as Amazon EC2)
- The CSP owns physical servers that host VMs belonging to tenants
- A malicious user build a botcloud (one tenant)
- Monitoring in a black box way
- Monitoring metrics available at the hypervisor level
  CPU (%), MEM (KB/s), TX (Kb/s), RX (Kb/s)

Experimental framework and scenarios

Parameter	Description
Attack types	UDP Flood , TCP SYN flood
Experimetation time	1h (normal state) $\rightarrow$ 1h (attack) $\rightarrow$ 1h (back to normal)
Monitoring frequency	1 minute
Environment	LXC-Linux (Planet-Lab)
Botcloud (botnet)	Hybrid_V1.0
Data collected	16,65 GB

TABLE: Summary of the scenario numerical parameters

Experimental framework and scenarios

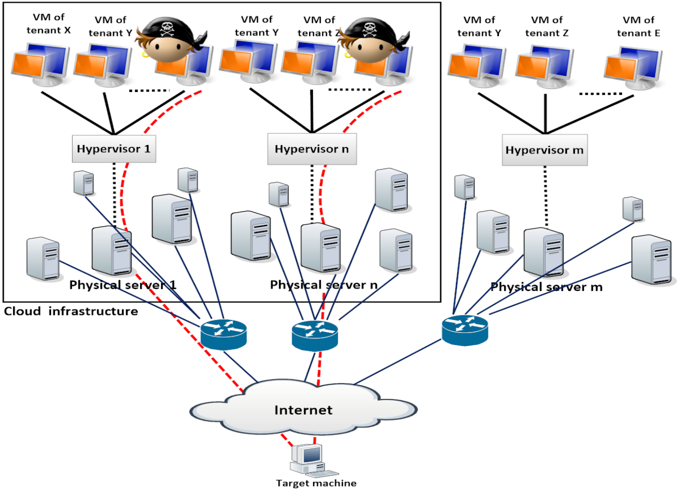

FIGURE: Overview of the experimentation environment

Our previous work

Understanding botclouds over a Principal Component Analysis [NOMS 2014]

- Exhaustively understanding and featuring a botcloud in its execution environment
- Highlighting the strong similarity of bots' behavior
- Highlighting correlations between the different metrics of a botcloud
- Detection and separation of the attack phase from the idle one for two study-cases (UDP flood and TCP SYN flood)

A factorial space for a system-based detection of botcloud activity [NTMS 2014]

- Highlighting constant metrics' contributions in eigenvectors' matrices of botcloud activity for whatever attack rate
- Definition of a factorial space as a reference to detect DDoS

Principal Component analysis (PCA)

- PCA is a descriptive statistical method belonging to the factorial category
- Explains the variance-covariance matrix of a set of variables through a few new variables (Principal Components)
- Aims at
  - Easing the exploration and analysis of high dimensional data by reducing their dimensions
  - Visualization and interpretation of multi dimensional data
  - Understanding the relationship between the different variables
- Benefit : does not require any distribution assumption on the data to process

Detection approach

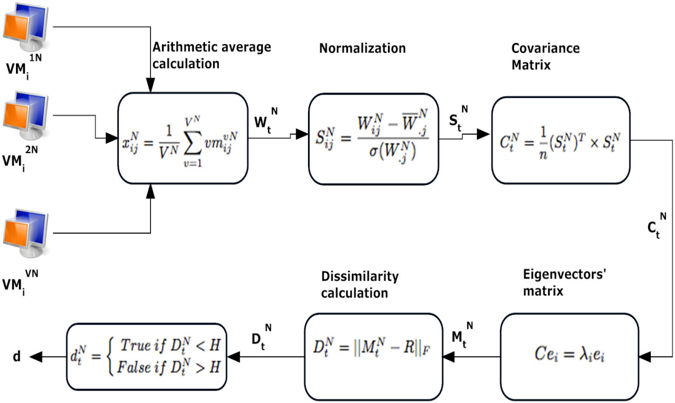

FIGURE: Detection algorithm steps

Plan

Scenario : UDP Flood case

UDP flood attack rate	#Physical servers	#Tenants (incl.)	#VMs (incl.)	#Attacking VMs
8 Mb/s	41	123	1,288	41
16 Mb/s	41	118	1,261	41
40 Mb/s	43	123	1,310	43
56 Mb/s	41	114	1,241	41
80 Mb/s	40	103	1,198	40

TABLE: Summary of the scenarios numerical parameters

ROC curves

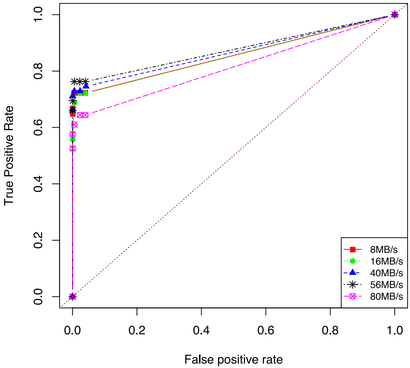

FIGURE: Roc curves of the five cases (compared to 25 legitimate tenant)

Matthews Correlation Coefficient and Accuracy

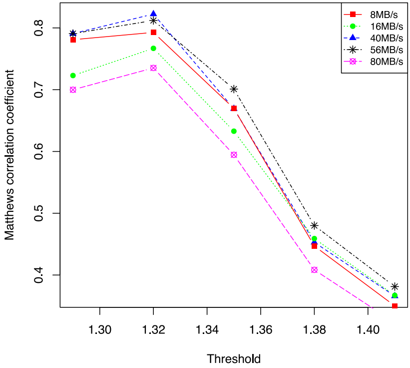

FIGURE: Matthews correlation coefficient

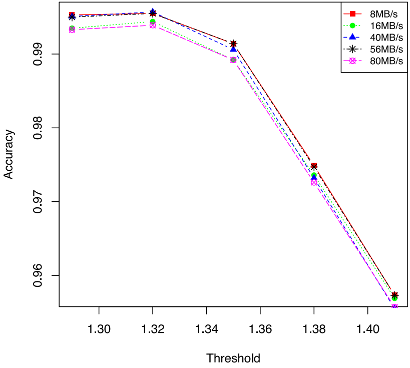

FIGURE: Accuracy

Plan

Conclusion

Conclusion

- Presentation of novel source-based detection approach based on PCA and system metrics
- Validation of the approach through extensive simulations relying on real traces obtained through *in situ* experimentations
- Proving efficiency and resiliency of our detection algorithm over different statistics that took into account a large workload amount

Future work

- Proposing a distributed approach of our detection algorithm
- Extending the study in order to consider other attacks such as application level attacks
- Development of an autonomous self-protection system for CSPs against DDoS attacks leveraged by a cloud infrastructure

Daniel Joseph Dean, Hiep Nguyen, and Xiaohui Gu.
Ubl : unsupervised behavior learning for predicting performance anomalies in virtualized cloud systems.
In *Proceedings of the 9th international conference on Autonomic computing*, ICAC '12, pages 191–200. ACM, 2012.

botcloud an emerging platform for cyber-attacks, October 2012.
http ://baesystemsdetica.blogspot.fr.

Kassidy Clark, Martijn Warnier, and Frances M. T. Brazier.
The evolution of cloud computing markets.
Closer 11, 2011.

Ji Li, Dah-Yoh Lim, and Karen Sollins.
Dependency-based distributed intrusion detection.
In *Proceedings of the DETER Community Workshop on Cyber Security Experimentation and Test on DETER Community Workshop on Cyber Security Experimentation and Test 2007*, DETER, pages 8–8. USENIX Association, 2007.

S. Ried, H. Kisker, and P. Matzke.
The evolution of cloud computing markets.
Forrester research paper, 2010.

J. Mirkovic and P. Reiher.
D-ward : a source-end defense against flooding denial-of-service attacks.

Dependable and Secure Computing, IEEE Transactions on, 2(3) :216– 232, july-sept. 2005.

Hammi Badis, Guillaume Doyen, and Rida Khatoun.
Understanding botclouds from a system perspective : a principal component analysis.
In *Network Operations and Management Symposium (NOMS 2014)*. IFIP/IEEE, may 2014.

Hammi Badis, Rida Khatoun, and Guillaume Doyen.
A factorial space for a system-based detection of botcloud activity.
In *Sixth IFIP International Conference on New Technologies, Mobility and Security (NTMS'2014)*. IFIP/IEEE, March 2014.

Jérôme François, Issam Aib, and Raouf Boutaba.
Firecol : a collaborative protection network for the detection of flooding ddos attacks.
IEEE/ACM Trans. Netw., 20(6) :1828–1841, 2012.


Merci

Annexe

- $MCC = \frac{TP*TN-FP*FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
- $ERR = \frac{FalsePositive+FalseNegative}{TruePositive+FalsePositive+TrueNegative+FalseNegative}$
- $ACC = \frac{TruePositive+TrueNegative}{TruePositive+FalsePositive+TrueNegative+FalseNegative}$