Survey on Machine learning Applications in Cyber Security Problems

[1]Ajay Singh, [2]Dr. Jitendra Sheetlani

Department of Computer Science
Sri Satya Sai University of Technology and Medical Sciences Sehore Bhopal (M.P.)

***Abstract*: This study suggested that the works previously done on machine learning (ML) based applications in Cyber Security. The organization of review on the existing published works is not historic and instead thematic. The paper comprises into the contains related works based on ML for DGA generated domain name detection, intrusion detection, network traffic data analysis, spam, and phishing activities data analysis, malware data analysis, and Android malware data analysis respectively. This study helps to understand cybersecurity and use of Machine learning application.**

***Keywords*: Machine learning, Cybercrime, malware, data analysis**

Inroduction

As of late, the nitty gritty review on different procedures to distinguish DGA created space names is finished by (Vinaykumar R, 2018; Zhauniarovich et al., 2018). In prior days, boycotting is the most normally utilized strategy. These strategies totally neglect to recognize new sorts or variations of DGA based space name (Kuhrer et al., 2014). Afterward, many methodologies have been presented in light of ML. The current deals with building ML based DGA classifier is ordered into 2 kinds like review and ongoing. Review techniques can be utilized as a traditionalist framework and can't be utilized for ongoing location and preven-tion (Antonakakis et al., 2012; Yadav et al., 2010, 2012). Fundamentally, the review strategies utilize a bunching way to deal with bunch the area names into various groups and gauge the factual properties for each bunch. Utilizing the measurable test, classifica-tion is completed. Furthermore, this sort of framework depends on relevant data like HTTP headers, NXDomains across an organization, and detached DNS to promote im-demonstrate execution. In the vast majority of constant endpoint framework, getting the relevant information is undeniably challenging. Ongoing recognition techniques work on a for each area premise utilizing just the space names alone rather than extra relevant data. The strategies display less execution for a genuine organization (Yadav et al., 2012). Also, the exhibitions got by these strategies are extremely less (Yadav et al., 2012) and these techniques depend on include designing. The definite examination of different component designing strategies with ML calculations to recognize DGA produced area name identification is supportive of presented by (Ashok et al., 2017). The proposed strategy is executed progressively at a network access supplier (ISP) level to screen the DNS occasions. The system is exceptionally versatile in high item equipment server and it can deal with 2 million occasions each second. The ML put together strategies based with respect to include designing for DGA space name identification techniques are helpless in an ill-disposed climate. The ML based DGA classifiers have been sentenced because of high misleading positive rates. To shield DGA created space, the strategy needs to grasp hidden area attributes, yet in addition develop and adjust quickly. As the techniques utilized by the enemy have been continually changing, the models that are prepared in disconnected utilizing earlier day information can not recognize promotion advanced DGAs that arose as of late. Further, the Expense Unfeeling ML models are inclined to multiclass imbalanced DGA family order. To conquer the current issues, this examination work fosters an extensive data set and a definite execution assessment of different Expense Inhumane and Cost-Delicate profound learning (DL) structures involving the data set in DGA identification and arrangement is accomplished in this work.

Homoglyph or name satirizing assaults are utilized to muddle space names and the ongoing methodologies depend on string coordinating and Levenshtein or alter distance which are computationally weighty and delivers high bogus positive outcomes. A significant measure of work has been done on concentrating on productive string matching strategies. (Deng et al., 2013,2014) are a portion of the works that attention on growing quick and productive string matching strategies and different works like (Wang et al., 2011) centers around improving the nature of closest neighbor search strategy. The traditional methodologies for recognizing name ridiculing in view of string matching are not successful (Woodbridge et al., 2018). Consequently, a ton of studies propose name satirizing recognition methods in light of visual examinations. In (Dark, 2008) and (Linari et al., 2009), the creators have made a custom alter distance which takes visual similitude of the characters into the record. Substitution of character by an outwardly comparable person will result in more modest alter distance than an outwardly dissimi-lar character. These strategies rely upon physically determined similitude measures between the characters and it did exclude the enormous Unicode set. These custom alter distance approaches are contrasted and the standard alter distance technique in (Woodbridge et al., 2018) and it very well may be seen that is no critical improvement in the outcomes it actually prompts high misleading positive rate. Following (Woodbridge et al., 2018), in this examination work, the presentation assessment of different Siamese organizations is done and most impor-tantly as opposed to changing over the characters into picture design, the characters are changed over into numeric arrangement and implanting is utilized to become familiar with the thick component portrayal. Advertisement ditionally, the area name ridiculing and DGA recognition and characterization modules are incorporated to lessen the deception rate.

Interruption Location

The examination on security issues connecting with network interruption location framework (NIDS) and have based interruption recognition framework (HIDS) exists since the introduction of PC models. As of late, applying ML based answers for NIDS and HIDS are of prime interest among security analysts and subject matter experts. An itemized study on existing ML based arrangements are talked about exhaustively by (Mishra et al., 2018)

Business NIDS essentially utilize either factual measures or figured limits on highlight sets like bundle length, between appearance time, stream size, and other organization traffic boundaries to successfully show them inside a particular time-window (Mishra et al., 2018). They experience the ill effects of high pace of misleading positive and bogus negative alarms. A high pace of misleading negative alarms shows that the NIDS could neglect to identify goes after more as often as possible and a high pace of bogus positive cautions implies the NIDS could superfluously alarm when no assault is really occurring. Thus, these business arrangements are inadequate for present day assaults.

Self-learning framework is one of the powerful techniques to manage the current day at-tacks. This utilizations directed, semi-administered, and solo systems of ML to become familiar with the examples of different typical and noxious exercises with an enormous corpus of nor-mal and assault organization and host-level occasions. However different ML based arrangements are found in the writing, the relevance to business frameworks is in beginning phases. The current ML based arrangements yield high bogus positive rate with a high computational expense (Staudemeyer, 2015). This is on the grounds that ML classifiers become familiar with the attribute of basic TCP/IP includes locally.

An enormous investigation of scholastic exploration utilized the true standard benchmark information, KD-DCup 99 to work on the viability of interruption recognition rate. This informational index was utilized in KDDCup 98 and KDDCup 99 test. Absolutely, 24 sections were submitted in the KD-DCup 98, in that 3 winning passages utilized variations of choice tree (DT) to whom they showed just the peripheral insights importance in execution. The ninth winning section in the challenge utilized the 1-closest neighbor classifier. The principal massive contrasts execution was found somewhere in the range of seventeenth and eighteenth passages. This gathered that the initial 17 entries technique were vigorous and were profiled by (Staudemeyer, 2015). After the test, the majority of the distributed consequences of KDDCup 99 have utilized a few component engi-neering strategies for dimensionality decrease. While few investigations utilized specially constructed informational indexes, greater part utilized similar informational index for recently accessible ML classifiers.

Generally, a far reaching writing audit shows not very many examinations utilized current DL approaches for NIDS and the regularly utilized benchmark informational indexes for exploratory butt-centric ysis are KDDCup 99 and NSL-KDD (Javaid et al., 2016; Kim et al., 2016; Staudemeyer, 2015; Yin et al., 2017). The IDS in view of repetitive brain organization (RNN) outflanked other traditional ML classifiers in recognizing interruption and interruption type on the NSL-KDD informational collection (Yin et al., 2017). Two-level methodology was proposed for IDS in which the main level concentrates the ideal highlights utilizing scanty autoencoder (AE) in a solo manner and grouped utilizing softmax relapse (Javaid et al., 2016). The use of stacked AE was proposed for ideal component extraction in a solo manner where the supportive of presented strategy is totally non-symmetric and characterization was finished utilizing Arbitrary woodland (RF). Novel LSTM engineering was proposed and by displaying the organization traffic data in time series acquired better execution. The proposed technique performed very much contrasted with every one of the current strategies as well as KDDCup 98 and 99 test passages (Staudemeyer, 2015). However these DL put together techniques showed better execution with respect to the KDDCup 99 and NSL-KDD informational indexes, organization in business framework is in many cases restricted. The essential explanation is that the informational indexes are extremely old and in particular execution assessment of these structures is expected on different informational indexes. Another significant trademark is that this design can gain the elements from the crude organization pack-ets. Nonetheless, in the current investigations, the removed highlights are passed as contribution to become familiar with the attributes among authentic and noxious organization exercises. Following, in this work, the utilization of profound brain organizations (DNNs) is read up for NIDS and HIDS and the presentation of DNNs is assessed on different informational collections of NIDS and HIDS.

Network Traffic Information Investigation

The organization traffic is expanding dramatically. Recognizing and checking network traffic is of critical assignment towards distinguishing the pernicious exercises. As of late, the nitty gritty overview on different methods for network traffic investigation is finished by (Fadlullah et al., 2017).

Most usually utilized approaches are port based, signature based, and measurable fea-tures based network traffic distinguishing proof. Port based technique is the at first utilized strategy which depends on predefined explicit port numbers (Contact et al., 2013). The per-formance of port based strategy is extremely restricted in the present organization traffic distinguishing proof because of the explanation that the conventions are enlisted without the standard of port enrollments and the ports of conventions are dynamic in nature. From the year 2002 onwards, signature based technique was utilized. The marks are hexadecimal qualities or it tends to be an arrangement of strings which changes as per every application (Park et al., 2008). This is a straightforward strategy and the exhibition generally high when contrasted with the port based proto-col arrangement. The blunder rate is lesser than 10% and these techniques are more viable. At the point when detail of a convention changes or another one is planned, the time has come consuming to begin once again for tracking down significant marks. Profound parcel review (DPI) is another tech-nique utilized for network traffic grouping. nDPI is publically accessible which depends on DPI. This can recognize standard application successfully and creates issues with intriguing application and encoded applications. Measurable highlights and ML based arrangement is another strategy (Alshammari and Zincir-Heywood, 2007, 2008, 2009, 2011; Zuev and Moore, 2005). These are more well known lately. This depends on the measurable fea-tures in transmission of the traffic, for example, the time span between bundles, parcel size, rehashing design, etc. These highlights are then taken care of into old style ML calculations like Guileless Bayes (NB), DT, Backing vector machine (SVM), and Brain organization (NN) (Tan and Collie, 1997). These strategies can be executed continuously or close to constant, however the central issue is that the time and experience that we really want to choose

the proper elements which is to be taken care of into ML calculation. The assaults connected with port maltreatment, arbitrary port utilization, and burrowing is expanding quickly step by step. These assaults are not completely ready to destroy which causes huge security issues in networks. These frameworks require broad space information and all the more critically they are not exact.
Lately, payload bytes data is passed into DL models (Wang, 2015). (Wang, 2015) showed the initial thousand bytes is adequate to recognize conventions successfully. In this writing, they gathered TCP stream information from inner organization and the full payload is removed from each bundle. Then, at that point, they joined the payload bytes for each TCP meeting. A byte is addressed by a number from 0 to 255. It was then standardized [0, 1] scale. The length of every payload grouping was 1,000. They got around 0.3 million records in the wake of pruning information for tests. The quantity of convention types in the information taken was 58. The completely extricated payloads of the parcels gathered were made into a picture design. This actually intends that, every byte was addressed as a pixel. DL has performed well in PC vision assignments and utilized something very similar to organize traffic examination (Wang, 2015). (Wang, 2015) proposes a technique for highlight extraction by utilizing stacked AE. The principal benefit of utilizing AE is that we can take care of both regulated and Spam and Phishing Exercises Information Examination solo information; obviously the marked information will give more exact highlights. Alongside that, the AE technique will accomplish the objective of dimensionality decrease moreover. Here the highlights are planned to new space and the repetitive data is likewise separated. This examination work contrasts from the past works in the accompanying focuses: (1) similar investigation of different DL designs and traditional ML calculations are completed for SSH traffic investigation (2) The mix of AE and DL structures system is proposed for application network traffic characterization, pernicious traffic arrangement, and noxious traffic location Spamming and phishing assaults are the most well-known security challenges we face in the present digital world. The current strategies for the spam and phishing recognition depend on boycotting and heuristics method. These techniques require human intercession to refresh assuming any new spam and phishing action happens. Additionally, these are totally wasteful in identifying new spam and phishing exercises. These methods can identify malevolent movement solely after the assault has happened. As of late, ML strategies have been utilized for spam and phishing movement discovery utilizing electronic mail (email) and uniform asset finder (URL) information investigation. These methodologies have the capacity to distinguish new sorts of spam and phishing assaults. The nitty gritty review of spam and phishing movement location utilizing email and URL information investigation is talked about exhaustively by (Almomani et al., 2013; Khonji et al., 2013; Mujtaba et al., 2017; Sahoo et al., 2017). Due to a large portion of the information of spam and phishing movement as messages, in this work the adequacy of different regular language handling (NLP) message portrayal techniques are planned to spam and phishing exercises portrayal to get familiar with the semantic, primary, and syntactic highlights naturally and following DL structures for ideal element extraction and grouping.

Malware Information Investigation

Vindictive programming or malware keeps on representing a significant security worry in this computerized age as PC clients, partnerships, and state run administrations witness a dramatic development in malware assaults. Current malware identification arrangements take on static and dynamic investigation of malware marks and ways of behaving that are tedious and ineffectual in recognizing obscure malware. Ongoing malware utilizes polymorphic, transformative, and other shifty procedures to change the malware ways of behaving rapidly and to produce countless malware. Since new malware is overwhelmingly variations of existing malware, ML calculations are being utilized as of late to lead a viable malware investigation. As of late, the itemized overview on different ML strategies for malware examination is finished by (Ucci et al., 2018).

Malware Characterization utilizing Static Investigation

A few security specialists have applied area level information on convenient exe-cutable (PE) for static malware recognition. As of now, examination of byte N-grams and strings are the two most usually involved strategies for static malware discovery without space level information. Notwithstanding, the N-gram approach is computationally costly and the exhibition is significantly exceptionally low (Raff et al., 2018b). It is frequently hard to apply area level information to extricate the important elements while building a ML

Malware Information Investigation model to recognize the malware and harmless records. This is because of the way that the Windows working framework (operating system) doesn't reliably force its own determinations and principles (Raff et al., 2017). Because of continually changing determinations and stan-dards occasionally, the malware recognition framework warrants corrections to meet future security necessities. To address this, (Anderson and Roth, 2018) has applied ML al-gorithms with the highlights got from parsed data of PE document. They embraced arranging of freethinker elements, for example, crude byte histogram and byte entropy histogram which was taken from (Saxe and Berlin, 2015), and moreover utilized string extraction. MalConv (Raff et al., 2018a) contrasted these old style ML models and the DL approach. They have made the informational index with highlights as well as crude documents and the related code openly accessible since DL models require more investigation and require further examination.
An old style completely associated network and RNN model of DL was generally utilized to distinguish malware with 300 bytes data from the PE header document (Raff et al., 2017). Consequently, (Krˇc'al et al., 2018) has utilized convolutional brain organization (CNN) on countless byte long executables and got predictable outcomes across 2 unique tests in light of a past report (Raff et al., 2018b). Utilizing space level information, (Anderson and Roth, 2018) has separated a few highlights and showed that its exhibition is practically identical to the MalConv (Raff et al., 2018a) DL approach. The presentation of Malconv model was improved by making alteration to the current design (Krˇc'al et al., 2018). This exploration work accepts that the DL capacities have not been completely understood and this work proposes the utilization of Windows-Static-Cerebrum Droid (WSBD) model for

consolidating DL.
In this work, by utilizing WSBD, this examination work assesses the exhibition of seat checked models (Anderson and Roth, 2018), (Raff et al., 2018a), and (Krˇc'al et al., 2018) on the freely accessible informational index from (Anderson and Roth, 2018) alongside secretly gathered examples of harmless and malware. This exploration work presents a few variations of the current DL models from (Raff et al., 2018a) and (Krˇc'al et al., 2018). In advertisement dition, this examination work looks at the presentation of different old style ML classifiers on the space level elements got from (Anderson and Roth, 2018) utilizing DL strategies.

Malware Order utilizing Dynamic Investigation

Malware examination strategies in light of dynamic investigation are more strong to confusion techniques when contrasted with static examination. In (Tobiyama et al., 2016), highlights from 5 minutes Programming interface calls were separated and given to CNN for order utilizing dynamic examination. They utilized around 170 examples and got 0.96 for region under the bend (AUC) as the quality measure. In (Huang and Stirs up, 2016), shallow feedforward brain organization (FFN) with highlight sets of Programming interface calls were gotten from countless examples of harmless and malware that were gathered secretly. It performs well when contrasted with the current methodologies yet it comes up short on concentrate on the speed of execution which is significant for continuous organizations. In (Pascanu et al., 2015), explores different avenues regarding reverberation state organizations (ESNs) and RNN were led to get familiar with the language of malware. In the majority of the examinations, the ESNs performed well in contrast with RNN. In (Shibahara et al., 2016), tests were led to decide when to stop the malware execution concerning network correspondence. This technique has diminished the all out time taken by 67% contrasted and ordinary strategies. In (Kolosnjaji et al., 2016), the utilization of RNN and its variation long transient memory (LSTM) and CNN were utilized for malware order with Programming interface call long successions as highlights. The serious issue with the current strategies are that they require some investment to examine the ways of behaving during execution. In (Kolosnjaji et al., 2016), a mixture of CNN and RNN was utilized for malware order utilizing framework call arrangements. These framework calls were acquired from dynamic investigation and their technique was accounted for to outflank recently utilized calculations, for example, SVM and secret markov model (Well). Be that as it may, this exploration work recognizes the principal disadvantage as the inability to talk about the significance of execution time towards discovery of malware progressively. (Rhode et al., 2018), has proposed a technique in light of RNN with two unique informational indexes. They additionally assessed the presentation of other traditional ML classifiers. They had announced 94% precision with 5s execution time.
Many examination studies have looked at malware location procedures in view of static, dynamic, and crossover investigation. In (Damodaran et al., 2017), utilization of Well on both static and dynamic examination of capabilities and a near report on identification rates was led over a significant number of malware families. They detailed that unique examination by and large yielded the best identification rates. This examination work proposes Windows-Dynamic-Mind Droid (WDBD) model that assesses the viability of the different traditional ML calculations and DL designs to know which calculation is generally fitting for Windows malware arrangement. This exploration work utilizes two distinct informational collections that contain various quantities of malware and harmless examples that are caught during dif-ferent execution time.

Malware assaults are on the ascent and lately, new malware is effectively created as variations of existing malware from a known malware family. To defeat this issue, it is critical to gain proficiency with the comparative attributes of malware that can assist with characterizing it into its loved ones. A few examinations directed in (Nataraj, 2015) enjoy taken benefit of the way that most malware variations are comparable in structure, with advanced sign, and picture handling strategies utilized for malware order. They have changed the malware doubles into grayscale pictures and revealed that malware from the equivalent malware family is by all accounts very comparable in format and surface. Since picture handling methods require neither dismantling nor code execution, it is quicker in contrast with the static and dynamic investigation. The primary benefit of such a methodology is that it can deal with pressed malware and can chip away at different malware regardless of the operating system. Trial results have shown 98% grouping precision on an enormous malware data set and it is additionally strong to famous muddling strategies specifically, encryption. They have made benchmarked information, Malimg as open for additional examination.
Lately, the Malimg informational index is utilized to assess the adequacy of cutting edge ML calculations over old style ML calculations. Rather than following different sign and picture handling methods, the utilizations of DL calculations are changed into malware order utilizing Malimg informational collection (Agarap and Pepito, 2017; Rezende et al., 2018). In (Agarap and Pepito, 2017), they have applied the blend of SVM and DL architec-tures like CNN and RNN varieties. They have separated the informational collection arbitrarily into 80% for preparing and 20% for testing and guaranteed that the blend of GRU and SVM performed well in contrast with different strategies. As of late, (Rezende et al., 2018) did the point by point examination of various CNN models like ResNet-50, VGG16, VGG-19, and the exchange learning applied on both the Malimg and secretly gathered informational collection. This work proposes DeepImageMalDetect (DIMD) that use DL with picture process-ing approach for malware classification. The presentation of the proposed design is contrasted and the other DL models and old style ML classifiers. This multitude of techniques are assessed on the benchmark informational index and furthermore the presentation of those strategies are displayed on the as of late gathered private malware tests. As the utilization of cell phones flood past the PCs (Pc's), the malware authors additionally went with the same pattern, concentrating making malware for the cell phones. The long known security components of Pc's, like interruption discovery framework (IDS), firewalls, encryption antivirus, and other endpoint put together safety efforts with respect to Android cell phones are simply starting to be presented. This adds to the simplicity of spreading malware on the cell phones than the Pc's. As of late, applying ML ways to deal with the examination and classification of Android malware has turned into a functioning area of exploration. As of late, the definite overview of different ML strategies for Android malware investigation is finished by (Suarez-Tangil et al., 2014).

In the Android applications, a record named AndroidManifest.xml characterizes every one of the consents and the Programming interface calls made by the comparing application (Oberheide and Mill operator, 2012). Most work in static examination depends on AndroidManifest.xml record for elements, for example, authorization and Programming interface calls. Notwithstanding these two elements, other meta-information highlights, for example, adaptation name, form number, distributer name and so on have been utilized (Urcuqui-L'opez and Cadavid, 2016). A static examination system has been presented for the identification of pernicious ways of behaving by involving authorizations as elements with six ML classifiers (Urcuqui-L'opez and Cadavid, 2016). To perceive how we can improve the Android malware discovery rate, this exploration works has utilized similar informational index in this work.

Most work in unique examination of Android malware utilized framework call occasions at portion level as highlights. For instance, the techniques (Isohara et al., 2011; Xie et al., 2010) guaranteed that the framework calls are more powerful with the investigations that dissected 230 applications utilizing a model. Moreover, the model framework made more progress in de-tecting the vindictive examples of obscure applications. As of late, (Egele et al., 2012) studied the dynamic malware investigation strategies utilized in the previous years. (Bente et al., 2012), proposed correlative procedure called as Reliable, Setting related Inconsistency De-tection for Cell phones (TCADS) that utilized comparable arrangement of highlights of Andromaly with the trust and setting related data. However TCADS itself showed a sensible compelling security instrument for Android malevolent applications, it totally bombed when it was tried on genuine climate. (Kim et al., 2012), involved comparative elements as of Andromaly to distinguish the vindictive exercises in cloud foundation of cell phones. (Dini et al., 2012), utilized a blend of both the framework calls and comparative highlights of Andromaly.

All of the previously mentioned investigations were done exclusively on few harmless and malware applications. (Amos et al., 2013) had shown enormous scope investigation of ML classifiers on huge number of genuine applications. What's more, the creators proposed Framework for Consequently Preparing and Assessing Android Malware Classifiers (STREAM) climate for col-lection of component vector. The bundle runs Android investigate span (adb) scripts in Linux operating system to lead robotized examination of harmless and malware applications. To proceed with additional re-search, STREAM bundle gives informational collection (both preparation and testing highlight vectors) in Quality Connection Document Arrangement (ARFF) design. Utilizing WEKA device, the creators had the option to accomplish exactness in the reach 70-81.25% with the static ML classifiers specifically, RF, NB, Multi-facet Perceptron (MLP), Bayes net, LR, and J48. They revealed 81.25% malware discovery rate by utilizing Bayes net. This work expands the examination work of (Amos et al., 2013) by utilizing LSTM.

Towards DL based Android malware discovery, Droid-Sec (Yuan et al., 2014) device was presented. Creators announced DL techniques are appropriate for Android malware classifica-tion in contrast with static ML classifier with complete 202 highlights from 3 classifications like consent, touchy Programming interface, and dynamic way of behaving in static and dynamic examination. The outcomes showed 96.5% exactness involving DNN in genuine Android malware identification task. Fur-ther, they proposed Droid Detector (Yuan et al., 2016) in view of the past examination and utilized static ML and DL classifier with the harmless informational collection from 20,000 harmless applications and 1,760 malware applications from Contagio People group and Genome Task. It accomplished 96.76% as the most elevated exactness involving DL with 192 highlights in static and dynamic investigation. In (McLaughlin et al., 2017), study extricated opcode succession portrayal from .apk records and passed to the implanting layer to become familiar with the semantic data among them and again passed into DL layers to gain proficiency with the ideal component portrayal. Following, in this work, the adequacy of DL and ML calculations are read up for Android malware location.

References

1. Vinayakumar, R., Soman, K. P., Poornachandran, P., & Sachin Kumar, S. (2018). Evaluating deep learning approaches to characterize and classify the DGAs at scale. Journal of Intelligent & Fuzzy Systems, 34(3), 1265-1276.
2. Vinayakumar, R., Soman, K. P., Poornachandran, P., & Sachin Kumar, S. (2018). Evaluating deep learning approaches to characterize and classify the DGAs at scale. Journal of Intelligent & Fuzzy Systems, 34(3), 1265-1276.
3. Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2018). Evaluating deep learning approaches to characterize and classify malicious URLs. Journal of Intelligent & Fuzzy Systems, 34(3), 1333-1343.
4. Vinayakumar, R., Soman, K. P., Poornachandran, P., & Sachin Kumar, S. (2018). Detecting Android malware using long short-term memory (LSTM). Journal of Intelligent & Fuzzy Systems, 34(3), 1277-1288.
5. Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2017). Evaluation of recurrent neural network and its variants for intrusion detection system (ids). International Journal of Information System Modeling and Design (IJISMD), 8(3), 43-63.
6. Vinayakumar, R., & Soman, K. P. (2018). DeepMalNet: Evaluating shallow and deep networks for static PE malware detection. ICT Express, 4(4), 255-258. 303 List of Publications based on the research work
7. Vinayakumar, R., Soman, K. P., Poornachandran, P., Mohan, V. S., & Kumar, A. D. (2019). ScaleNet: Scalable and Hybrid Framework for Cyber Threat Situational Awareness Based on DNS, URL, and Email Data Analysis. Journal of Cyber Security and Mobility, 8(2), 189-240.
8. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., & Venkatraman, S. (2019). Robust Intelligent Malware Detection Using Deep Learning. IEEE Access, 7, 46717-46738.
9. Vinayakumar, R., Alazab, M., Soman, K. P., Poornachandran, P., Al-Nemrat, A., & Venkatraman, S. (2019). Deep Learning Approach for Intelligent Intrusion Detection System. IEEE Access, 7, 41525-41550.
10. Vinayakumar, R., Soman, K. P., & Poornachandran, P. (2019). A Comparative Analysis of Deep Learning Approaches for Network Intrusion Detection Systems (N-IDSs): Deep Learning for N-IDSs. International Journal of Digital Crime and Forensics (IJDCF), 11(3), 65-89.

11. Vinayakumar, R., & Soman, K. P. (2019). Siamese neural network architecture for homoglyph attacks detection. ICT Express.
12. Vinayakumar R, Soman KP, Simran K, Prabaharan Poornachandran, Mamoun Alazab , Deep Learning and Visualization for Botnet Detection in the Internet of Things of Smart Cities, Journal of Future Generation Computer Systems. (Accepted)
13. R, Soman KP, Sriram S, Prabaharan Poornachandran, and Mamoun Alazab, Deep Learning Based Two-Level Framework for Domain Name Systems 304 List of Publications based on the research work Data Analysis, IEEE Transactions on Information Forensics and Security. (under major revision)
14. Vinayakumar R, Soman KP, Prabaharan Poornachandran, and Mamoun Alazab, Malicious URL Detection using Deep Learning, IEEE Transactions on Emerging Topics in Computational Intelligence. (under major revision)
15. Vinayakumar R, Soman KP, Sriram S, Prabaharan Poornachandran, Mamoun Alazab, Simran K, A Comprehensive Tutorial and Survey of Applications of Deep Learning for Cyber Security, IEEE Communications Surveys and Tutorials. (under major revision)
16. Vinayakumar R Soman KP, Prabaharan Poornachandran, Mamoun Alazab, Muhammad Ajmal Azad and Ala' M. Al-Zoubi, Spam Emails Detection based on Deep Learning. (ready for submission)
17. In Proceedings 13th Annual Computer Security Applications Conference, pages 99–107. IEEE, 1997.
18. Shun Tobiyama, Yukiko Yamaguchi, Hajime Shimada, Tomonori Ikuse, and Takeshi Yagi. Malware detection with deep neural network using process behavior. In 2016 IEEE 40th Annual Computer Software and Applications Conference (COMPSAC), volume 2, pages 577–582. IEEE, 2016.
19. Joe Touch, M Kojo, E Lear, A Mankin, K Ono, M Stiemerling, and L Eggert. Service name and transport protocol port number registry. The Internet Assigned Numbers Authority (IANA), 2013.
20. Daniele Ucci, Leonardo Aniello, and Roberto Baldoni. Survey of machine learning techniques for malware analysis. Computers & Security, 2018.
21. Christian Urcuqui-L´opez and Andr´es Navarro Cadavid. Framework for malware analysis in android. Sistemas & Telem´atica, 14(37):45–56, 2016.
22. Sitalakshmi Venkatraman and Mamoun Alazab. Use of data visualisation for zero-day malware detection. Security and Communication Networks, 2018, 2018.
23. Prabaharan Poornachandran Mamoun Alazab Vinayakumar R, Soman KP and Sabu M. Thampi. Amritadga: A comprehensive data set for domain generation algorithms (dgas). In Big Data Recommender Systems: Recent Trends and Advances, Institution of Engineering and Technology (IET), 2019.
24. Jiannan Wang, Guoliang Li, and Jianhua Fe. Fast-join: An efficient method for fuzzy to300 BIBLIOGRAPHY ken matching based string similarity join. In 2011 IEEE 27th International Conference on Data Engineering, pages 458–469. IEEE, 2011.
25. Zhanyi Wang. The applications of deep learning on traffic identification. BlackHat USA, 24, 2015. Jonathan Woodbridge, Hyrum S Anderson, Anjum Ahuja, and Daniel Grant. Detecting homoglyph attacks with a siamese neural network. In 2018 IEEE Security and Privacy Workshops (SPW), pages 22–28. IEEE, 2018.
26. Liang Xie, Xinwen Zhang, Jean-Pierre Seifert, and Sencun Zhu. pbmds: a behaviorbased malware detection system for cellphone devices. In Proceedings of the third ACM conference on Wireless network security, pages 37–48. ACM, 2010.
27. Yang Xin, Lingshuang Kong, Zhi Liu, Yuling Chen, Yanmiao Li, Hongliang Zhu, Mingcheng Gao, Haixia Hou, and Chunhua Wang. Machine learning and deep learning methods for cybersecurity. IEEE Access, 6:35365–35381, 2018.
28. Sandeep Yadav, Ashwath Kumar Krishna Reddy, AL Reddy, and Supranamaya Ranjan. Detecting algorithmically generated malicious domain names. In Proceedings of the 10th ACM SIGCOMM conference on Internet measurement, pages 48–61. ACM, 2010.
29. Sandeep Yadav, Ashwath Kumar Krishna Reddy, AL Narasimha Reddy, and Supranamaya Ranjan. Detecting algorithmically generated domain-flux attacks with dns traffic analysis. IEEE/Acm Transactions on Networking, 20(5):1663–1677, 2012.
30. Chuanlong Yin, Yuefei Zhu, Jinlong Fei, and Xinzheng He. A deep learning approach for intrusion detection using recurrent neural networks. Ieee Access, 5:21954–21961, 2017. 301 List of Publications based on the research work Bin Yu, Jie Pan, Daniel Gray, Jiaming Hu, Chhaya Choudhary, Anderson Nascimento, and Martine De Cock. Weakly supervised deep learning for the detection of domain generation algorithms. IEEE Access, 2019.
31. Zhenlong Yuan, Yongqiang Lu, Zhaoguo Wang, and Yibo Xue. Droid-sec: deep learning in android malware detection. In ACM SIGCOMM Computer Communication Review, volume 44, pages 371–372. ACM, 2014.
32. Zhenlong Yuan, Yongqiang Lu, and Yibo Xue. Droiddetector: android malware characterization and detection using deep learning. Tsinghua Science and Technology, 21(1): 114–123, 2016.
33. Yury Zhauniarovich, Issa Khalil, Ting Yu, and Marc Dacier. A survey on malicious domains detection through dns data analysis. ACM Computing Surveys (CSUR), 51 (4):67, 2018.
34. Denis Zuev and Andrew W Moore. Traffic classification using a statistical approach. In International workshop on passive and active network measurement, pages 321–324. Springer, 2005.