Student Paper of the Year

Predicting Cyber-Attacks Using Publicly Available Data

George Onoh
Bowie State University
Bowie, Maryland

Abstract - Cyber-attacks are often detected too late. According to reports on reported cyber-attack incidents, most victim organizations do not know that their systems have been breached until they are informed by organizations or individuals external to the victim organization's physical or logical network. This is a significant problem for cyber security professionals and organizations. To further understand this problem, I investigated the following questions in this study: How are external organizations able to detect cyber-attack incidents using only publicly available information? How can cyber-attacks be predicted based on only publicly available data? I collected data on indices representing mentions of a certain type of attack (brute-force/ password guessing attack) from public data repositories as well as ground truth data for a target organization. I extracted and stored the data daily. I used the collected data as training data in a machine learning algorithm. After limited training, the system was able to predict future attacks. The results suggest that it is possible to predict cyber-attacks based on publicly available data.

Keywords

Network security, cyberattack prediction, social network, Bayesian networks, probabilistic warning, open source intelligence, cyber threat intelligence, cognitive computing

1 PREDICTING CYBER-ATTACKS USING PUBLICLY AVAILABLE DATA

Cybersecurity breaches happen in phases. Typically, these breaches are only detected at their later phases or after they have been completed. Being able to detect early stages of an attack will provide organizations significant advantage in their effort to secure the organization's information. Most reported cyber-attacks are detected by parties external to the victim organizations. Automated detection of cyber-attacks using probabilistic warning systems, that fuse both Internal and External Sensors, is a subject of on-going research efforts funded by various governments such as the Chinese government (Wu, Yin, & Guo, 2012) and the United States government (Okutan, Yang, & McConky, 2017). Such systems will draw from multiple data sources including open source data such as Twitter, analyze the data and predict attacks, thereby giving organizations some lead-time to take actions to protect themselves adequately.

1.1 Statement of the Problem

Detection of cyber-attacks often happen too late for most victims. According to reports on cyber-attack incidents, most victim organizations do not know that they have been hacked until they are informed by organizations or individuals external to the victim organization's corporate network. This is a significant problem for cyber security professionals and organizations.

1.2 Research Questions

For this study, the following questions were investigated:

- How are external organizations able to detect cyber-attack incidents using only publicly available information?
- How can cyber-attacks be detected using readily available hardware and software as well as publicly available data?

1.3 Purpose of the Study

The purpose of this study is to investigate current approaches to cyber-attack detection, as well as prediction, based on publicly available data.

2 BACKGROUND

Analyzing openly available data can yield useful and actionable information that can be used to predict cyber-attacks. Many studies showing promising results in this area are based on theories of Bayesian network. This is also known as Bayes network, belief network, Bayes(ian) model or probabilistic directed acyclic graphical model. It is a probabilistic graphical model (a type of statistical model) that represents a set of random variables and their conditional dependencies via a directed acyclic graph (DAG).

2.1 Scope

The open source signals used in this study are generated from crowdsourced data. Not every cyber-attack is reported to the public and therefore this study could not have examined data on all cyber-attacks. Also, the available data was filtered based on key phrases such as "brute-force attack", "password guessing" and other combinations of the words and related words. The filter criteria can be further refined for improved accuracy. Also, only data presented in English language were collected. There is a huge amount of data in other languages which were not used in this study. No doubt, accuracy can be improved by including data in more languages. Also, data was collected for only thirty days which is a limited sample size. Increasing the sample size should improve the accuracy. Finally, only one form of cyber-attack was considered in this study – "brute-force password guessing attack".

3 LITERATURE REVIEW

Cybersecurity breaches happen in phases. Typically, these breaches are only detected at their later phases or after they have been completed. Being able to detect

early stages of an attack will provide organizations significant advantage in their effort to secure the organization's information. Most reported data breaches are detected by parties external to the victim organizations. Automated detection of cyber-attacks using approaches such as probabilistic warning systems, that use both Internal and External Sensors, is a subject of on-going research efforts funded by various governments such as the Chinese government (Wu, Yin, & Guo, 2012) and the United States government (Okutan, Yang, & McConky, 2017). Such systems will draw from multiple data sources including open source data, such as Twitter, and analyze the data and predict attacks, thereby giving organization some lead-time to take actions to protect itself adequately.

Detection of cyber-attacks often happen too late for victims. According to reports on cyber-attack incidents, most victim organizations do not know that they have been hacked until they are informed by organizations or individuals external to the victim organization's physical or logical network. This is a significant problem for cyber security professionals and organizations.

There is therefore a growing interest in finding ways to detect cyber-attacks at an early phase of the attack or, better still, before the attack happens. It is therefore a subject of growing research interest to governments and organizations.

In pursuance of such a solution, on July 17, 2015 Intelligence Advanced Research Projects Activity (IARPA), a United States government agency published a Broad Agency Announcement (BAA) calling for teams to participate in its research program titled Cyber-Attack Unconventional Sensor Environment (CAUSE). In the BAA, the program indicated that it was seeking to research on multidisciplinary methods for accurate and timely forecast of cyber-attacks. The program started in February 2016 and is expected to continue until August 2019 (IARPA, 2015).

In 2013 Axelrad, E. T., Sticha, P. J., Brdiczka, O., & Shen, J. in their work titled "A Bayesian network model for predicting insider threats" proposed a model for predicting insider threat by analyzing psychological data on the individuals. Some of the variables they measured include Agreeableness, Neuroticism,

Conscientiousness, Excitement seeking, Perceived stress, Hostility and Job Satisfaction. They concluded that these factors could be used to predict counterproductive behavior. Also, their study provided a concise discussion of Bayesian networks, a key concept used in this study.

In another study titled "A Bayesian Method for the Induction of Probabilistic Networks from Data" authored by Cooper, G. F., & Herskovits, E., the authors provided mathematical basis for an algorithm for constructing a belief network or Bayesian network from a database. Their study rigorously demonstrated mathematical basis for the principles behind the algorithm used in this study.

In 2011, Lee, K., Palsetia, D., Narayanan, R., Patwary, M., Agrawal, A., & Choudhary, A. authored a study titled "Twitter trending topic classification". In the study, they proposed a model for use in classifying Twitter trending topics using Text-based and Network-based Classifications. A similar approach to theirs was used in analyzing text-based data and extracting critical indices used in this study.

Also, in a study funded by the National Natural Science Foundation of China and published in 2012, Wu, Yin & Guo proposed a model for predicting cyberattacks based on Bayesian Network. They used Attack Graphs to model the threat scenario while accounting for environmental factors of the network. This approach enabled them to more effectively represent the environmental factor in their model than previous studies. Their model considered factors such as the vulnerabilities in the network, the value of the assets in the network, the usage condition of the network (for example traffic) and the attack history of the network. Using their model, they were able to predict attack probability with a higher accuracy than previous models (Wu, Yin, & Guo, 2012). One problem with this approach, however, is that it is static and seems to ignore the dynamic nature of the threat. The cyber threat landscape is constantly and rapidly changing, and they did not use any mechanism to adapt to those changes as they happen. This means that the accuracy of the probability of attack prediction made using the model will deteriorate as the network's threat landscape continues to change over time.

However, their multi-factor approach was adopted in this study and adapted for real-time use.

In a study presented at the International Workshop on Biometrics and Forensics in 2017, Hernández, A., Sanchez, V., Sánchez, G., Pérez, H., Olivares, J., Toscano, K., Nakano, M., & Martinez, V. proposed a sentiment analysis method for Twitter data. They were able to use this method to predict actual security events by analyzing sentiments of tweets on Twitter. They used Twitter API Stream to collect live tweets from Twitter. They also used tools like SentiWordNet compendium to enable them to synthesize the sentiments in the tweets. They also used a Support Vector Machine (SVM) classifier to predict events after training it with data they have collected. (Hernández, et al., 2016). This study is relevant to the current study because it makes use of social media data to predict security events such as hacktivists attacks. It however could not be used to provide actionable early warning to cybersecurity professionals tasked with protecting an organizations data.

Another study published in February 2017 by Khandpur, et al proposed a framework for security event detection from social media data. They opined that social networks can be used as crowdsourced sensors for detecting cyberattacks. This approach is important because it gives the organization the capability to reach beyond its walled garden to have a sense of what is going on outside its boundaries and beyond their firewalls. They also demonstrated a weakly supervised approach which requires no training phase. This in my opinion is profound because it eliminates the need for collecting large volumes of data to train the system. This also means that systems built with this approach can be used immediately without requiring lengthy system training. The main contributions of their study are: they proposed a framework for detecting cyberattack from social media, they also proposed a query expansion strategy that is based on dependency tree patterns, and they also performed empirical evaluation of three kinds of cyberattacks (which are the Distributed Denial of Service attack, data breaches and account hijacking). Their study achieved high accuracy in retrieving cyberattack related content from social media and using the content to detect security related events (Khandpur, et al.,

2017). The result of the study is significant to this study because it gives organizations a sense of what is going on (in terms of cyberattacks) beyond the corporate network's boundaries in real-time. One weakness of this approach is that it just tells us what is currently happening. It does not warn us of impending attacks. Also, while it is nice to know what is going on around the world regarding cyber-threats, I am sure most security professionals will be more interested in knowing about cyber-events targeted at the organizations they are charged with protecting.

Also, another study by Okutan, A., Yang, S. J., and McConky, K. published in 2017 went a step further. Their model was able to predict cyber-attacks before they happened, giving cybersecurity professionals time to take steps to avoid such attacks or minimize the potential impact of the attack. They used a Bayesian classifier and non-conventional signals. These signals are available to the public and are from sources such as Twitter, Global Database of Events, Language, and Tone (GEDLT) and Hackmageddon. In the study they used signals obtained from open sources of data to calculate measurable indices such as Twitter Attack Mentions (TAM), GDELT Event Mentions (GEM), GDELT Event Tone (GET) and Hackmageddon Number of Attacks (HNA). These indices are calculated for the potential target daily. Also ground truth of the potential target (information about actual attacks) is also recorded. All the data is fed into a Bayesian network as training data, daily for about five months. After the training period, the system was able to predict future attacks with high levels of accuracy. This result is significant to my current study because it not only enables organizations to detect cyberattacks going on outside the boundaries of the organization's network, it also gives the cybersecurity professionals an early warning before their network is attacked. This then raises the question; how can we apply this in real-life?

One of the aims of this study is to test the replicability of the study published by Okutan, A., Yang, S. J., and McConky, K. I took a similar approach, using readily available software, hardware and data. I used anonymized data from the website of an organization for the training data. For easier reference, I will call the organization JEWEL.

The findings of the studies reviewed here, suggest that the following assumption can be made:

On a given day i and for a particular attack type a (say password guessing brute-force attacks) the following relationship exists:

$$g^{ia} = f(t^{ia}, m^{ia}, e^{ia}, h^{ia}, n^{ia})$$

Where:

g^{ia} = Ground Truth value for day i

t^{ia} = Twitter Attack Mentions (TAM) value for attack type a on day i

m^{ia} = Facebook Event Mentions (FEM) value for attack type a on day i

e^{ia} = Instagram Event Mention (IEM) value for attack type a on day i

h^{ia} = Blogs Event Mention (BEM) value for attack type a on day i

n^{ia} = Forums Event Mention (FoEM) value for attack type a on day i

4 RESEARCH DESIGN

Businesses, governments, organizations and individuals are increasingly taking advantage of the rapid advancements in information technology. Our homes, offices and appliances are getting smarter and the world is getting more connected. As a result, more and more organizations, governments and individuals have come to

depend heavily on information systems. These information systems now hold critical information that must be protected.

As a result, cybersecurity is an increasingly important subject. This is because, as the importance and value of information systems and the information they hold increase, so does the associated risks posed by cyber-attacks.

However, detection of cyber-attacks often happens too late for victims. According to reports on cyber-attack incidents, most victim organizations do not know that they have been hacked until they are informed by organizations or individuals external to the victim organization's physical or logical network. This is a significant problem for cyber security professionals and organizations.

Recent studies in the area of cybersecurity have shown that it is possible to detect cyber-attacks based on publicly available data. Studies have also shown that it is possible to predict future cyber-attacks by analyzing publicly available data. This study investigates current approaches to cyber-attack detection using probabilistic warning systems based on publicly available data.

For the purpose of this study, the following questions were investigated:

- How are external organizations able to detect cyber-attack incidents using only publicly available information?
- How can cyber-attacks be predicted using only publicly available data?
- How can cyber-attacks be detected using readily available hardware and software as well as publicly available data?

Data was collected using brand24.com. It provides tools to enable measuring and monitoring social media sentiments and chatter.

4.1 Sampling

The unit of analysis is a day. I collected TAM, FEM, IEM, BEM, FoEM and Ground Truth for thirty days from the respective data sources described below.

Ground Truth is Yes (if there was an attack) or No (if there was no attack) for each day.

4.2 Data Collection

4.2.1 Limitations and Trustworthiness

The open source signals used in this study are generated from crowdsourced data. Not every cyber-attack is reported to the public and therefore this study could not have examined data on all cyber-attacks. Also, the available data was filtered based on key phrases such as "brute-force attack", "password guessing" and other combinations of the words and related words. The filter criteria can be further refined for improved accuracy. Also, only data presented in English language were collected. There is a huge amount of data in other languages which were not used in this study. No doubt, accuracy can be improved by including data in more languages. Also, data was collected for only thirty days which is a limited sample size. Increasing the sample size will improve the accuracy. Finally, only one form of cyber-attack was considered in this study – "brute-force password guessing attack".

4.2.2 Validity and Reliability

To ensure reliability, automated tools were used to query the data repositories for the raw data. Also, to further ensure validity of results, ten-fold cross-validation was used in analysis.

4.2.3 Open Source Signals and Ground Truth

The publicly available data used are accessible to anyone without being in direct contact with the organization. The signals (variables) used, are further explained below:

- Twitter Attack Mentions (TAM): That is the number of raw tweets that contain the attack type keyword. This was determined every day for the attack type "password guessing / brute-force attack".

- Facebook Event Mentions (FEM): This is the number of public Facebook posts that contain the attack type keywords. This was determined every day for the attack type "password guessing / brute-force attack".
- Instagram Event Mention (IEM): This is the number of public Instagram posts that contain the attack type keywords. This was determined every day for the attack type "password guessing / brute-force attack".
- Blogs Event Mention (BEM): This is the number of public blog posts that contain the attack type keywords. This was determined every day for the attack type "password guessing / brute-force attack".
- Forums Event Mention (FoEM): This is the number of public discussion forums posts that contain the attack type keywords. This was determined every day for the attack type "password guessing / brute-force attack".
- Ground Truth: Data for the ground truth is automatically collected and logged by the Intrusion Detection System protecting the web server. Only data concerning events referring to "password guessing brute-force" attacks were extracted for the date range of interest. For each date, if there were any "password guessing brute-force" attacks that day then the Ground Truth is recorded as "Y" meaning Yes. If there was no such attack, ground Truth for the day is recorded as "N" for No.

On a daily basis, data for each of the four external signals were stored and loaded into a Bayesian classifier along with the ground truth (Yes or No) input provided by the organization as training data for the Bayesian classifier.

4.3 Analytical Method

The data was imported into Weka for analysis. Weka stands for Waikato Environment for Knowledge Analysis. It is a suite of machine learning software written in Java, developed at the University of Waikato, New Zealand. It is a free software tool. Weka comes with a suite of tools and machine learning algorithms, but the one used is the Naïve Bayes Classifier. This was used to build and test the model. I also used ten-fold cross-validation to ensure accuracy of the analysis. The results are discussed below.

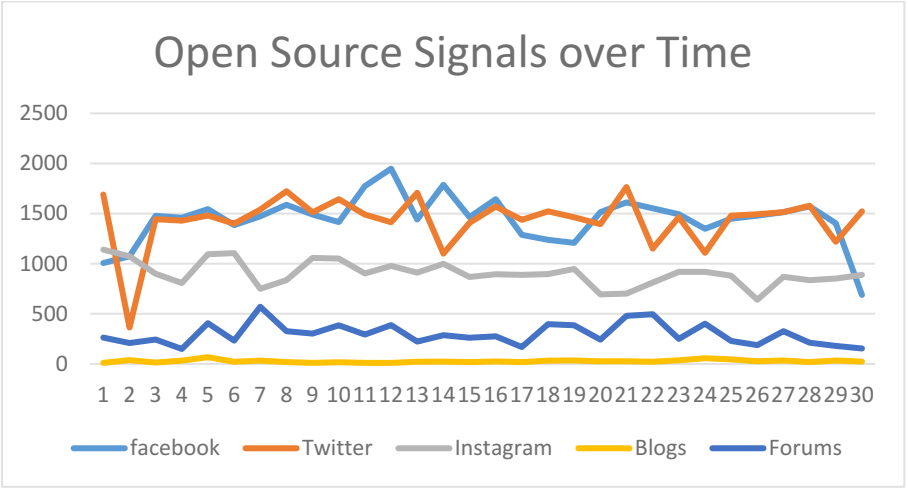

Figure 1: Showing trend in open source signal over time

Figure 1 shows a graphical representation of the open source signals over time measured in days. Data analysis was conducted using Weka and the test mode used in the model is ten-fold cross-validation to ensure accuracy. The results are summarized as follows:

Correctly Classified Instances	19	63.3333 %
Incorrectly Classified Instances	11	36.6667 %
Total Number of Instances	30	

Table 1
Breakdown of full training set

Attribute	Yes	No
	0.69	0.31
Facebook		
mean	1396.55	1564.14
std. dev.	238.461	183.194
weight sum	21	9
precision	43.4483	43.4483
Twitter		
mean	1421.07	1463.2
std. dev.	295.41	128.804
weight sum	21	9
precision	50.0714	50.0714
Instagram		
mean	890.653	938.272

Table 1
Breakdown of full training set

Attribute	Yes	No
std. dev.	131.308	85.9777
weight sum	21	9
precision	18.5185	18.5185
Blogs		
mean	29.5556	20.7593
std. dev.	13.728	9.4543
weight sum	21	9
precision	3.1667	3.1667
Forums		
mean	308.882	277.202
std. dev.	118.713	57.3039
weight sum	21	9
precision	15.5926	15.5926

Table 2
Detailed Accuracy by Class for the model

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.714	0.556	0.75	0.714	0.732	0.154	0.714	0.853	Y
	0.444	0.286	0.4	0.444	0.421	0.154	0.714	0.586	N
Weighted Avg.	0.633	0.475	0.645	0.633	0.639	0.154	0.714	0.773	

The model classified 63.333% of the instances correctly. Also, F-Measure for predicting an attack is 0.732.

5 CONCLUSIONS

I set out to find answers to the following questions: How are external organizations able to detect cyber-attack incidents using only publicly available information? How can cyber-attacks be predicted based on only publicly available data? The following conclusions can be drawn:

- This study has demonstrated how an external organization can detect as well as predict cyber-attack incidents using only publicly available information.
- An accuracy of 63.333% and an MCC (Matthews Correlation Coefficient) of 0.154 implies a lack of randomness, especially considering the very small dataset. This is significant because, there is a lot of room for improvement in the model used. For example, the accuracy is likely to increase if more training data were collected. Also modifying the model to account for

environmental factors peculiar to the target organization as well as particular assets should also further increase the accuracy as demonstrated by Wu et al. (Wu, Yin, & Guo, 2012)

- There is a relationship between the variables which can be summarized as follows:

  On each day i for a particular attack type a:

  $g^{ia} = f(t^{ia}, m^{ia}, e^{ia}, h^{ia}, n^{ia})$

  Where:

  g^{ia} = Ground Truth value for day i

  t^{ia} = Twitter Attack Mentions (TAM) value for attack type a on day i

  m^{ia} = Facebook Event Mentions (FEM) value for attack type a on day i

  e^{ia} = Instagram Event Mention (IEM) value for attack type a on day i

  h^{ia} = Blogs Event Mention (BEM) value for attack type a on day i

  n^{ia} = Forums Event Mention (FoEM) value for attack type a on day i

6 FUTURE WORK

In the future, I will like to use other external variables that include elements that are unique to the target organization – such as a measure of sentiments towards the organization, as well as a daily assessment of the organization's information security

risk posture. It will be interesting to see how this affects the accuracy of the prediction.

REFERENCES

[1] Cybersecurity Ventures. (2017). 2017 Cybercrime Report. Retrieved from https://cybersecurityventures.com/2015-wp/wp-content/uploads/2017/10/2017-Cybercrime-Report.pdf

[2] Hernández, A., Sanchez, V., Sanchez, G., Pérez, H., Olivares, J., Toscano, K., . . . Martinez, V. (2016). Security attack prediction based on user sentiment analysis of Twitter data. Industrial Technology (ICIT), 2016 IEEE International Conference on, pp. 610-617.

[3] IARPA. (2015, July 17). Cyber-attack Automated Unconventional Sensor Environment (CAUSE). Retrieved from https://www.iarpa.gov: https://www.iarpa.gov/index.php/research-programs/cause

[4] Khandpur, R. P., Ji, T., Jan, S., Wang, G., Lu, C.-T., & Ramakrishnan, N. (2017). Crowdsourcing Cybersecurity: Cyber Attack Detection using Social Media. Cryptography and Security (cs.CR). ACM.

[5] Okutan, A., Yang, S., & McConky, K. (2017). Predicting cyber attacks with bayesian networks using unconventional signals. Proceedings of the 12th Annual Conference on Cyber and Information Security Research (CISRC '17). New York, NY, USA: ACM.

[6] Verizon. (2014). 2014 Data Breach Investigation report. Retrieved from www.verizonenterprise.com: http://www.verizonenterprise.com/resources/reports/rp_Verizon-DBIR-2014_en_xg.pdf

[7] Wu, J., Yin, L., & Guo, Y. (2012). Cyber Attacks Prediction Model Based on Bayesian Network. Parallel and Distributed Systems (ICPADS), 2012 IEEE 18th International Conference on. Singapore, Singapore: IEEE.