Insider threat classification using deep learning

SUPERVISOR :

Dr.Urooj Ainuddin

Assistant Professor

Presented By:

- Muhammad Ali

What is an insider threat ?

- Insiders are the trusted employees of the company having legitimate access to resources.
- Insider threats are the risks that stems from users' access to critical data, storage and other resources when it becomes misappropriated.

Types of insider threats

Negligent

Insiders who pose an unintentional threat due to human error and lack of security awareness

Malicious

Current or former employees who abused their access to steal intellectual property for personal gains

Third - Party

Vendors who misuse their access and compromise the security of critical data

How big is this problem?

- According to FBI survey insider attacks are 50% more costly then external attacks.
- Average time to identify an insider attacks is 75-80 days.
- They know the configuration of system and weaknesses.
- Insider threat incidents have risen 44% over the past two years.
- The cost of credential theft to organizations increased 65% from $2.79 million in 2020 to $4.6 million at present.

Source: Ponemon Institute “Cost of insider threats global report 2022”

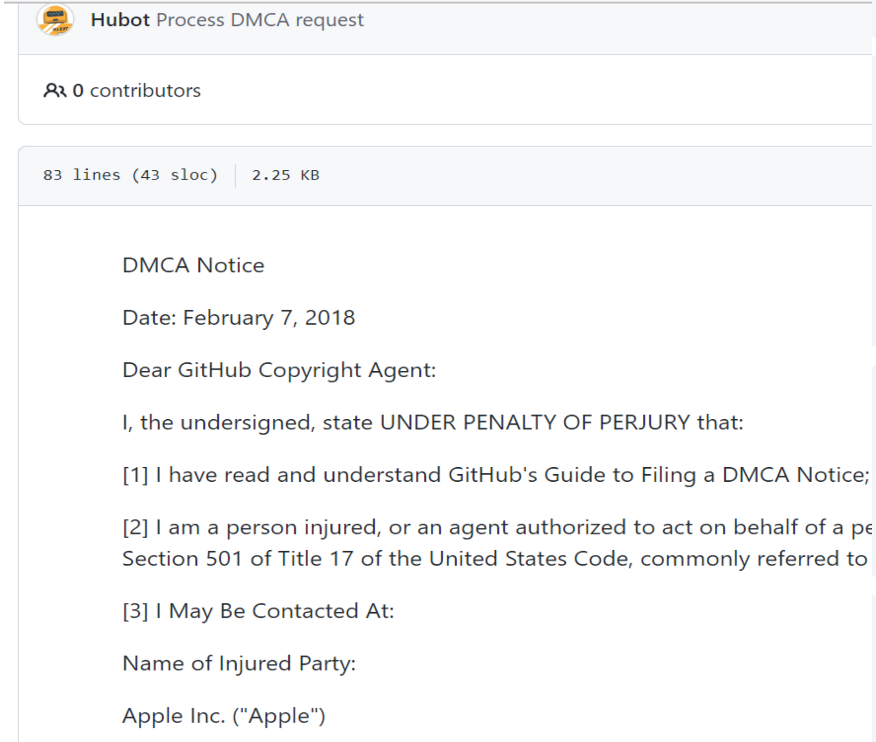
Twitter insider attack incident

- In July 2020 high-profile accounts on Twitter were hacked and used for illicit bitcoin transactions.
- Profiles including of Barack Obama and Elon Musk.
- Hackers were able to get into the Twitter admin Slack channel and from there got access to administrative tools.
- Losses accrued are estimated to be $250 million.

Source: https://resources.infosecinstitute.com/topic/8-of-the-worlds-biggest-insider-threat-security-incidents/

Apple insider attack incident

- In 2018, Apple’s IOS source code was leaked in a non-malicious way by an intern.
- Apple put a Digital Millennium Copyright Act (DMCA) takedown notice on GitHub.



Hubot Process DMCA request

0 contributors

83 lines (43 sloc) | 2.25 KB

DMCA Notice

Date: February 7, 2018

Dear GitHub Copyright Agent:

I, the undersigned, state UNDER PENALTY OF PERJURY that:

[1] I have read and understand GitHub's Guide to Filing a DMCA Notice;

[2] I am a person injured, or an agent authorized to act on behalf of a pe
Section 501 of Title 17 of the United States Code, commonly referred to

[3] I May Be Contacted At:

Name of Injured Party:

Apple Inc. ("Apple")

Source: https://resources.infosecinstitute.com/topic/8-of-the-worlds-biggest-insider-threat-security-incidents/

Solution to the problem

- It is found that insiders taking part in these threats tends to exhibit certain characteristics and observable behaviours.
- Machine Learning and security analytics techniques can used to classify those behaviours as malicious or non malicious.

User entity behaviour monitoring

- From where the data is coming from.
- How information is being accessed.
- Logon patterns of users.
- External devices usage such as USBs.
- How the browser is being accessed.
- Email exchanges.
- Uploading/downloading of files.

Dataset

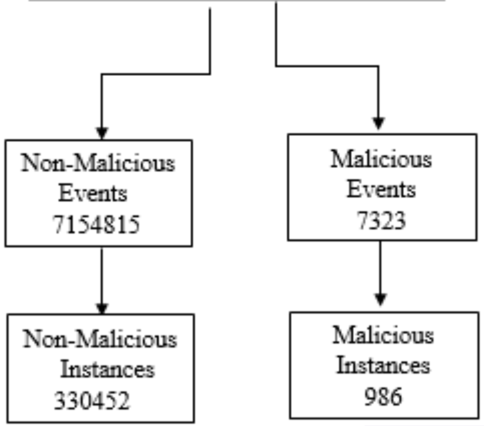
- Finding a real-world data is quite challenging as companies don’t share real world data.
- CMU published an insider threat dataset in 2020.
- CMU CERT insider threat dataset is a synthetic benchmark raw dataset for this problem.
- Total size is around 90 GigaBytes.
- Version v4.2 is used in this project.

Dataset

- This synthetic dataset generation uses various scenarios to define the malicious activities.
- The log files contains:

  1. Logon-logoff data
  2. Web browsing history
  3. File access logs,
  4. Emails sent and received
  5. Device usage
  6. Psychometric Scores

LOGS	COUNT.OF ACTIONS
Email file (.csv)	2629979
Http file (.csv)	1048575
Device file (.csv)	405380
File file (.csv)	445581
Logon file (.csv)	854859

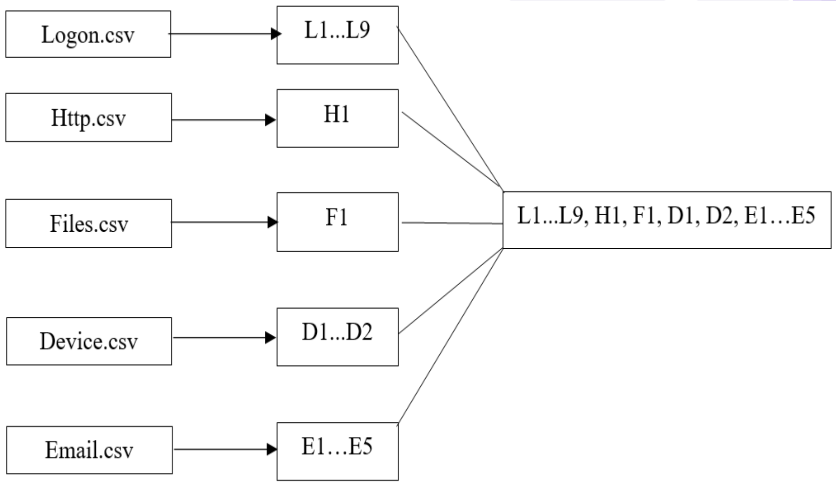


Feature Set Extracted from the raw data

Log file	Features	Description
Login	L1	Difference between office start time and first login time
	L2	Difference between last login time and office end time
	L3	Average difference in time between office start time and number of logins before office hours
	L4	Average difference in time between office end time and number of logins after office hours
	L5	Total number of logins
	L6	Number of logins outside office hours
	L7	Number of systems accessed
	L8	Number of systems used outside office hours
	L9	Average session length outside office hours
Email	E1	Count of emails sent outside the domain of organization
	E3	No. of attachments
	E4	Average email size
	E5	Number of recipients
Device	D1	Count of thumb drive usage outside office
	D2	Count of external device usage
File	F1	Number of .exe files downloaded
Http	H1	Count of usage of wikileaks.org

Pre-processing: Feature Vector Construction

- An array of employees per day usage/behaviour is known as a feature vector.
- These feature vectors are used for behavioural analysis.



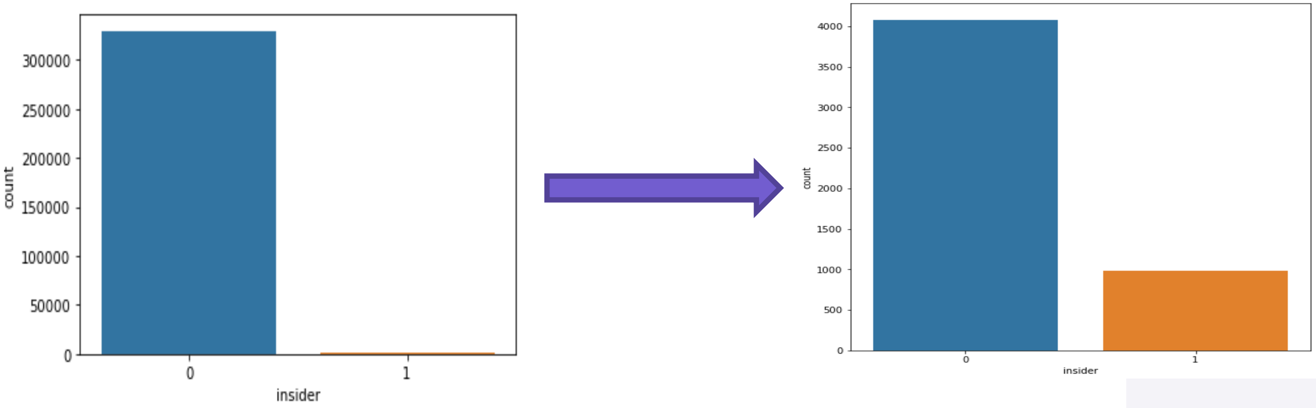
Feature Vectors

user	day_num	L1	L2	L3	L4	L5	L6	L7	L8	L9	E1	E3	E4	E5	D1	D2	H1	F1	insider
JCR0770	192	61.000000	539.000000	0.000	0.0	1	0	1	0	0.0	2	5	32177.666667	3	0	0	0	0	0
PGC0066	126	33.000000	567.000000	0.000	0.0	1	0	1	0	26.0	1	0	28616.818182	11	0	0	0	0	0
SJC0333	337	11.000000	611.000000	11.000	0.0	1	1	1	1	53.0	0	0	0.000000	0	0	0	0	0	0
FHS0837	483	63.000000	537.000000	0.000	0.0	1	0	1	0	119.0	0	0	0.000000	0	0	0	0	0	0
PLH0715	211	18.000000	582.000000	0.000	0.0	1	0	1	0	0.0	0	0	0.000000	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
JJM0203	274	18.000000	286.500000	18.000	0.0	2	1	1	1	18.0	0	0	0.000000	0	0	0	0	0	1
LDG0459	249	30.000000	335.400000	0.000	0.0	2	0	1	0	31.0	0	0	0.000000	0	0	0	0	0	0
HIW0536	7	16.000000	584.000000	0.000	0.0	1	0	1	0	40.0	10	0	24949.454545	11	0	0	0	0	0
SVS0871	487	13.000000	296.316667	0.000	0.0	2	0	1	0	25.0	0	0	0.000000	0	0	0	0	0	0
SOB0360	270	175.783333	402.116667	90.525	0.0	4	2	4	2	346.2	0	0	0.000000	0	0	0	0	0	0

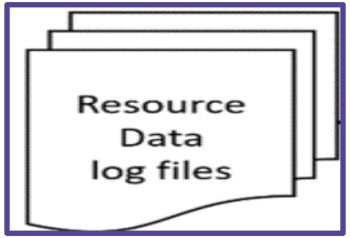
Pre-Processing: Handling with class imbalance problem

The original data has an imbalance ratio of 1:340 for non-malicious and malicious classes.

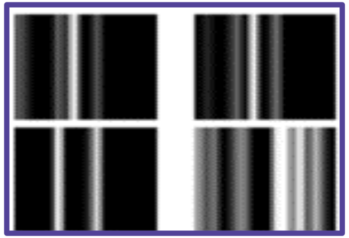
Sampling ratio of 25 is used to perform random undersampling technique which gives optimal results.

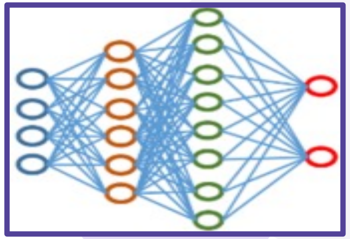


Proposed Methodology



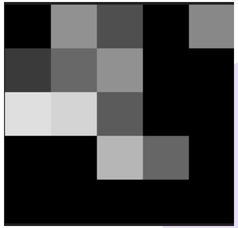
Feature
Extraction





Grayscale Image Generation

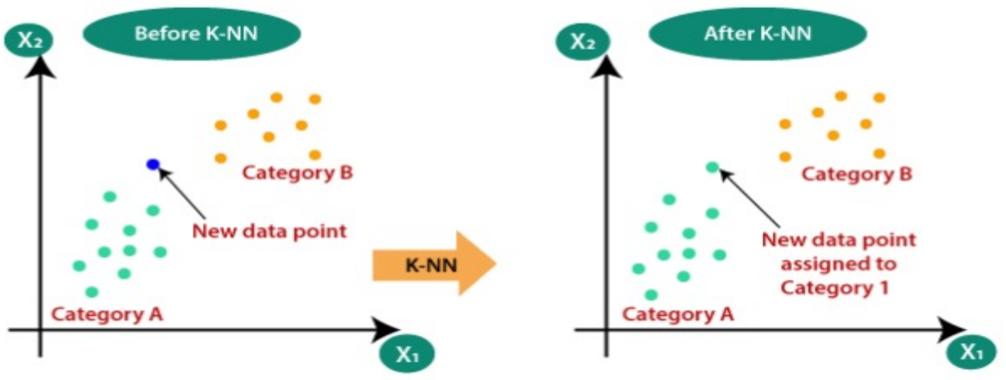
- All the feature vectors need to be converted into grayscale images.
- Image is composed of pixels values ranging from 0 to 255.
- Computer vision techniques can be used to create grayscale images.
- Tab2Img library is used for image generation.
- Then DCNN approach of image classification is performed for anomaly detection.



Model training and testing

1. K-Nearest Neighbour (KNN)

- KNN stands for ”k-nearest neighbors”, a traditional machine learning algorithm used for classification.
- In KNN, the predicted output for a new data point is based on the labels of its k-nearest neighbors in the training data.



Evaluation Results for KNN

- 1D feature vectors are used in this model.
- Optimal results obtained by choosing the K value at 5.

Class	Precision	Recall	F1-Score	Accuracy
Non-malicious	0.95	0.96	0.96	0.93
Malicious	0.85	0.81	0.83	

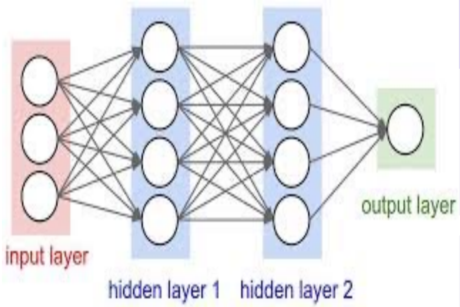
2. Random Forest

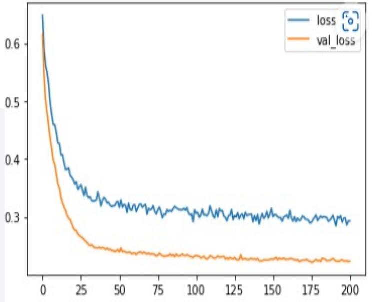
- 1D feature vector is used in this approach.
- Random Forest is an ensemble learning method that constructs multiple decision trees and outputs the mode of the classes (classification).
- **Evaluation Results for Random Forest**

Class	Precision	Recall	F1-Score	Accuracy
Non-malicious	0.96	0.96	0.96	0.93
Malicious	0.83	0.83	0.83	

3. Deep Neural Networks (DNNs)

- They are composed of layers of interconnected nodes, or neurons, that process and transmit information.
- Input layer contains 17 nodes and 2 hidden layers.
- Hidden layers contains 9 and 5 neurons.
- Output layer uses sigmoid activation function.
- To avoid overfitting early stopping criteria and a dropout layer with a rate of 0.5 is used.
- 1D feature vector array is used for training and testing.





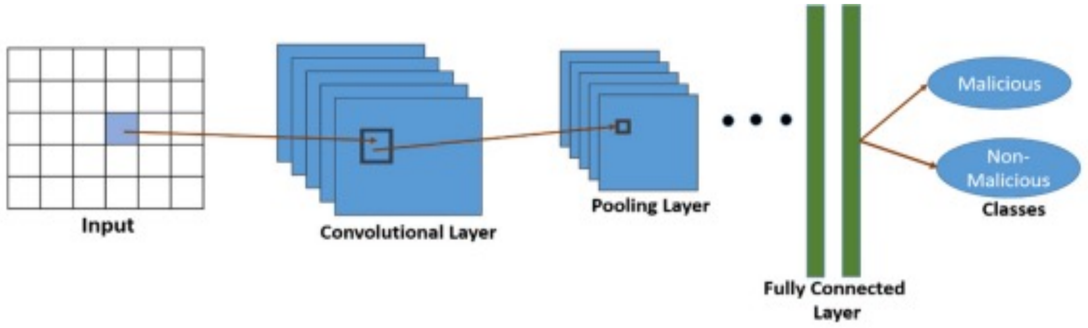
3. Deep Neural Networks (DNNs)

- **Evaluation Results for DCNN**

Class	Precision	Recall	F1-Score	Accuracy
Non-malicious	0.95	0.96	0.95	0.93
Malicious	0.85	0.79	0.82	

3. Deep Convolutional Neural Networks (DCNNs)

- DCNNs are a type of artificial neural network commonly used in computer vision tasks, such as image classification and object detection.
- They are called "convolutional" because they use convolutional layers, which apply filters to the input data to extract features and patterns.



3. Deep Convolutional Neural Networks (DCNNs)

- Grayscale images from our generated grayscale image dataset for training and testing.
- In this approach 3 convolutional layers and 2 fully connected layers are used, with the number of nodes as [32x64x128x64x1].
- Sigmoid activation function is used in the output layer.
- Dropout rate of 0.5 is used in dropout layer to avoid overfitting.

3. Deep Convolutional Neural Networks (DCNNs)

- **Evaluation Results for DCNN**

Class	Precision	Recall	F1-Score	Accuracy
Non-malicious	0.96	0.94	0.95	0.93
Malicious	0.80	0.87	0.83	

Evaluation of performances of all Models

Model	Accuracy	Precision	F1-Score	Recall
KNN	0.93	0.85	0.83	0.81
Random Forest	0.93	0.83	0.83	0.83
DNN	0.93	0.85	0.82	0.79
DCNN	0.93	0.80	0.83	0.87

- A high precision indicates that the classifier has a **low rate of false positives**, means it rarely flags non-malicious insiders as malicious.
- A high recall indicates that the classifier has a **low rate of false negatives**, means it rarely misses actual malicious insiders.
- It's important to have a high recall to avoid missing any malicious insider, as the consequences can be severe, which is ensured by DCNN.

Reasons why DCNN outperformed

- Convolutional layers or filters have captured the non-linear boundaries or behaviours from the dataset efficiently as the features are not really correlated with the label.
- Additionally, CNNs are able to learn hierarchical representations of the input data. This allows CNNs to automatically and adaptively learn spatial hierarchies of features from the input image.

Key takeaways

- Image based approach with CNN outperformed other models in our problem.
- Converting a numeric data into images allow us to use a variety of DCNN models such as Transfer learning, Yolo, MobileNet, VGG.
- Accuracy is not the right matrix for imbalanced dataset.
- But still, DCNN doesn't guarantee the best results in every case as it depends on specific problem and dataset.
- Being self-sufficient is really important in life !

MILESTONES ACHIEVED

Summary

Milestones Achieved

- Literature review
- Grayscale image conversion
- Data acquisition & extraction
- Model training and testing
- Feature extraction & Feature vector construction
- Evaluation of models

Future Plans

1. Use of Transfer Learning
2. Research Paper
3. Attend Conferences
4. Application development
5. Insider threat management software

Competitors:
Exabeam
Proofprint

THANK-YOU

Any questions & suggestions?