International Journal of Research
Available at http://internationaljournalofresearch.org

p-ISSN: 2348-6848
e-ISSN: 2348-795X

Volume 02 Issue 04
April 2015

Open Source Threat Intelligence System

Sabari Girish Nair
MBA-ITBM Student, SCIT, Pune, India
sabari.nair@associates.scit.edu

Dr. Priti Puri
Assistant Professor, SCIT, Pune, India
priti@scit.edu

Abstract— With more and more business switching to e-business model, cyber security has become a top priority. With the rise of more sophisticated attacks and lack of capabilities to detect, respond and prevent such attacks, the need for next level cyber security solutions has raised. To counter these everyday cyber-attacks, the organizations now need a much dynamic solution; a solution which is smart and intelligent just likes the security threats. So, there is a great emphasis on threat intelligence solutions. The Information Security market is buzzing about threat intelligence with many vendors coming up with new solutions. Often these solutions are very costly and are generally out of the reach of small firms and startups. This paper tries to overcome this obstacle by proposing a more cost effective way to achieve threat intelligence solutions. So, the main objective of this paper is to propose a way to leverage the benefits of threat intelligence solutions without the huge costs involved.

Keywords-Threat Intelligence, Open Source Security, Analytics and Security

I. INTRODUCTION

Threat intelligence is evidence-based knowledge, including context, mechanisms, indicators, implications and actionable advice, about an existing or emerging menace or hazard to assets that can be used to inform decisions regarding the subject's response to that menace or hazard. We know that the major component in threat intelligence solution is its analytic engine. [2] This component is the one which process all the information, correlates them with attack vectors and derives intelligence. It is the part for which big security companies charge exorbitant prices as better the analytics engine better would be the derived intelligence. So the solution must focus on providing this analytic engine in a cost effective manner to the small firms.

II. PROBLEM UNDER STUDY

Based on the above problem statement the major thing to remember here is that we are looking ways to provide small firms with threat intelligence solutions. So we need to consider that finance will be a huge pressure point. This is why the small firms can't afford complex and elaborate threat intelligence software which is available in the market.

Now a complete threat intelligence solution will have a complex structure, which further translates into high costs. So our solution needs to be one which is simple and easy to build and implement.

III. PROPOSED SOLUTION

As stated above that the heart of threat intelligence solution is its analytic engine. So to make a cost effective solution we need to have an analytic engine which is cost effective and efficient. So we propose to use open source analytic tools for this purpose. Based on this, we have proposed the following architecture of the threat intelligence solution.

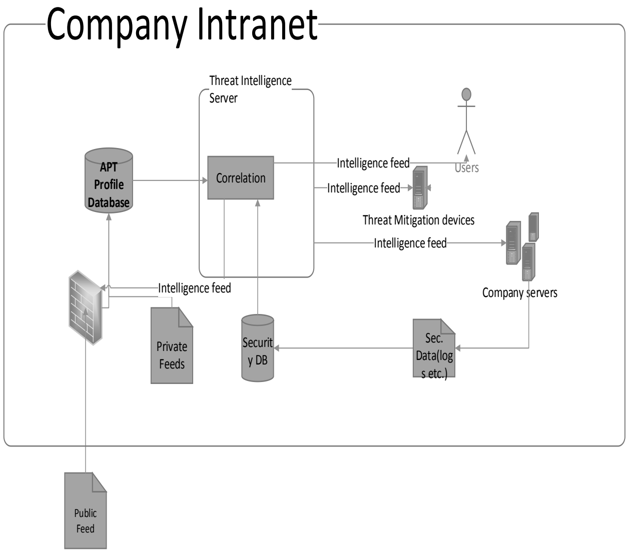

Figure 1: Proposed Threat Intelligence Architecture

The detailed description of each part of the solution is as follows:

A. *Feed Data*

This is the data which is used to derive the intelligence. It can be further classified as public and private feeds.

Public feeds are the threat profile data which we can obtain from other companies, government organizations etc. These contain details such as

- Malware signatures
- Infected URL lists
- Potential attack IPs

Based on this and security logs, the analytic engine will process and derive the intelligence on various threats or threat vectors.

Some of the sources for such data are:

- Feeds from CERTs
- Feeds from other security agencies like MacAfee, Norton, Atlas, Red sky alliance etc.

Private Feeds

These are the APT profile data which is maintained by the companies themselves based of their experience. It also includes previous intelligence feeds from the threat intelligence server which have been stored.

If the organization is new then this would not be their initially. Slowly the company can build their own database of various threat vectors through continuous intelligence feeds and other logs. Data from SIEM or any other monitoring systems could also be qualified as private feeds.

B. *APT Profile database*

The feeds about the APT profiles need to be stored somewhere. For this, we suggest to create a database for the same. This database would be normalized and would be filled by transforming the public and private feeds.

It is essential that all the feeds be stored in a consistent format so that the analytic engine could easily analyze the data. Use of RDBMS with normalized tables also improves the efficiency of the threat intelligence solution.

We propose to use open source DBMS software for the solution.

C. *Security Database*

This database will contain all the network logs, application logs, and system logs which are generated by all the systems in the firm. This is the actual data which is analyzed by correlating with the APT profile data to derive the intelligence.

Similar to the APT Profile database this too needs to be normalized and indexed so as to improve the efficiency of the threat intelligence solution.

D. *Threat Intelligence Server- Correlation/Analytic Engine*

This part of the proposed solution is the core of any threat intelligence solution. It basically consists of an analytic engine which takes raw data from multiple sources, processes it and gives the intelligence feeds based on it.

For this part, we propose to use open source analytic tools.

Some of the tools which could be used in our solution are:

- SpagoBI [3]

  SpagoBI is an Open Source Business Intelligence suite, belonging to the free/open source. Spago World initiative, founded and supported by Engineering Group. It offers a large range of analytical functions, a highly functional semantic layer often absent in other open source platforms and projects, and a respectable set of advanced data visualization features including geospatial analytics. SpagoBI is released under the Mozilla Public License, allowing its commercial use. SpagoBI is hosted on OW2 Forge managed byOW2 Consortium, an independent open-source software community.

- Openi [4]

  Open I is a web-based OLAP reporting application. OpenI is an out-of-box solution for building and publishing reports from XMLA-compliant OLAP data sources. The final release of OpenI as a standalone platform is 2.0-RC2.

- Rapidminer [5]

  RapidMiner is a software platform developed by the company of the same name that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics. It is used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process including results visualization, validation and optimization. RapidMiner is developed on a business source model which means the core and earlier versions of the software are available under an OSI-certified open source license on Sourceforge.

- Mondrian [6]
  Online Analytical Processing server (OLAP). Allows business users to analyze large and complex amounts of data in real-time. Its Online Analytical Processing server (OLAP) is written in JAVA which is why the system responds to queries fast enough to allow an interactive exploration of the data - even if they have millions of records, occupying several gigabytes. It brings multidimensional analysis to the masses, allowing users to examine business data by drilling and cross-tabulating information.

E. *Intelligence feeds*

This is the final product through which the firms could improve their information security. Based on these feeds, the firms could modify the rules on their mitigation devices like firewall, IDS, IPS etc. thereby using the feeds to get a proactive security defense.

IV. Proof Of Concept

We tried to implement the proposed solution. For that purpose, Software used

- MySql- Database for the log data
- SpagoBI Server- Analysis
- SpagoBI Studio- Creating Business models for analysis

All three are open source software.

A. *Scenario*

We tested our solution for two different scenarios as given below:

1. We assumed that there are nine machines with different hostnames in a company. In these machines, logging is enabled such that if an application is executed then the log saves the name of the application, host name, date and ip address of the system. The company is closed on weekends so ideally there should be no applications running at weekends.

2. We monitored the network connections for these nine machines. The company is closed on weekends so ideally there should be no network connections during weekends.

B. *Findings*

Scenario 1:

Based on the analysis of the log data, it was found that there were two applications which were running all seven days. This shows an anomaly which means it could be malwares.

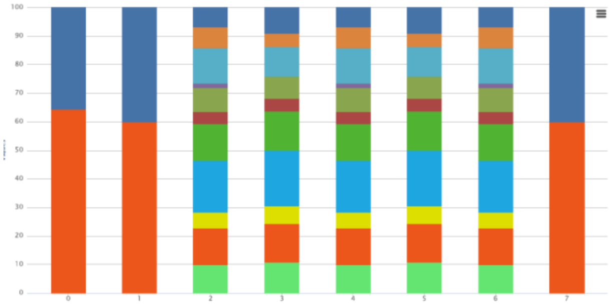

Figure 2: Scenario 1 Findings

The screen shows that 2 apps namely CS (deep blue) and WOT (orange) were executed in all seven days which probably means that they are malwares. Based on this intelligence, we can set the firewall rules to block the execution.

Scenario 2:

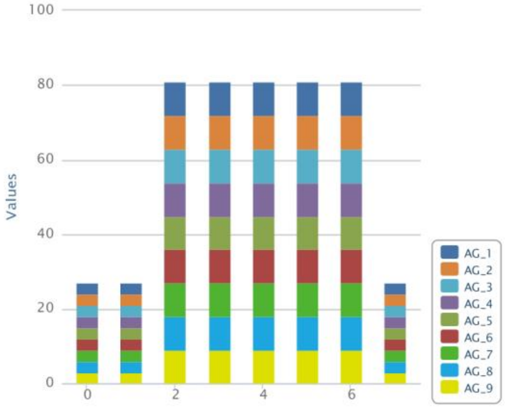

Figure 3:Scenario 2 Findings

The above analysis shows that on weekends (column 0, 1,7) there are network connections from all the hosts for at least 3 hours. As the company is closed on weekends this behavior is suspicious. It is safe to assume that there is some kind malware which is generating these connections.

To summarize both the above findings, open source solution could be used to derive intelligence and can be a viable option for threat intelligence solutions.

V. Pros and Cons of the Proposed Solution

A. *Pros [7]*

1. It's almost free because the company may need to pay only for subscription based APT profile feeds. Apart from this, as we are using open source software sans databases, it is free.

2. Developers can add to it or modify it due to open source analytics, which means better quality, more secure and less prone to bugs than proprietary systems

3. Using open source software, we are not locked in to a particular vendor's system which only works with their systems. This means that our solution can be integrated even with legacy systems.

4. Business can modify and adapt open source software according to their own requirements, something that is not possible with proprietary systems.

B. *Cons [7]*

The proposed solution will not be as effective as proprietary threat intelligence products due to open source analytic tools to derive intelligence.

This solution will be semi-automatic which means a dedicated team must be there to continuously derive intelligence and work on it by updating the rules on mitigation devices. To make it completely automatic the firm may need to build custom interfaces.

Although having an open system means that many people can identify bugs and fix them, it also means that malicious users can potentially view it and exploit any vulnerabilities.

Conclusion

The main motivation behind this research work was the need to protect the IT landscape from ever evolving cyber-attacks in a proactive manner. The research was focused to provide threat intelligence capabilities to small firms by leveraging the open source community. The findings of the research showed promising results which shows that open source analytic software can be utilized for threat intelligence.

VI. References

[1] R. Holland, "Five steps to build effective threat intelligence capability," *Forrester,* 2013.

[2] "CIF," [Online]. Available: https://code.google.com/p/collective-intelligence-framework/.

[3] "Spagobi," [Online]. Available: http://en.wikipedia.org/wiki/SpagoBI.

[4] "OpenI," [Online]. Available: http://wiki.openi.org/.

[5] "RapidMiner," [Online]. Available: http://en.wikipedia.org/wiki/RapidMiner.

[6] "Mondrian," [Online]. Available: http://community.pentaho.com/projects/mondrian/.

[7] "Open source advantages," [Online]. Available: http://www.entrepreneurhandbook.co.uk/open-source-software/.

[8] "Red Sky Alliances," [Online]. Available: http://redskyalliance.org/?page_id=35.

[9] "Pentaho Comunity Edition," [Online]. Available: http://wiki.pentaho.com/display/COM/Community+Edition+Downloads.

[10] "Open Source Analytic tools," [Online]. Available: http://www.predictiveanalyticstoday.com/open-source-free-business-intelligence-solutions/.

[11] K. Grutzmacher, "CISCO OpenSOC Hadoop Design," 2014.

[12] D. o. FireEye. [Online]. Available: http://www.allaboutgovernance.com/uncategorized/fireeye-part-4-drawbacks-of-fireeye-current.

[13] "Fireeye Advantages," [Online]. Available: ca.westcon.com/documents/49972/fireeyebattlecard_final.pdf.

[14] C. D. R. L. Greg Rattray, "Building an Effective Cyber Threat Intelligence Practice".

[15] "Atlas feeds," [Online]. Available: http://www.arbornetworks.com/resources/research/atlas-aif-feed.

[16] i. Partners, "What is cyber threat intelligence".