Insider Threats in Verizon

LaChandra Ash

11.7.2021

The greatest challenge is to figure out who the insider threats are, why they are a threat to businesses and other domains, how can business insider threats remain undetectable for long periods of time, what are the common threats presented by a insider threat, how are insider threats finally detected, what methods were used to detect them, unknowingly evidence left behind in physical and information systems within domains. Detecting and preventing insider threats are the two major challenges in businesses. The teams that work in information systems find it challenging to detect and reveal insider threats, because there is a lack of data and other evidence to prove who the insider threat is.

Verizon Data Breaches in 2019

An insider threat is a security risk that was created within the victim organization. The attacker may be someone in senior management, external or internal contractor, company officer, and employees. Approximately 34% of the data breaches that occurred atVerizon in 2019 involved internal threat actors [12]. There were 17% data files that allowed anyone to access them.

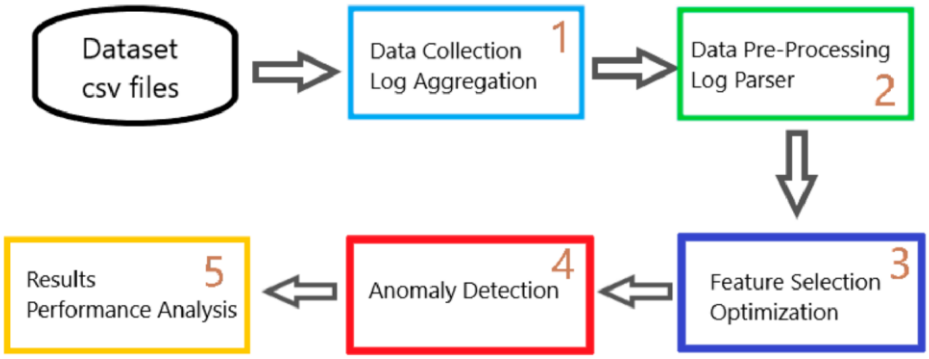
Hypothesis Questions:

- Can an algorithm be created to prevent the attacks?
- Why did the threats occur?
- How long did the threats occur at Verizon before they became detectable?
- How can data science play a role in reducing insider threats, reducing impacts of it, protect the business reputation and its assets.
- How can machine learning be used to accurately detect the insider threats and provide businesses with the information so they can prevent or reduce the number of attacks.
- How can an insider threat impact Verizon, management, employees, assets, reputation, customers, and stakeholders?
- Which data science tools can be used to detect the insider threat?
- How can the tools and outcomes help business to decide on the methods they should use to prevent the attacks?
- How to provide better security for their data, their revenue, etc.

Methods:

Step 1: Exploratory Analysis

- The first method I used was exploratory analysis. I imported libraries modules that were necessary for the project. I defined the specific functions that were needed to plot the data. The figure below illustrates the model that could be used to detect malicious behaviors at Verizon. The data is obtained from different sources. The user information was obtained. Below is an example of data flow within the model.



Step 2: The second step involves obtaining the data and pre-processing it. The data plays a very important roles in creating the good models that can detect the breaches. To protect the privacy of Verizon's data, a synthetic type of dataset was used to create a program that will detect a breach in the version system. The dataset was generated form the Carnegie Mello University. The release r4.2 dataset was used for this project. I decided to split the dataset into 7 different parts. For example:

File	Description
Device.csv	Log of user's activity regarding connecting and disconnecting a thumb drive
Email.csv	Log of user's e-mail communication
File.csv	Log of user's activity regarding copying files to removable media devices
http.csv	Log of user's internet browsing history
Logon.csv	Log of user's workstation logon and logoff activity
psychometric.csv [1]	Users' psychometric scores based on big five personality traits
Ldap [1]	A set of eighteen (18) files regarding the organizations' users and their roles

Step 3: The features within the dataset were encoded during the pre-processing. The dates, times, and additional features were converted into numbers, so the machine learning algorithms can understand it, machine learning only understands integers. For example:

Feature	Possible Values
Day	0, 1, 2, 3, 4, 5, 6
Time	1, 2, 3, 4, . . . , 24
User	String Type
PC	String Type
Activity [1]	1, 2, 3, 4, 5, 6, 7

Step 4: Machine learning was used to search and discover various patterns within the data to predict and detect anomalies. The machine learning method was the best method used to detect the various anomalies. The algorithm was created to detect normal and malicious activities. The Local Outlier Factor (LOF) is an algorithm classified as unsupervised. Every data point was assigned a score. If the point has lower density versus the other neighboring points, it is considered to be anomaly. The dataset that created the highest precision was selected for the first optimal. The insider detection rate (DR) and precision was determined during the use of two equations:

$$DR = \frac{TP}{TP + FN}$$

$$Precision = \frac{TP}{TP + FP}$$

The TP stands for true positive, and it represents an integer of a true threat that was detected. FP stands for false positive and it represents integers with normal types of users that did not present a threat to the data. A false negative means the malicious threat was not detected and was falsely determined as a normal type of user. Below is a figure of the samples that were obtained from initial synthetic type of dataset:

Step 5: The feature selection results were fitted on a thousand rows of the data and the results are shown below. The results are represented by the seven columns. The accuracy for testing and train was based on the parts that were sliced for every value. For example:

	BPSO	BMVO	BGWO	BMFO	BWOA	BFFA	BBAT
Time Taken [1]	9.4016	11.3000	8.8428	8.5710	12.0246	10.5799	9.3706
Train Accuracy [1]	0.9985	0.9939	0.9939	0.9939	0.9939	0.9939	0.8030
Test Accuracy [1]	0.9941	1	1	1	1	1	0.8235
Number of Iterations x Features Selected [1]	7 × 4 Feat. 4 × 5 Feat. 2 × 6 Feat. 4 × 7 Feat.	2 × 4 Feat. 15 × 5 Feat. 2 × 6 Feat.	0 × 4 Feat. 19 × 5 Feat.	0 × 4 Feat. 19 × 5 Feat.	2 × 4 Feat. 15 × 5 Feat. 3 × 6 Feat.	3 × 4 Feat. 14 × 5 Feat. 0 × 6 Feat. 3 × 7 Feat.	5 × 4 Feat. 8 × 5 Feat.

The split ratio of the dataset was 0.34, testing was 34% accuracy, and training set had 66% accuracy. All features were selected for example:

	BPSO	BMVO	BGWO	BMFO	BWOA	BFFA	BBAT
Time Taken [1]	9.4016	11.3000	8.8428	8.5710	12.0246	10.5799	9.3706
Train Accuracy [1]	0.9985	0.9939	0.9939	0.9939	0.9939	0.9939	0.8030
Test Accuracy [1]	0.9941	1	1	1	1	1	0.8235
Number of Iterations x Features Selected [1]	7 × 4 Feat. 4 × 5 Feat. 2 × 6 Feat. 4 × 7 Feat.	2 × 4 Feat. 15 × 5 Feat. 2 × 6 Feat.	0 × 4 Feat. 19 × 5 Feat.	0 × 4 Feat. 19 × 5 Feat.	2 × 4 Feat. 15 × 5 Feat. 3 × 6 Feat.	3 × 4 Feat. 14 × 5 Feat. 0 × 6 Feat. 3 × 7 Feat.	5 × 4 Feat. 8 × 5 Feat.


```
#select all features
selected_features = ['day', 'time', 'Logon', 'Logoff', 'Connect', 'Disconnect', "email", "file", "http"]
```

The local outlier factor results that contains optimization is below,

Experiment [1]	Features	Time_Taken	TP	FP	FN	DR	Precision
1	9.0	8.604782	70.0	918.0	0.0	1.000000	0.070850
2	5.0	6.609488	69.0	910.0	1.0	0.985714	0.070480
3	4.0	5.928126	62.0	774.0	8.0	0.885714	0.074163
4	7.0	7.686987	70.0	918.0	0.0	1.000000	0.070850
5	6.0	6.994491	69.0	919.0	1.0	0.985714	0.069838

Step 6: The results displays swarm algorithms are great to use for detecting threats. The swarm algorithms should be used to enhance the speed and accuracy of detecting the malicious users' behaviors.

Potential Issues:

- Lack of accuracy of data provided in results
- Data in dataset is not accurate
- No solutions can be drawn from dataset
- The insider threat can not be detected due to lack of data

Conclusion

Insider threats can occur in any type of business, industry, and healthcare domains. The purpose of the project is to bring the awareness of insider threats to businesses so they can obtain tactics to monitor and prevent the insider and its threat. The impacts that the insider threat can cause to a business may be high, medium, and low. It is important for businesses to know this information to prevent out of pocket expenses, stolen assets of data, and protect their own reputation.

References

Catescu, G. (2018). *Detecting insider threats using Security Information and Event Management (SIEM)* (Doctoral dissertation, University of Applied Sciences Technikum Wien).

Kim, A., Oh, J., Ryu, J., & Lee, K. (2020). A Review of Insider Threat Detection Approaches With IoT Perspective. *IEEE Access*, *8*, 78847-78867.

Rizzo, P., Jemmali, C., Leung, A., Haigh, K., & El-Nasr, M. S. (2018, July). Detecting betrayers in online environments using active indicators. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation* (pp. 16-27). Springer, Cham.

Kim, A., Oh, J., Ryu, J., & Lee, K. (2020). A Review of Insider Threat Detection Approaches With IoT Perspective. *IEEE Access*, *8*, 78847-78867.

Catescu, G. (2018). *Detecting insider threats using Security Information and Event Management (SIEM)* (Doctoral dissertation, University of Applied Sciences Technikum Wien).

Saxena, N., Hayes, E., Bertino, E., Ojo, P., Choo, K. K. R., & Burnap, P. (2020). Impact and Key Challenges of Insider Threats on Organizations and Critical Businesses. *Electronics*, *9*(9), 1460.

Väisänen, T. (2017, September). Categorization of cyber security deception events for measuring the severity level of advanced targeted breaches. In *Proceedings of the 11th European Conference on Software Architecture: Companion Proceedings* (pp. 125-131).

Rizzo, P., Jemmali, C., Leung, A., Haigh, K., & El-Nasr, M. S. (2018, July). Detecting betrayers in online environments using active indicators. In *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction and Behavior Representation in Modeling and Simulation* (pp. 16-27). Springer, Cham.

Weber, K., Schütz, A. E., & Fertig, T. (2020). Insider Threats–Hidden Enemies. *HMD Praxis der Wirtschaftsinformatik*, *57*, 613-627.

https://www.code42.com/resources/ebook-usual-suspects-of-insider-threat/?utm_content=usual%20suspects%20of%20insider%20threat&utm_source=google&utm_medium=cpc&utm_campaign=ENT_Insider%20Threat%20-%20Ge neral%20-%20Search&utm_term=insider%20threat&gclid=EAIaIQobChMIwKDZzYG77QIVkueGCh2b6wamEAAYASAAEgLXx_D_BwE

https://www.isdecisions.com/insider-threat/statistics.htm

https://kilthub.cmu.edu/articles/dataset/Insider_Threat_Test_Dataset/12841247/

Verizon Data Breach Investigations Report 12th Edition. Available online: https://www.verizonenterprise.com/verizon-insights-lab/dbir/ (accessed on 30 August 2019). 2. Widup, S.; Spitler, M.; Hylender, D.; Bassett, G. Verizon Data Breach Investigations Report 11th Edition. Available online: https://www.verizonenterprise.com/verizon-insights-lab/dbir/

Verizon Enterprise. Verizon Data Breach Investigations Report 10th Edition. Available online: https://www.verizonenterprise.com/verizon-insights-lab/dbir/ Schultz, E.E. A framework for understanding and predicting insider attacks. Comput. Sec. 2002, 21, 526–531.