Fighting an unfair battle: Unconventional defences against sophisticated threats

Nikos Virvilis

September 2015

Overview

- Motivation
- Current challenges
- Related work
- Contribution
- Advanced Persistent Threat manifestations
- Redefining the security architecture
  - Proposed model
  - Evaluation results
- Future work

Motivation

- Global spending of more than 71B USD in 2014 for cyber security
  - Yet, the number and complexity of cyber attacks is increasing
- Most developed nations have recognized cyberspace as the fifth domain of warfare
- Considering the severe cyber attacks that we have witnessed in recent years, it is becoming evident that traditional security solutions are ineffective against sophisticated attackers

Motivation Cont.

- *"Forget anything you know about current defensive technologies and state-of-the-art security solutions. We have them all, and we know that are not effective against our threat actors. We need something new, robust and practical. Academic solutions are not acceptable. Think out of the box."*
  – NCI Agency, Cyber Defense Director

Current Challenges

- The vast majority of security solutions in use today were designed in the 80's and 90's
  - High-speed networks, mobile users, complex web platforms, social networks and cloud services have created new challenges
  - Security solutions are continuously getting improved, yet they keep failing over and over again
  - The root cause lies in their design: Focused on prevention & real time response
- Defense is an unfair game:
  - Defenders will always have to find and fix all vulnerabilities and attack paths, while the attackers only need to exploit one

Related work

- Limited research on APT
- Significant research on the use of deception, but only for very specific tasks:
  - Honeypots for worm detection, identification of 0-day exploits
  - Honey files or honey records for the detection of insiders
- Proposed deception models are only focused on insider threat detection
- No comprehensive models which combine multiple deception-based attack indicators focusing on APT

Contribution

- A holistic view on APT, and its three distinct manifestations (i.e. attack paths)
  - The unique characteristics of each attack path are presented in detail, along with countermeasures for limiting the exposure
- Proposal of a sophisticated APT detection model
  - Combines anomaly detection with multiple deception-based attack indicators
  - Correlates both current and historical events over a wide period of time
  - It has been implemented and evaluated; the results highlight its superior effectiveness in attack detection

APT Manifestations - External

- External APT attacks
  - The most common type of APT attack, which originates from outside the targeted organization's perimeter
  - These attacks entail very little risk for the attackers, as it is very difficult to attribute them to a specific individual
- In the vast majority of external APT attacks, sophisticated malware was used:
  - It enabled the attackers to control and interact with the compromised systems, use them as hopping points for attacking other systems, or in some cases (i.e. Stuxnet) as an autonomous weapon

Sophisticated APT malware

	Stuxnet	Duqu	Flame	Red October	Mini Duke	Regin
Active since	June 2009 (2005)	Nov. 2010	May 2012 (2006)	May 2007	June 2011	Possibly 2003
Detected	June 2010	Sept. 2011	May 2012	Oct. 2012	Feb. 2013	Nov. 2014
PE Type	DLL		OCX	EXE	EXE	SYS/DLL
Initial infection	Unknown	MS Word	Unknown	MS Excel / Word, Java	PDF	Unknown
Replication	Removable drives, network	Manual replication only				
Rootkit module	Yes		No			Yes
Key logging	No	Yes			No	Yes
Evasion	Yes			No	Yes	Yes
Encryption	XOR	XOR, AES-CBC	XOR, RC4, Substitution	XOR	Unique per victim, XOR, ROL	RC5 Variant, XOR
Target	Sabotage	Information gathering				

External attacks - Contribution

- Detailed technical review of sophisticated malware:
  - A technical review of Stuxnet, Duqu, Flame, Red October, Mini Duke and Regin is presented, highlighting the common characteristics and techniques of such advanced malware
  - This allowed a better understanding of why our existing security defenses have failed to protect us against such threats

Related Publications:

- Virvilis N., Gritzalis D., "Trusted Computing vs. Advanced Persistent Threats: Can a defender win this game?", in Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (ATC-2013), pp. 396-403, IEEE Press, Italy, December 2013
- Virvilis N., Gritzalis D., "The Big Four - What we did wrong in Advanced Persistent Threat detection", in Proc. of the 8th International Conference on Availability, Reliability and Security (ARES-2013), pp. 248-254, IEEE, Germany, September 2013

APT Manifestations - Internal

- Insider attacks and APT have been considered as two different threat groups in the literature
  - However there is (at least one known) incident, where an APT group has used an insider to perform malicious actions on their behalf (Stuxnet)
  - Thus, this thesis considers the insider threat as a subset of APT, and recommends that robust APT detection models should also be able to detect insider attacks

Internal attacks - Contribution

- Robust insider threat prediction models
  - Kandias M., Mylonas A., Virvilis N., Theoharidou M., Gritzalis D., “An Insider Threat Prediction Model”, in Proc. of the 7th International Conference on Trust, Privacy, and Security in Digital Business (TrustBus-2010), pp. 26-37, Lopez J., et al. (Eds.), LNCS-6264, Springer, Spain, August 2010
  - Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in Proc. of the 6th International Conference on Critical Infrastructure Security (CRITIS-2011), pp. 93-103, Springer (LNCS 6983), 2013
- Ineffective against APT attacks that use an insider for performing a limited set of malicious actions

APT Manifestations - Indirect

- Attacks which have as an end goal the exploitation of a specific target, but instead of attacking directly the target's infrastructure, they focus on alternative entry points that are easier to compromise
- Mobile devices and more specifically smartphones and tablets are ideal entry points:
  - Access to corporate resources
  - Few security solutions for mobile platforms, very basic
  - Weaker OS level protection mechanisms
  - Auditing is non-existent

Indirect Attacks - Contribution

- Our experiments reveal that the level of security offered on mobile platforms against web based attacks is significantly lower than that for desktop platforms
  - Web based attacks are a very common entry point for APT

Related Publications:

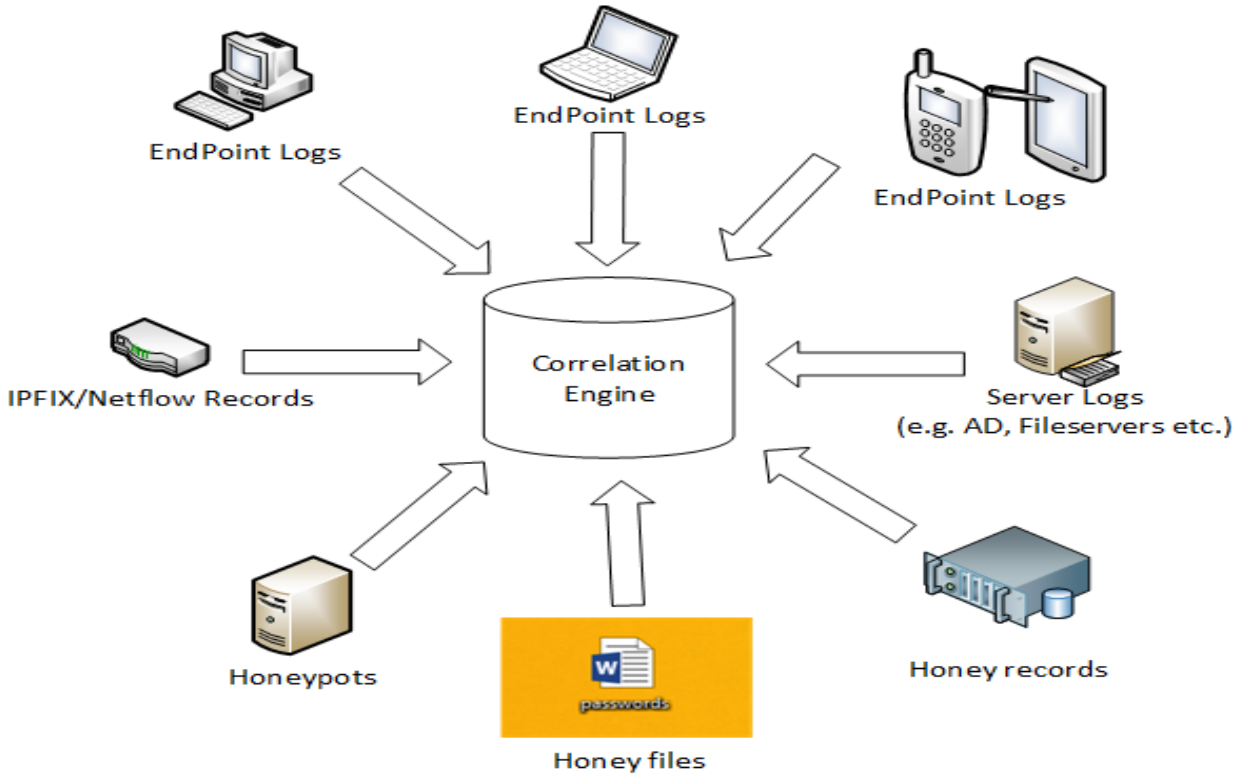
- Virvilis N., Mylonas A., Tsalis N., Gritzalis D., "Security Busters: Web Browser security vs. rogue sites", Computers & Security, 2015 (to appear)
- Virvilis N., Tsalis N., Mylonas A., Gritzalis D., "Mobile devices: A phisher's paradise", in Proc. of the 11th International Conference on Security and Cryptography (SECRYPT-2014), pp. 79-87, ScitePress, Austria, August 2014

Redefining the Security Architecture

- Based on the aforementioned topics, its is evident that a radical change in the way we defend our assets is required
- The use of deception, is a very promising approach for attack detection, regardless of the attacker's skills and capabilities
  - Honeypots on operational segments
  - Honey tokens on DB's
  - Honey files on file servers
  - Fake user accounts and authentication tokens
  - Fake social network personas

A novel APT detection model

- The model combines anomaly detection with multiple deception-based attack indicators and correlates both current and historical events over a wide period of time



Attack indicators

Multiple, successful logins from an account to multiple systems	High volume of traffic between two systems
Successful logins during irregular business hours and/or non-working days	Connections to countries where the organization does not do business with
Multiple login failures for one or more accounts	High resource utilization
A large number of network connections originating from a system	Access to multiple files from a single account within a small time period
Network connections between workstations	Multiple or large (multiple records) database queries from a single account within a small time period
Network connections from a workstation to multiple servers	Network traffic destined to unused IP ranges
Network connections originating from server(s), towards workstation(s) or the Internet	Network traffic or interaction with honeypots
Protocol anomalies	Authentication attempt using a honey-account
TCP connections which last for several minutes	Communication attempts (e.g. emails) towards a honey-account
TCP connections for which the transmitted traffic from the internal system is significantly more than the received traffic	Access of honey files (e.g. on a file server)
Periodical connections	Access of honey records (e.g. database records, DNS records)

Threat ratings

- As not all of the aforementioned attack indicators have the same severity or are prone to false positives, each indicator has been assigned a "Threat Rating" (TR) with values: **Low**, **Medium**, **High** or **Critical**
- Indicators that when triggered there it is a high probability of malicious activity (true positive), have received a higher threat rating
- If multiple indicators are triggered from the same origin (e.g. same IP address or username) then, the Aggregated Threat Rating is equal to the sum of the independent Threat Ratings *(TR)*:

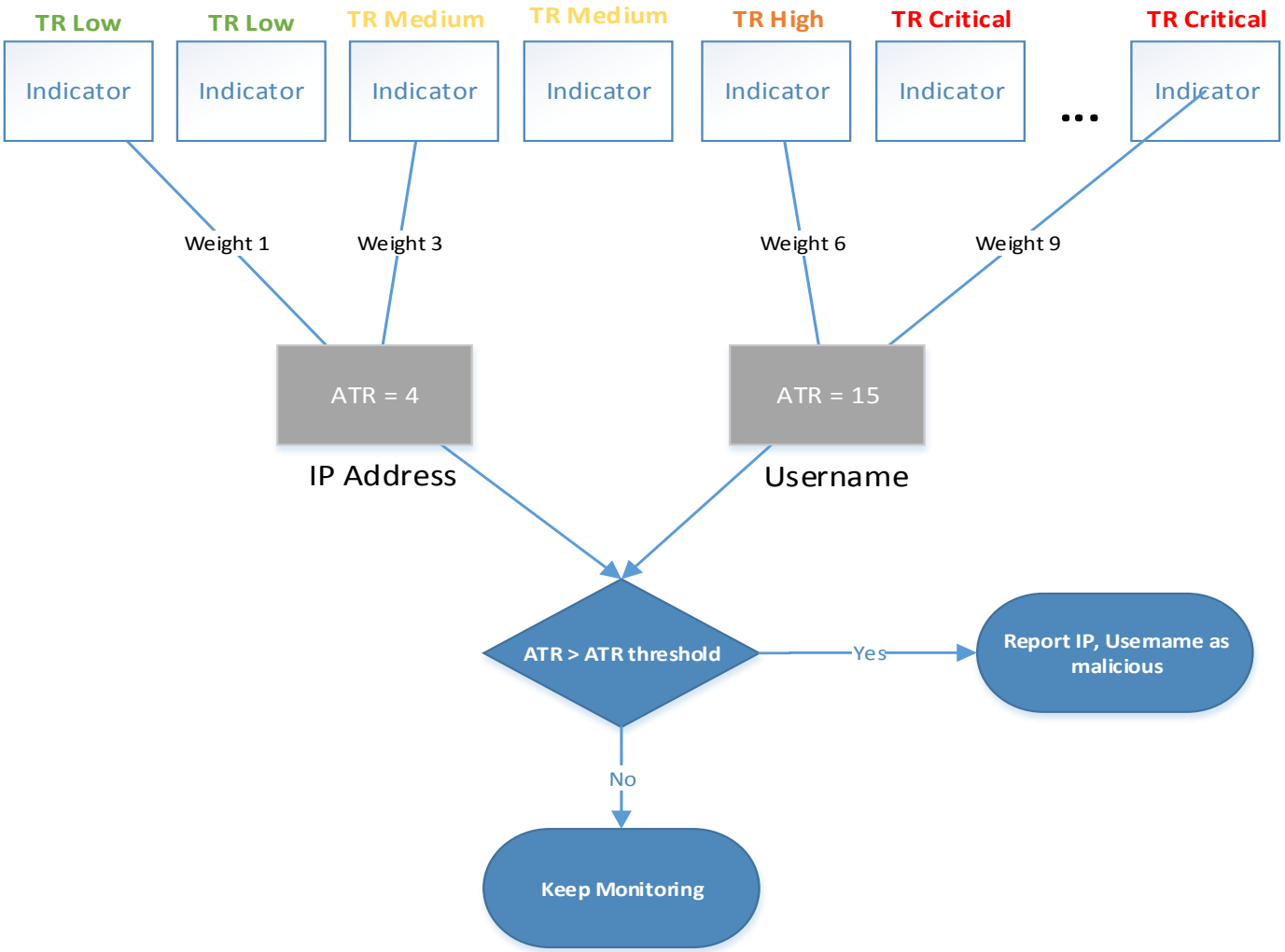
$$ATR = TR_1 + TR_2 + \ldots + TR_n$$

- The higher the value of the *ATR*, the higher the probability of an actual attack
- If **ATR >= ATR_Threshold** then **alert**

Correlation Window

- Three correlation windows have been defined:
  - Short (i.e. minutes), where all relevant captured data between the current time minus Nm (minutes) will be analyzed
  - Medium (i.e. hours): all relevant captured data between the current time minus *Nh* (hours) will be analyzed
  - High (i.e. days): all relevant captured data between the current time minus *Nd* (days) will be analyzed
  - The default *N* values (based on which this model has been evaluated are: *Nm*=30, *Nh*=24, *Nd*=15 (configurable)

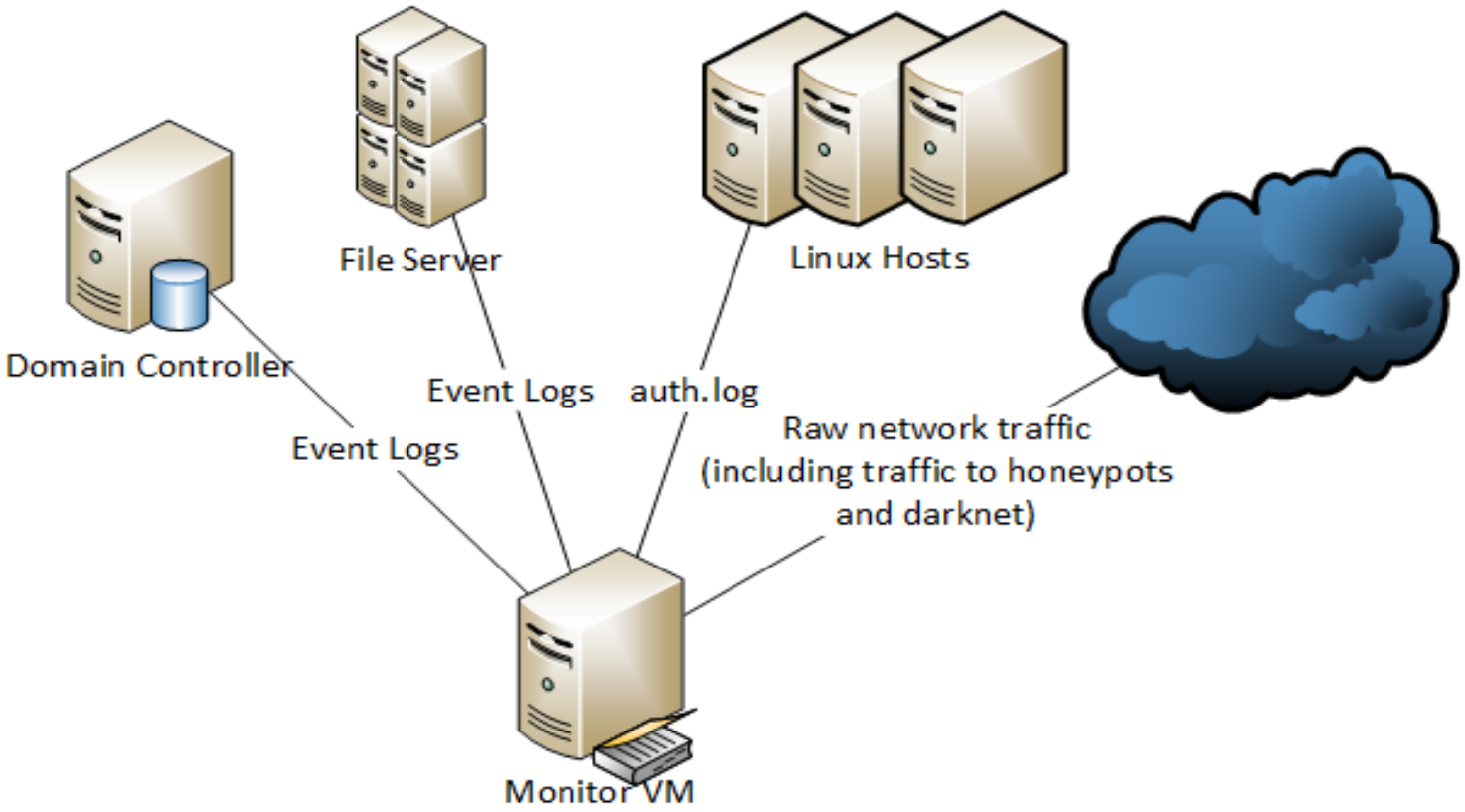
ATR Calculation Process



Proof of Concept Implementation

- A subset of the proposed model was built and evaluated. The model consists of an SQLite database and three modules:
  - The first module is responsible for parsing the Windows binary XML logs (.evtx) and Linux log files (syslog)
  - The second module is responsible for the analysis of network traffic
  - The results from both modules are stored in the database and are analyzed by the correlation engine (third module)
  - Set up on an Ubuntu Linux 14.04 LTS VM, with 2 GB RAM, 1 vCPU and 20 GB of disk space
  - The Bro network security monitor (Bro 2015) was used for the creation of network traffic metadata
  - The honeyd service was used for the creation of honeypots

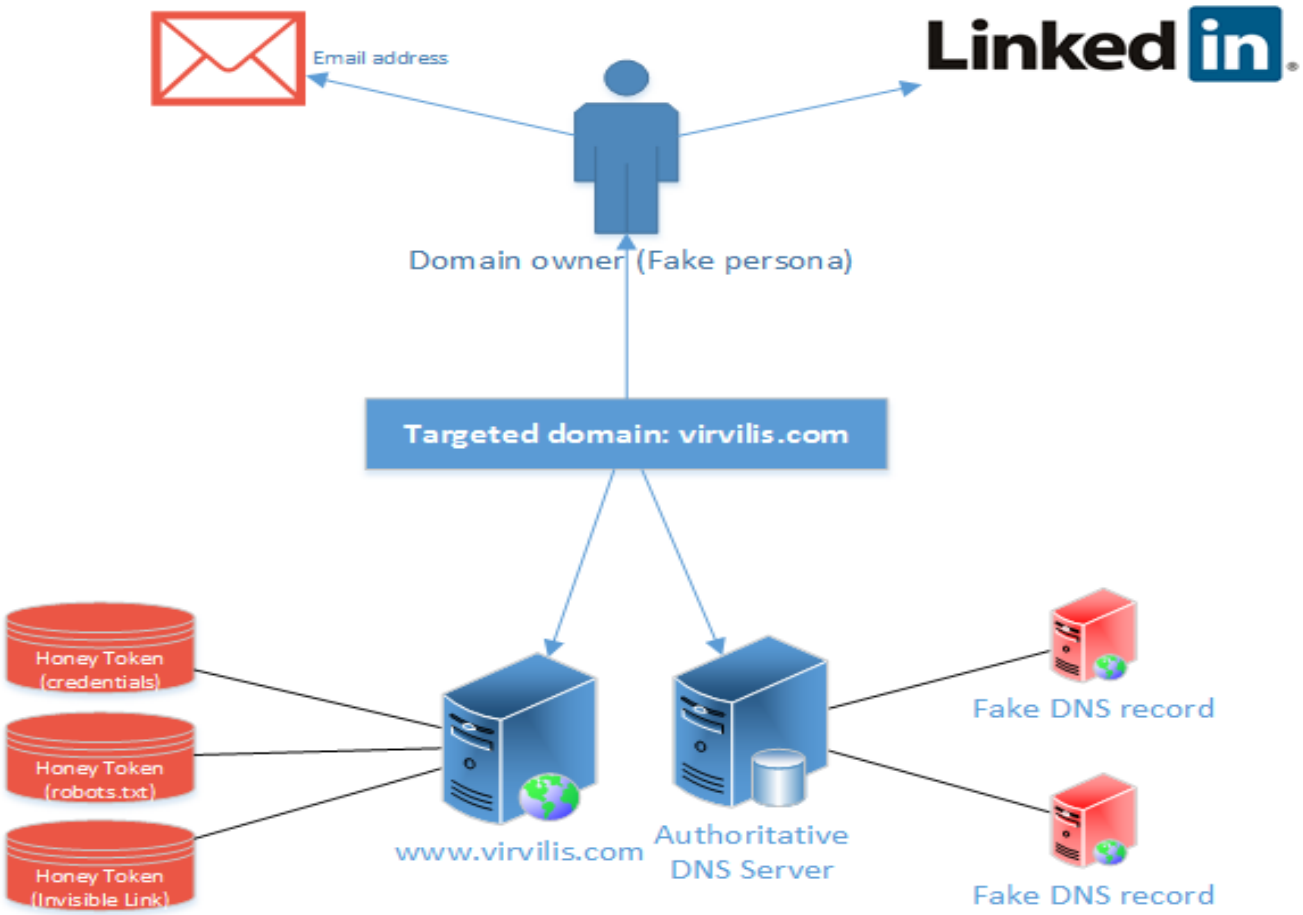
Proof of Concept Implementation – Data input



Model Evaluation

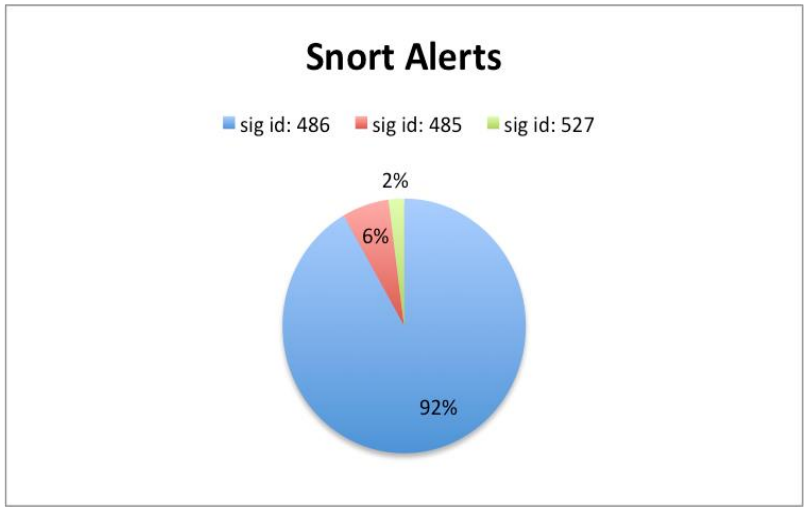
- Four expert penetration testing teams (two military and two from the industry) participated in two attack scenarios (CTF's)
- The first scenario was simulating an external attack (one flag)
- The second scenario was simulating an attack on an internal network (two flags)
- Two of the teams were informed about the use of deception countermeasures in advance

External Scenario Architecture



External Evaluation Results

Detection Mechanism	Number of True Positive alerts (True malicious IP addresses)	Number of False Positive alerts
Proposed model	**3/3**	**16**
Snort IDS	**0/3**	**397**



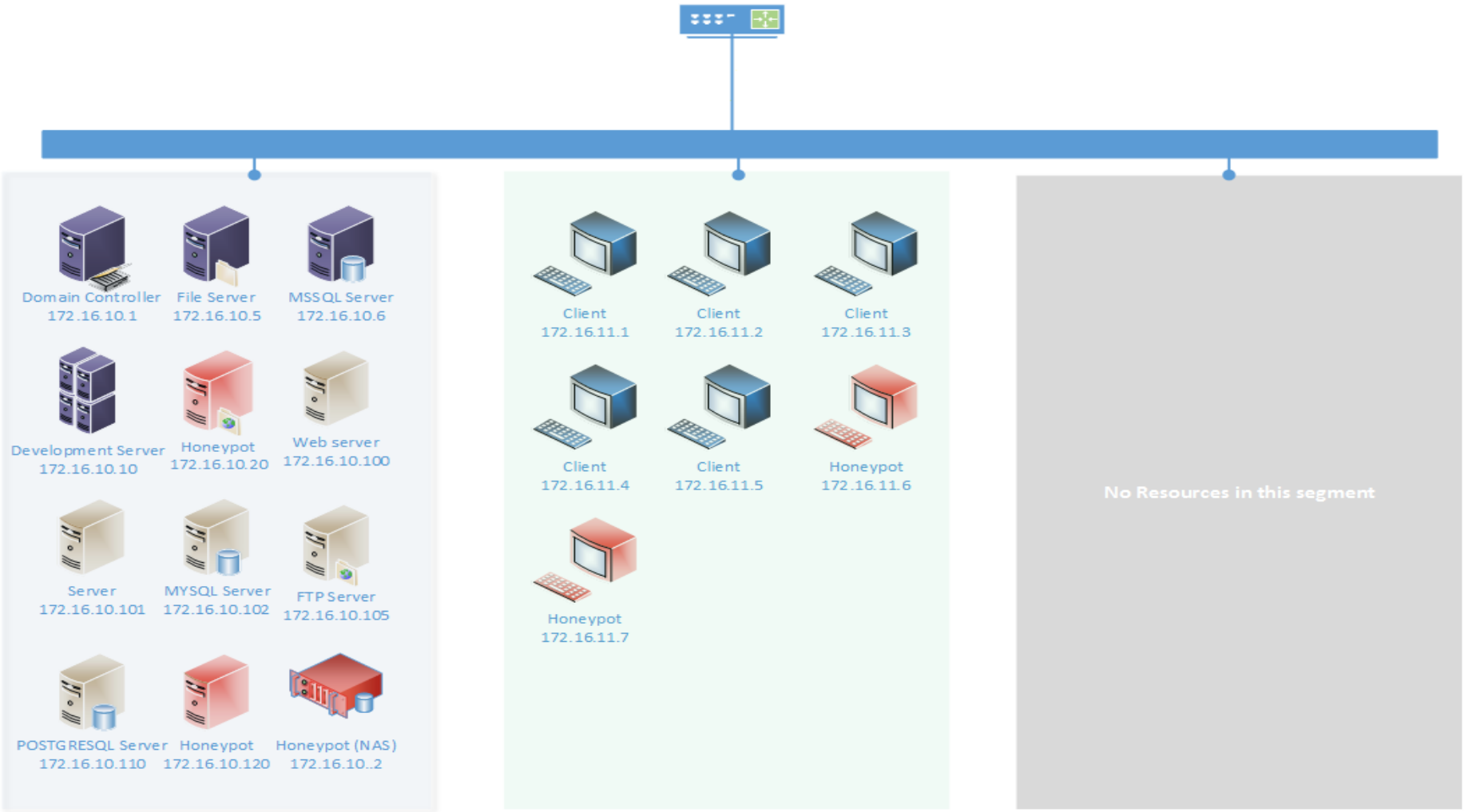
Snort rules that were triggered:

485: ICMP packet filtered

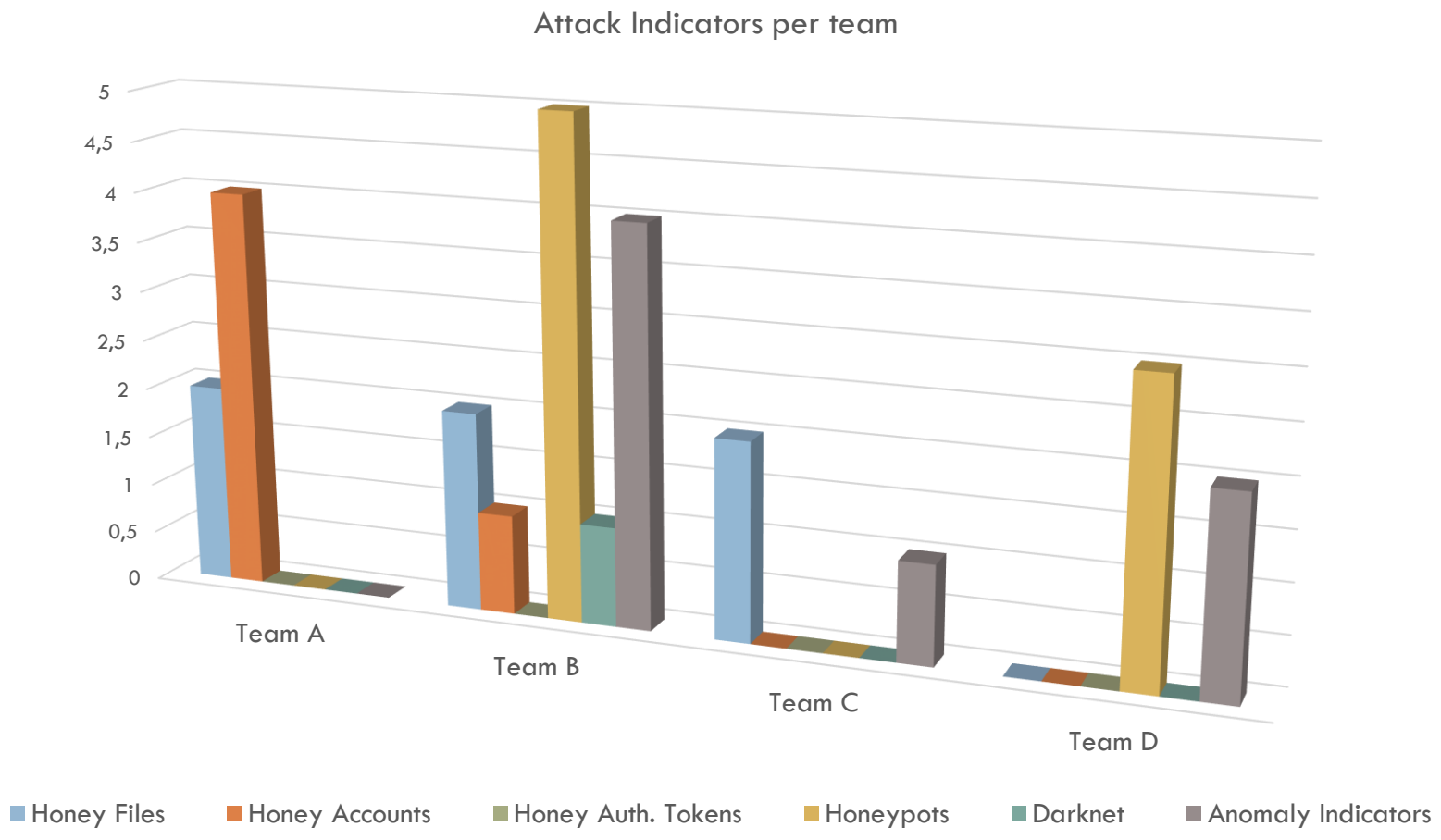
486: ICMP destination unreachable

527: Packets with same source and destination IP

Internal Scenario Architecture



Internal Evaluation Results



Internal Evaluation Results – Cont.

Team	Snort Alerts (13 True Positive out of 2930 alerts in total)	Proposed model's Alerts
A	**(Nothing detected by Snort)**	Access to honeyfiles
		Authentication attempts to other systems using honey accounts
B	Failed authentication attempt to telnet server	Access to honeyfiles
		Authentication attempts to other systems using honey accounts
	Failed authentication to the FTP server	Connections to honeypots
		Connections to darknet
		Multiple (4) anomaly based attack indicators
C	Access of the IIS server	Access to honeyfiles
		One anomaly based attack indicator
D	Failed authentication attempts	Connections to honeypots
	Access of the IIS server	Multiple (2) anomaly based attack indicators

Discussion

- The model accurately reports offensive IP's / usernames, and the type of offenses that were performed
  - Saves the analysts from a flood of (in most cases) meaningless alerts
  - In contrast with traditional security solutions, deception-based attack indicators are **very hard to evade**
  - The teams that were informed about the use of such indicators triggered the same indicators as the teams that were not
  - Teams reported that knowing that such indicators were in place made their task more challenging, as it slowed them down significantly in the fear that every action they made could trigger an alert

Related Publications:

- Virvilis N., Serrano O., Dandurand L., "Big Data analytics for sophisticated attack detection", ISACA Journal, Vol. 3, 2014
- Virvilis N., Vanautgaerden B., Serrano O., "Changing the game: The art of deceiving sophisticated attackers." in Proc. of the 6th International Conference on Cyber Conflict (CYCON-014), pp. 87-97, Estonia, June 2014

Limitations

- Significant effort for setting up, configuring, fine-tuning and maintaining
- It complements existing security solutions (does not replace them)
- Evaluation introduces potential bias as APT groups had to be simulated by penetration testing experts
  - Short time
  - No 0-days or spear phishing
  - Direct comparison with existing security solutions is hard

Future work

- Future work should focus on the likely benefits of introducing psychological profiling, to complement the proposed APT detection model
  - This could potentially allow the detection of skillful insiders, who, owing to their thorough knowledge of the infrastructure, are aware of all the attack indicators (including the deception-based ones) and thus are able to evade them
- On the more practical side, future work should investigate ways to automate the model's deployment, and especially the deployment of deception-based indicators
- Proposed for funding as a NATO Innovation project

References

1. Gritzalis D., Katos V., Katsaros P., Soupionis Y., Psaroudakis J., Mentis A., “The Sphinx enigma in critical VoIP infrastructures: Human or botnet?”, in Proc. of the 4th International Conference on Information, Intelligence, Systems and Applications (IISA-2013), IEEE Press, Greece, July 2013.
2. Gritzalis D., Stavrou V., Kandias M., Stergiopoulos G., “Insider Threat: Enhancing BPM through Social Media”, in Proc. of the 6th IFIP International Conference on New Technologies, Mobility and Security (NMTS-2014), Springer, UAE, April 2014
3. Gymnopoulos L., Dritsas S., Gritzalis S., Lambrinoudakis C., "Grid security review", in Proc. of the MMM-ACNS-2003 2nd International Workshop on Mathematical Methods, Models and Architectures for Computer Networks Security (MMM '03), V. Gorodetski (Ed.), Lecture Notes in Computer Science (LNCS 2776) , Springer, St. Petersburg, September 2003.
4. Kandias M., Galbogini K., Mitrou L., Gritzalis D., "Insiders trapped in the mirror reveal themselves in social media", in Proc. of the 7th International Conference on Network and System Security (NSS 2013), pp. 220-235, Springer (LNCS 7873), Spain, June 2013.
5. Kandias M., Mylonas A., Theoharidou M., Gritzalis D., “Exploitation of auctions for outsourcing security-critical projects”, in Proc. of the 16th IEEE Symposium on Computers and Communications (ISCC ‘11), Tinnirello I., et al (Eds.), pp. 646-651, Greece, June 2011.
6. Kandias M., Mylonas A., Virvilis N., Theoharidou M., Gritzalis D., “An Insider Threat Prediction Model”, in Proc. of the 7th International Conference on Trust, Privacy, and Security in Digital Business (TrustBus-2010), pp. 26-37, Lopez J., et al. (Eds.), LNCS-6264, Springer, Spain, August 2010.
7. Kandias M., Stavrou V., Bosovic N., Mitrou L., Gritzalis D., “Proactive insider threat detection through social media: The YouTube case”, in Proc. of the 12th Workshop on Privacy in the Electronic Society (WPES-2013), pp. 261-266, ACM Press, Berlin, November 2013
8. Kandias M., Stavrou V., Bozovic N., Mitrou L., Gritzalis D., "Can we trust this user? Predicting insider’s attitude via YouTube usage profiling", in Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (ATC-2013), pp. 347-354, IEEE Press, Italy, December 2013.
9. Kandias M., Virvilis N., Gritzalis D., "The Insider Threat in Cloud Computing", in Proc. of the 6th International Conference on Critical Infrastructure Security (CRITIS-2011), pp. 93-103, Springer (LNCS 6983), 2013.
10. Kokolakis S., Gritzalis D., Katsikas S., "Generic Security Policies for Healthcare Information Systems", Health Informatics Journal, Vol. 4, No. 3 . 4, pp. 142-159, 1999.
11. Kotzanikolaou P., Theoharidou M., Gritzalis D., "Accessing n-order dependencies between critical infrastructures", International Journal of Critical Infrastructures, Vol. 9, Nos. 1-2, pp. 93-110, 2013.
12. Kotzanikolaou P., Theoharidou M., Gritzalis D., “Cascading effects of common-cause failures on Critical Infrastructures, in Proc. of the 7th IFIP International Conference on Critical Infrastructure Protection (CIP-2013), pp. 171-182, Springer (AICT 417), USA, March 2013
13. Lekkas D., Gritzalis D., "e-Passports as a means towards the first world-wide Public Key Infrastructure", in Proc. of the 4th European PKI Workshop (EuroPKI '07), J. Lopez, P. Samarati (Eds.), pp. 34-48, Springer (LNCS 4582), Spain, June 2007.
14. Lekkas D., Gritzalis D., "Long-term verifiability of healthcare records authenticity", in Proc. of the 9th IMIA Working Conference on Data Protection, Roger-France F., et al. (Eds.), IOS Press, France, April 2006.
15. Mylonas A., Dritsas S., Tsoumas B., Gritzalis D., “On the feasibility of malware attacks in smartphone platforms”, in E-Business and Telecommunications, pp. 217-232, Lecture Notes (CCIS 314), Springer, 2012.
16. Mylonas A., Gritzalis D., Tsoumas B., Apostolopoulos T., “A qualitative metrics vector for the awareness of smartphone security users”, in Proc. of the 10th International Conference on Trust, Privacy & Security in Digital Business (TRUSTBUS-2013), pp. 173-184, Springer (LNCS 8058) Czech Republic, August 2013.

17. Mylonas A., Kastania A., Gritzalis D., “Delegate the smartphone user? Security awareness in smartphone platforms”, Computers & Security, Vol. 34, pp. 47-66, May 2013.

18. Mylonas A., Tsalis N., Gritzalis D., "Evaluating the manageability of web browsers controls", in Proc. of the 9th International Workshop on Security and Trust Management (STM-2013), pp. 82-98, Springer (LNCS 8203), United Kingdom, September 2013.

19. Stavrou V., Kandias M., Karoulas G., Gritzalis D., "Business Process Modeling for Insider threat monitoring and handling", in Proc. of the 11th International Conference on Trust, Privacy & Security in Digital Business (TRUSTBUS-2014), pp. 119-131, Springer (LNCS 8647), Germany, September 2014.

20. Stergiopoulos G., Theoharidou M., Gritzalis D., "Using logical error detection in Remote-Terminal Units to predict initiating events of Critical Infrastructures failures", in Proc. of the 3rd International Conference on Human Aspects of Information Security, Privacy and Trust (HCI-2015), Springer, USA, August 2015.

21. Theoharidou M., Kotzanikolaou P., Gritzalis D., “Risk assessment methodology for interdependent critical infrastructures”, International Journal of Risk Assessment and Management (Special Issue on Risk Analysis of Critical Infrastructures), Vol. 15, Nos. 2/3, pp. 128-148, 2011.

22. Theoharidou M., Papanikolaou N., Pearson S., Gritzalis D., “Privacy risks, security and accountability in the Cloud”, in Proc. of the 5th IEEE Conference on Cloud Computing Technology and Science (CloudCom-2013), pp.177-184, IEEE Press, United Kingdom, December 2013.

23. Theoharidou M., Tsalis N., Gritzalis D., "In Cloud we Trust: Risk-Assessment-as-a-Service", in Proc. of the 7th IFIP International Conference on Trust Management (IFIP TM-2013), pp. 100-110, Springer (AICT 401), Spain, June 2013.

24. Virvilis N., Dritsas S., Gritzalis D., “A cloud provider-agnostic secure storage protocol”, in Proc. of the 5th International Conference on Critical Information Infrastructure Security (CRITIS-2010), Wolthusen S., et al. (Eds.), pp. 104-115, LNCS-6712, Springer, Greece, September 2010.

25. Virvilis N., Dritsas S., Gritzalis D., “Secure Cloud Storage: Available Infrastructure and Architecture Review and Evaluation”, in Proc. of the 8th International Conference on Trust, Privacy & Security in Digital Business (TRUSTBUS-2011), Furnell S., et al. (Eds.), pp 74-85, LNCS-6863, Springer, France, August 2011.

26. Virvilis N., Gritzalis D., “The Big Four - What we did wrong in Advanced Persistent Threat detection?”, in Proc. of the 8th International Conference on Availability, Reliability and Security (ARES-2013), pp. 248-254, IEEE, Germany, September 2013.

27. Virvilis N., Gritzalis D., “Trusted Computing vs. Advanced Persistent Threats: Can a defender win this game?”, in Proc. of 10th IEEE International Conference on Autonomic and Trusted Computing (ATC-2013), pp. 396-403, IEEE Press, Italy, December 2013.

28. Virvilis N., Tsalis N., Mylonas A., Gritzalis D., "Mobile devices: A phisher's paradise", in Proc. of the 11th International Conference on Security and Cryptography (SECRYPT-2014), pp. 79-87, ScitePress, Austria, August 2014.

29. N. Virvilis, O. Serrano, "Changing the game: The art of deceiving sophisticated attackers", in Proc. of the 6th International Conference on Cyber Conflict (CYCON-014), pp. 87-97, IEEE, Estonia, June 2014.

30. Virvilis N., Serrano O., Dandurand L., "Big Data analytics for sophisticated attack detection", ISACA Journal, Vol. 3, pp. 22-25, 2014

31. Virvilis N., Tsalis N., Mylonas A., Gritzalis D., “Security Busters: Web browser security vs. suspicious sites”, Computers & Security, Vol. 52, pp. 90-105, July 2015