Deep Learning-Based Framework for Detecting Malicious Insider-Inspired Cyberattacks Activities in Organisations

Gibson Chengetanai[1], Tendai Chandigere[1], Pepukai Chengetanai[2] and Rachna Verma[1]
[1]School of Computing and Information Systems, BAC, Gaborone, Botswana
[2]School of Accounting and Finance, GIPS, Gaborone, Botswana

gibsonc@bac.ac.bw

Abstract: Cyberattacks are happening at an alarming rate both in developed and developing countries. This is due to more users now being connected to the global village (internet). Significant strides have been taken by organisations to protect information technology assets together with data, by doing defense-in-depth, using firewalls and access control approaches collectively. These approaches work well in detecting attacks by outsider cyber-attackers. In recent cyberattacks the perpetrators have been those within the organisation, as they can easily bypass security measures especially those with high privileges and they can go undetected for quite a long time. We propose a deep learning approach termed Automatic_ IDS_ Deep model (framework) that is infused with intrusion detection systems to give timely detection of malicious activities by those within the organisation. Experiments were conducted and averaging of results was done to determine accuracy, recall, and precision of the proposed model. The model (framework) offers better results on its performance in detecting attacks that are perpetrated within the organisation.

Keywords: Deep learning, Intrusion detection system, Cyberattacks, Malicious users

1. Introduction

Cyberattacks are happening every day on a larger scale as more and more people get connected to the internet. Users in the workplace or in their homes, are all now going online increasingly on social networking sites such as X formerly known as Twitter, Facebook, TikTok among the many such web services available these days. Employees in organisations can divert some of their time to network with fellow professionals, friends, and other people around the globe (Smitha, Siano and Parenta, 2023; Lo et al, 2018).

Organisations through their IT specialists are putting measures to detect malicious attacks attempts towards the IT infrastructure sand its associated data and systems in a timely manner. Despite corporate attempts to protect valuable IT assets, cyber-attacks are still occurring, though business entities do not divulge these attempts for fear of reputational damage. It has been observed that many of the attempts to prevent attacks are effective for people outside the organisation. Insiders within organisations (usually disgruntled employees), and contract vendors can pose major a threat in the organisation (Insider threats[6]; Smitha, Siano and Parenta, 2023). Cyberattacks are due to the following (Figure 1):

Financial gain - employees can steal corporate data and sell it to the organization's rivals (Fujii, Kurima and Isobe, 2019). Hackers, for example can disable accounts and create new accounts once they gain privileged access to systems. Hackers can copy data and use that copied data which may include pricing models that have been used in the organisation and some can even go with the corporate data when they leave the organisation (Smitha, Siano and Parenta, 2023; Lo et al, 2018).

Theft of intellectual property – this usually occurs in state sponsored entities, whereby governments spy on each other on areas of research and development. Around 44% of cyberattacks are around intellectual property theft (Inayat et al., 2022).

Disgruntled employees and organizational politics can all contribute to individuals stealing corporate data as they are unhappy with the status quo in the organisation. Figure 1 shows the main motivation for attacks perpetrated within the organisation (Smitha, Siano and Parenta, 2023).

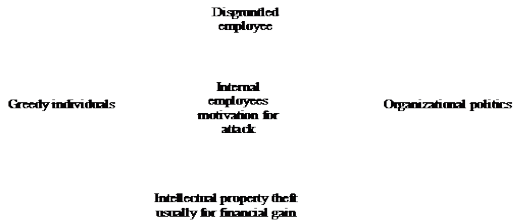

Figure 1: Motivations for insider cyberattacks in organisations

Organizational politics can lead to some employees wanting to punish the employer as they are disgruntled by how things are running in their workplace (Modini et al, 2020). This could be because such people have access rights to the system and can easily evade detection. Greedy individuals can be individuals who want to make money by selling data to competitors for a fee (Saxena et al., 2020).

2. Related Work

Machine learning approaches have been used to help in detecting malicious events in IT systems in organisations. Machine learning approaches help in detecting some abnormal trends with high accuracy. Most of these machine learning approaches are deployed on host or network perimeters to help in detecting an anomaly based on the signature-based approach. Machine learning approaches that have been used to detect anomalies in IT systems include support vector machines (SVM). SVM works in multiclass environment, and they are good at solving linear problems (Saxena et al., 2020; Modini et al., 2020). SVM approach does not scale well on huge data sets whilst artificial neural networks (ANNs) are prone to overfitting (Weng, Li and Zhu, 2020), and it requires more time for model training as compared to other machine learning approaches (Liu, 2021; Lang, 2019) . Naïve bayes is a supervised machine learning based approach that work based on Bayesian theorem and its advantage is of simplicity and easy of computation (Al-Shehariand Alsowail, 2021; Al-Mhiuai et al, 2020). K means clustering works by discovering clusters in the data object with k being number of clusters that have been formed (Valliammal and Shaju, 2018; Gayathri, Sajanhar and Xiang, 2020). When clustering is done, objects in a particular cluster will have similar characteristics. Logistics regression (LR) LR works well on linear data and thus fails to work well on non-linear data sets (Smitha, Siano and Parente, 2023). Decision trees are another machine learning method, however its weaknesses is that it does not consider correlation in decision making (Lo et al, 2018; Gayathri, Sajanhar and Xiang, 2020).

Anomaly based detection methods check on user behaviour and where there is a deviation from the expected behaviour, alert messages are sent to the administrators (Weng, Li and Zhu, 2020). The machine learning approaches each have some weakness in using them since most of them are associated with high false positive alert messages (Gayathri, Sajanhar and Xiang, 2020). From literature it is evident that machine learning approaches to finding malicious activities perpetrated within the organisation have a weakness of giving less accuracy. This can be attributed to machine learning approaches' inherent use of small data for training and testing data sets. A deep learning method is proposed which is a subset of machine learning. Deep learning approach was chosen mainly because it provides better accuracy relative to machine learning approaches since it uses huge data during model training.

3. Proposed Automatic_IDS_ Deep Learning Framework

Our contribution is as follows:

- Development of a deep learning approach called Automatic_IDS_Deep framework which takes into consideration a number of activities that are extracted from log files such as logon and logout times, devices used to connect, http sites visited, event log relating to removable devices such as flash memory and USB devices connected to the systems and times they have been connected, communication patterns, printer activity logs, database audit logs and access logs.
- New algorithm coupled with experimental results and evaluations to determine the accuracy of the proposed Automatic_IDS_Deep model.

Our proposed framework is based on user behaviour compilation looking into activities that users do on the system on a daily basis. These include checking and extracting information such as logon times, daily system logs including systems logs, web server logs and database audit logs. Event logs together with the device used to logon to the system and traffic send with particular attention to actions such as failed events from various IP addresses will be recorded. When activities are recorded, preprocessing is done to remove any redundancies, i.e. during the data cleaning process. Missing values are replaced with average values and an algorithm is infused into the system that does preprocessing. After data cleaning, feature extraction is done to determine on what is normal considered normal or otherwise, based on historical data. On feature extraction the following Table 1 shows the feature set variables that will be recorded to determine malicious users.

Table 1: Features to be selected when profiling inside user (internal employees) activities to determine malicious ones

Domain	Feature	Feature set to be observed
Network	HTTP	URLs, browser type, downloads
	Email	Source address, destination address, time, attachments
	TCP/IP	Source and destination IP addresses, size of data packets, data traffic size
Host	File and folders	Check on open, create, copy, move, delete operations
	USB	Device ID, duration, and associated actions on the device such as copy, delete, or paste
	System call	Process name and ID, registry, and file system
	Login or logout operations	User ID, timestamp, login duration, login frequency and logout time

After feature selection is done, we train and test data on the proposed model and evaluate its performance. Our proposed solution offers better solution of reducing false positives on alert messages because it considers all activities that users can do. The algorithm below shows the Automatic_IDS_Deep Learning based framework for detecting malicious events in the workplace.

Algorithm 1. Deep learning-based algorithm for detecting malicious events in the workplace

```
INPUT:    Check on data coming from different sources on the
            network β= {logon log, Database log, file log, print
             log, web server log, error log, file access log, TCP/IP,
              flow log, events log}
               iteration (t)
OUTPUT:   Network status on any cyberattacks showing machines
            that have been affected by malicious events
           Classification for malicous and non-malicious actions for review by network administrators
Procedure
Step 1   iteration:=0
Step 2: examine data packets in the network
      // data logs can be coming from logon, database file, print log,
      web server, error log, file access, TCP/IP flow log, events log
Step 3: pre-process using the configured algorithm
Step 4:  use feature selection algorithm
Step 5:  Training the model    //extract relevant features from data sets and train model with Epochs=200
          if result point to malicious activities then
          add the workstation or mobile nodes that has such
               activities
         else   //No malicious activities on the network
             save the workstation or mobile terminal as correct node
         end if
Step 6:  Evaluate the performance of the model
Step 7: Display the results on malicious nodes as generated in step 5
      t=t+1
end procedure
```

Organisations can also have the log file duplicated to trusted offsite parties that are credible. This will help in detecting malicious activities that might be done by administrators who might have privileges. These privileges include deleting data/evidence that might be needed to detect the culprits inside the organisation as they have capabilities for deleting user activities in the log file(s).

4. Experimental Results and Evaluation

In this section we present results showing the performance of the proposed Automatic_IDS_Deep model. The data set that was used to determine insider threat was synthetic data obtained from CERT, a project at Carnegie University. The items that make up the data set include login and logout, email and web, USB, system calls. The data set was generated using two processes to simulate an attack where malicious software can be injected into the system using removable devices such as memory sticks and lastly simulation using email attachments was done and injecting data to the Windows server 2019. Performance metrics that were explored were accuracy, recall, precision, and F1-score. The computation of each performance metric is below.

$$Accuracy = (TP + TN)/(TP + TN + FP + FN) \quad (1)$$

The F-measure of harmonic mean is computed as:

$$F1 - score = \frac{2 * precision * recall}{precision + recall} \quad (2)$$

with precision being computed as

$$Precision = TP/(TP + FP) \quad (3)$$

and recall also known as sensitivity computed as

$$Recall = TP/(TP + FN) \quad (4)$$

with TP being true positive that is an activity identified by the system as actually being malicious, TN is true negative being an activity that has been correctly identified as non-malicious and is actually non-malicious and lastly FN is for false negative which is an activity that is not detected by the system as malicious but is actually malicious

The results on Table 2 shows the performance of the proposed Automatic_IDS_Deep Model.

Table 2: Experimental results for the Automatic_IDS_Deep Learning framework

Training Accuracy	Training loss	Validation Accuracy	Precision	Recall	F1-Score
0.96	0.18	0.97	0.93	0.9	0.914

5. Conclusion

Cyberattacks in organisations have been observed to come from employees within the organisation. Internal employees who have privileged access to information that resides in the organisation can copy corporate data/information and give this information to the organization's competitors for a fee. This research has managed to come up with a framework , termed Automatic_ IDS_ Deep Framework. The model is novel in the sense that it infuses intrusion detection and combines it with deep learning which is based on convolutional neural networks. Synthetic data from Carnegie Mellon University was used to test our proposed model based on deep learning. Since most of the malicious activities happening in organisations are done by inside employees as seen from literature, we proposed a secure deep learning model called Automatic_ IDS_ Deep model. The adoption of this framework can help in mitigating attacks timeously by sending alert messages to administrators in the organisation for action. In future, we intend to do some comparison of our proposed model with other machine learning models such as (Naïve Bayes, SVM, etc.) and other deep learning models to check on the performance of our model relative to others.

References

Al-Shehari, T. and Alsowail, R.A., 2021. An Insider Data Leakage Detection Using One-Hot Encoding, Synthetic Minority Oversampling and Machine Learning Techniques. Journal of Entropy, Vol. 23 No. 10, pp.1-24

Fujii, S., Kurima, I. and Isobe, Y., 2019. Scoring Method for Detecting Potential Insider Threat based on Suspicious User Behavior using Endpoint Logs. Athens: The Steering Committee of The World Congress in Computer Science, Computer Engineering and Applied Computing (WorldComp), pp.1-7

Gayathri, R.G., Sajjanhar, A. and Xiang, Y., 2020. Image-Based Feature Representation for Insider Threat Classification. Applied Sciences, Vol .10 No. 14, pp4945-4953

Gheyas, I.A. and Abdallah, A.E., 2016. Detection and prediction of insider threats to cyber security: a systematic literature review and meta-analysis. Big Data Analytics, Vol .1 ,pp.1-29

Hammoudeh, M., Watters, P., Epiphaniou, G., Kayes, A.S.M. and Pinto, P., 2021. Special Issue "Security Threats and Countermeasures in Cyber-Physical Systems". Journal of Sensor and Actuator Networks, Vol. 10 No. 3, pp.1-4

https://techjury.net/blog/insider-threat-statistics/#gref Accessed 7 January 2024

Inayat, U., Muhammad, F.Z., Mahmood, S., Khalid, H.M. and Benbouzid, M., 2022. Learning-Based Methods for Cyber Attacks Detection in IoT Systems: A Survey on Methods, Analysis, and Future Prospects. Electronics, Vol. 11 No. 9, pp.1-42

Lang, B., 2019. Machine Learning and Deep Learning Methods for Intrusion Detection Systems: A Survey. Applied Sciences, Vol. 9 No. 20, pp.43-76

Liu, H., 2021. An insider threat detection system based on user and entity behavior analysis. Journal of Physics: Conference Series, Vol. 1999 No. 1, pp.1-11

Lo, O., Buchanan, W.J., Griffiths, P. and Macfarlane, R., 2018. Distance Measurement Methods for Improved Insider Threat Detection. Security and Communication Networks, pp.1-19

Modini, J., Vanzomeren, M., Fowler, S., Joiner, K. and Lynar, T., 2020. Rising to the Challenge of Insider Threats for Middle Powers. Reading: Academic Conferences International Limited, pp.1-6

Mohammed Nasser Al-Mhiqani, Ahmad, R., Z., Yassin, W., Hassan, A., Karrar, H.A., Ali, N.S. and Yunos, Z., 2020. A Review of Insider Threat Detection: Classification, Machine Learning Techniques, Datasets, Open Challenges, and Recommendations. Applied Sciences, Vol. 10 No. 15, pp.1-41

Saxena, N., Hayes, E., Bertino, E., Ojo, P., Kim-Kwang, R. and Burnap, P., 2020. Impact and Key Challenges of Insider Threats on Organizations and Critical Businesses. Electronics, Vol. 9 No 9., pp.1-29

Smitha, J.P., Siano, P. and Parente, M., 2023. Review of Cybersecurity Analysis in Smart Distribution Systems and Future Directions for Using Unsupervised Learning Methods for Cyber Detection. Energies, Vol. 16 No. 4, pp.1-24

Valliammal, N. and Shaju, B., 2018. Deep learning algorithm based cyber-attack detection in cyber-physical systems-a survey. International Journal of Advanced Technology and Engineering Exploration, Vol. 5 No. 49, pp. 489-494.

Weng, W., Li, W and Zhu, L. (2020). "Enhancing Medical Smartphone Networks via Blockchain-Based Trust Management Against Insider Attacks". IEEE Transactions on Engineering Management, Vol. 67, No. 4, pp.1377-1386