Filtering And Forwarding Approach: For Feature Reduction And Classification For Anomaly Detection In IoT Environment

Suresh B, Venkatchalam M and Saroja M

Abstract: Internet of Things (IoT) is rapidly evolving concept with some ability to change physical interaction amongst organizations and individuals. IoT attempts to swap “things” in reliable and secure manner. IoT has acquired application in numerous fields like learning, training, healthcare, resource management, information processing and so on. Moreover, real time implementation of IoT technology is to fulfill numerous privacy and security that are to be alleviated for large scale deployment. Prevention approach is anticipated to improve IoT device security and network against DoS attack which takes huge bandwidth in present IoT devices. As networks are self-configuring and wireless and it does not require pre-existing infrastructure and possess changeable node movements, security turns to be a most crucial crisis to be resolved. The anticipated model is sourced on investigations and analysis of bandwidth attacks that significantly concentrates on DoS which is considered to be significant challenge and complexity to identify; as well it reduces network performance. DoS comprise of set of malicious/vulnerable nodes and target certain node to stop users from accessing resources and services. Intrusion prevention devices are approaches considered as Add-ons to aggressively identify and avoid it for IDS detection process. Here, 3-tier (3T) intrusion prevention model, that is, at first, features dimensionality has to be reduced. Subsequently, Linear Discriminate Analysis (LDA) is for linear computation of intrusion features in order to carry out effectual classification. Thirdly, Linear filtering and forwarding (LFF) is to block unnecessary data that influences IoT devices. Simulation is carried out in MATLAB environment, performance metrics like accuracy, sensitivity, specificity, recall is computed for evaluation.

Index Terms: IoT, intrusion, Feature based dimensionality reduction, Linear filtering and forwarding, Linear computation

1. INTRODUCTION

In Recent times, Internet of Things (IoT) is emerging as a growing trend in internet that offers exchanging of services in global environment [1]. IoT is usually application domain that incorporates social arenas and technologies [2]. Various investigators consider IoT as “Network things, in which everyone is embedded with wireless sensors that are connected via world wide web” [3]. The preliminary attempt is to guarantee different things that can be operated and connected as it may interact with users. It is dynamic IT infrastructure with self-configuring competency for determining communication amongst virtual and physical identities via intelligent interfaces [4]. It assists continuous data exchange that triggers action as per real-world events automatically [5]. Significant confront encountered by IoT evolution is not its scalability however its security [6]. As known in prior, conventional networks are extremely secured than wireless counterparts. Traditional networks facilitate traffic to move diverse routing devices such as gateways, switches that are secured frequently with highly configured firewalls and numerous security approaches [7]. Therefore, networks are extremely equipped with Denial of Service (DoS) attacks or other attacks [8]. However, IoT are also considered as P2P that are wireless generally, and essentially susceptible to diverse kinds of attacks. Traditional protocols of wired networks are not appropriate to be executable [9], where topology varies recurrently; communication links amongst nodes are wireless and no centralized network control [10]. Therefore, it is essential for every node to integrate security method to eliminate attacks.

A. IDS Construction for IoT Devices

IoT comprises of wireless nodes that communicate via links [11]. There are some restrictions encountered in this type of networks like bandwidth constraint, battery life, security and so on. Security is measured as a significant factor in IoT [12]. With wireless communication, dynamically changing topologies are susceptible to various threats like eavesdropping, node compromise, DoS. Identifying these sorts of attacks are confronting task [13]. IDS are an appropriate method for recognizing these kinds of attacks in IoT. IDS are a wider that consistently monitors activities and makes suitable action when required [14]. In accordance to data collection and detection methods, IDS are partitioned to following categories; i) Anomaly ii) signature iii) specification. In last IDS, priori knowledge is utilized to identify attacks on networks. Disadvantage in these kinds of strategies are it cannot be used for unknown attacks. In former IDS, generally, system behaviour has to be observed, if changes from normal operation by certain threshold, it is identified. In specification, some limitations have to be fixed for protocols or operations. IDS observe functioning based on constraints.

Fig 1 IDS Prevention Architecture (Source 1)
B. Modeling IDS
The present IoT based IDS architecture comprises of three classifications: i) co-operative ii) stand-alone and iii) hierarchical. While in second architecture, each node is accountable for security of collaboration with nodes remaining in network. Subsequently, co-operative model have own systems [15]. It determines network intrusion co-operatively by sharing parameters and information. Finally hierarchical model is partitioned into clusters and nodes are chosen sourced on parameters like CHs accountability and in carrying out IDS. Significant advantages of these model is sufficient resource utilization, however has numerous demerits of choosing CH that are unfeasible in AD-HOC, where nodes progress in all directions. In this investigation, three tier (3T) approach is modeled for prevention of intrusion in IoT devices. Initially, features dimensionality has to be reduced. Subsequently, Linear Discriminate Analysis (LDA) is for linear computation of intrusion features in order to carry out effectual classification. Thirdly, Linear filtering and forwarding (LFF) is to block unnecessary data that influences IoT devices. Simulation is carried out in MATLAB environment. Rest is modeled as: Section II explains in detail about existing detection approach based on intrusion detection. Section III explains about three tier (3T) model for intrusion prevention, that specifically includes feature based dimensionality reduction, Linear Discriminate Analysis (LDA) and Linear filtering and forwarding (LFF). Section IV depicts numerical outcomes and discussion of anticipated model. Section V demonstrates conclusion and future direction of anticipated model.

2 RELATED WORKS
In this section, prevailing intrusion detection and prevention model usually utilizes statistical models like Bayes theory, Hidden Markov Model (HMM), signal processing, cluster analysis and distance measure to identify anomalies. It is extensively classified into unsupervised or supervised learning. In former, normal characteristics of networks or systems are designed with labeled dataset. Unsupervised approach considers characteristics of recurrent, and therefore, it is where no training data is needed. In [16], Casa et al. anticipated unsupervised NIDS sourced on sub-space clustering identification and illustration this model carries out effectually over unknown attacks. In [17], feature selection filter is anticipated, which uses FDR and PCA are to filter noises. In this method, SOMs neural approach is utilized to filter general activities. Moreover, it possess high FPA. Sheikhkan and Bostani et al. [18] anticipated an unsupervised architecture sourced on K-Means clustering and Optimum-path forest algorithm approach. This architecture designs normal and malicious characteristics of networks. In [19], Guo et al, anticipated two level IDS that initially identifies misuse and utilize KNN to diminish FAR. Toosi [20], anticipated attack classifier approach to execute mixture of Fuzzy NN, GA and FI model for intrusion detection. Indeed of higher accuracy rate while detecting normal characteristics and identifying simpler attacks like DoS attacks, it carries out ineffectual in identifying distribution attacks and low frequency like R2L. In [21], Horng et al, anticipated multi-classification attack of BRICH clustering model and support vector machines (SVM) to haul out appropriate attributes from KDD’99 dataset. This anticipated model has superior DR for probe and DoS is insufficient against R2L and U2R. In [22], Tan anticipated an approach for detection with MCA to enhance traffic accuracy. In [23], author depicts 2L classification to identify R2L and U2R with low complexity while optimizing feature diminishment. In [24], Osanaive et al depicted an ensemble sourced Multi-filter feature selection technique to identify DoS in cloud environment with filtering to acquire optimum selection in NSL-KDD. In [25], Iqbal illustrated attack taxonomy for cloud and recommended cloud sourced IDS. IDS have cast of for dealing security risks in industrial systems. For instance, Pan et al [26] demonstrated an automated and systematic model to construct hybrid IDS that shows temporal state sourced specification for electric PS to appropriately distinguish disturbances, cyber attacks and normal operations. In [27], Zhou et al, provided multi model driven IDS and an industrial anomaly based IDS sourced in HMM to filter attacks for suitable faults.

3 PROPOSED METHOD
The proposed method consists of classification module and dimensionality reduction which is discusses below:

A. Dimensionality Reduction
Here, dimensionality reduction module is depicted to resolve certain limitations as dimensionality may leads to wrong decisions while increasing computational complexity of classifiers. So as to handle higher dimensionality issue, supervised and unsupervised dimension reduction techniques are engaged. They are: LDA and PCA. Both approaches are deployed here for feature extraction and feature selection as in Eq. (1):

\[
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{bmatrix}
\xrightarrow{\text{feature extraction}}
\begin{bmatrix}
Y_1 \\
Y_2 \\
\vdots \\
Y_m
\end{bmatrix} =
\begin{bmatrix}
\mathbf{z}_1 \\
\mathbf{z}_2 \\
\vdots \\
\mathbf{z}_M
\end{bmatrix}
\begin{bmatrix}
X_1 \\
X_2 \\
\vdots \\
X_n
\end{bmatrix}
\] (1)

1) Feature selection: selection of subset of all features with respect to its effectiveness for superior classification

2) Feature Extraction: generation of new feature subset by merging prevailing features

In 3-Tier intrusion prevention, we use feature extraction approach for NSL-KDD mapping which comprises 41 features with lower feature space by eliminating lesser features. Generally, these approaches are restricted to transform Eq. (2):

\[
y = Wx
\] (2)

Let ‘x’ be ‘N’ random vector in NSL-KDD, and feature space comprises of M dimensional features, that is, M specifies new features that are transformed, here M < N. Matrix co-variance as in Eq. (3):

\[
\Sigma_x = \sum_{k=1}^{n}(x_k - m)(x_k - m)^T
\] (3)

Here, ‘m’ is mean vector as in Eq. (4):

\[
m = \frac{1}{n} \sum_{k=1}^{n} x_k
\] (4)

Vector decomposition is depicted as in Eq. (5):
\[ \sum \nu = \lambda \nu \]  
(5)

Where, \( \nu \) is Eigen vector and \( \lambda \) is Eigen value

Here, PCA is used to sort Eigen vectors in descending order. Subsequently, Eigen-vectors that have lower Eigen values holds least information regarding data distribution and these least information Eigen vectors will be eliminated. A general model is to rank Eigen vectors from high-low Eigen-value and selects top most \( k \) Eigen vectors sourced on Eigen values. Here in 3-tier intrusion prevention model, Eigen values that are more significant will be determined. Therefore, feature based matrix \( \mathbf{x}' \) utilizes linear transformation for testing and training dataset. In dimensionality reduction phase, Error Function is measured as performance analysis metrics. This EF is used for principal selection as depicted in Eq. (7), where \( p, c \) are principal components. \( p \) and \( c \) is for data representation and number of dimensions specifically, \( \lambda \) and \( N \) specifies Eigen values and number of samples respectively as in Eq. (6):

\[
\text{Error function} = \left[ \frac{1}{nm(m-1)} \sum_{i=1}^{m} \lambda_i \right]^{\frac{1}{2}}
\]

(7)

Cross validation is carried out to compute optimum principals with reduced errors, where selection criteria helps in reducing features and assists next level of dimensionality reduction module to evaluate spreadable objects and lower dimensionality matrix.

B. Linear Discriminant Analysis

It is used to speed up recognition of intrusion in a system. As samples in PCA are not specific, anticipated model uses DR to use labeled data in dimensionality transformation. LDA evaluates labels to diminish of huge datasets and LDA is extensively utilized in diverse domains like stock analysis and image processing. Transformation after LDA, mapped features shows four dimensions. Subsequently, dataset is transformed to \( c-1 \) dimensions, where \( 'C' \) is lass labels over dataset. Optimality is used to map higher to lower dimensional space when essential information is needed for classification.

There exist two scattered matrices that are attained from LDA, such as: \( S_B \) and \( S_W \) that is between class scatter matrix and within class scatter matrix. In 3-tier intrusion prevention model, LDA dimensionality reduction model moves dataset to least dimension. There are set of \( n_i \) vectors of \( x_{i1}, x_{i2}, \ldots, x_{in} \) that belongs to \( 'k' \) diverse class labels of \( C_i \) where \( i = 1, 2, 3, \ldots, k \) has samples \( 'n_i' \). For instance in NSL-KDD \( k = 5; \) that is, normal, Probe, DoS, L2R and U2R

Projection matrix \( 'W' \) is computed to maximize and reduce class scatter matrix in Eq. (8) and Eq. (9):

\[
S_B = \sum_{i} (\mu_i - \bar{x})(\mu_i - \bar{x})^T
\]

(8)

\[
S_W = \sum_{i} \sum_{i \in C_i} (x_i - \mu_i)(x_i - \mu_i)^T
\]

(9)

Where, \( \mu_i \) is mean value of class \( C_i \) samples as in Eq. (10):

\[
\mu_i = \frac{1}{n_i} \sum_{x \in C_i} x
\]

(10)

As ratio of \( J \) in Eq. (11), is in range \( S_B \) and \( S_W \), it is minimized as optimization crisis with projection matrix:

\[
J = \frac{W^T S_B W}{W^T S_W W}
\]

(11)

From these operations is conducted on training dataset to acquire transformation matrix to unknown instances or test sets.

a. Filter and Forward

C. Filter and Forward

The objective of preliminary filtering process is to recognize incoming packets that does not hold any intrusions and in case of any intrusions identified, it will be filtered out immediately. This process prevents system from unnecessary transmission to destination. Preliminary filtering process reduces load over destination and helps to enhance system performance, as process of transmitting filtered out packets from splitter to destination. To carry out filtering, default snort rule set has to be analyzed and it is found with 165 rules need only processing. This rule set is determined as error function rule set. If error function rule set is recognized, splitter functions as: When packets are attained, it is validated against error function. If no rule is recognized and it comprises no payload, packet has to be filtered as in fig 2. Else, it is forwarded to sensors for execution. It is forwarded to sensors of two classes: these classes are matched with one rule from error function or else it will not match with rules though it has payload. Packets are forwarded to sensors, when packet comes under second class, examine all snort rule set.

![Fig 2 Flow diagram of proposed model](image_url)
It is noted that certain instances of earlier filtering to systems are not configured to carry out statefull-system inspection. It needs entire TCP handshake before checking rules. If transmitted acknowledgement is dropped before filtering, then inspection is not performed. As well, TCP re-assembling will not function if control packets are dropped, (for instance, FIN flags, packets with SYN or ACK). Therefore, certain form of filtering has to be applied; some intrusion detection features will not function properly. To resolve this crisis, TCP reassembling has to be done over splitter which is more feasible.

D. Packet Distribution
The ultimate objective of load (packet) distribution is partitioned into network traffic between sensors to maintain them eventually loaded. In the mean time, distribution of network traffic has to guarantee that packets over network are analyzed by similar sensor, else system will loss an attack. For instance, an attack located in packet boundaries. If packets are transmitted to various sensors, attack will not be recognized. Moreover, pre-processing elements like TCP re-assembly needs complete flow to work correctly.

E. Buffering
Network buffering is an approach for using packet stream that accelerates node processing by enhancing buffering of memory accesses and therefore diminishing cache misses. Network buffering is sourced in subsequent observation. Every packet arrives in sensor is validated in snort and EM rules of application layer. For instance, packets provided to server will be validated against rule set which verifies web based attacks. This rule set is constant in execution time. As well, packet that reaches FTP will be validated over rule set to determine FTP based vulnerabilities. While validating packet against rule set, every node has to bring rule set initially and validate cache level of processor. With incoming stream, packets from network flows seem to be inter-leaved. As an instance, consider node that senses traffic stream comprising packets to web and it belongs to FTP. Packets are interleaved with FTP, which shows node, may alter rule set, that outcomes in cache miss and diminished recital. In order to validate and to resolve some cases that are not undesirable to carry out packets classification to rule set that reflects NIDS rule set. This work concentrates on simpler method sourced on heuristics for describing target network buffering for packets given. Source and destination: A packet is placed in network buffer sourced on results of hashing evaluated on source/destination packet ports. It is expected that it has various flows that terminates in diverse buffers, henceforth reduces packet interleaving. Destination: A packet is placed in network buffer sourced on outcomes of hashing evaluated on destination. Static destination: Here, network locally subset is allocated for known traffic and uses this destination for excess packets and buffers. Network buffer receives web based traffic, other buffer receive NNTP and last receives P2P traffic. Uncategorized packets are allocated to network buffers using destination, i.e. destination or hashing. Traffic types are considered by profiling traffic and verify how NIDS is used.

F. Acknowledgement
This method is modeled as: Splitter has to communicate to determine actions that have to be carried out, i.e. forwarding/dropping packets. This is carried out by receiving acknowledgments from nodes to splitter. ACK is packet. It comprises of header, two bytes specifying packets to be acknowledged (factors), then four byte integer specifying packet identifiers. There are some probable formats that are essential with lesser bytes and assisting ACK factors for configuration. Moreover, it is adequate.

1) Positive acknowledgement: ACK of packet that are not concerning any intrusion.
2) Positive accumulation: ACK for some packet set that are not concerning any intrusion.
3) Negative acknowledgement: ACK of packet of other session.
4) Negative accumulation: ACK of packet in attack session.
5) Packet attained.

Sessions have benefits and de-merits. Packet receiving approach does not need splitter to hold packets temporarily in memory, however it lacks in performance. Negative acknowledgments has two significant de-merits. Initially, to distinguish packet has to be forwarded; some timeout value has to be used. Recall should not drop packet when attacks may be missed. As an outcome, to determine timeout in worst case, results in pointlessly superior latency. Subsequently unfeasible to splitter for differentiate where packet comprises no attack where packet may drop owing to certain error. Henceforth, positive acknowledgements seem to be more appropriate. Accumulative ACK and simple ACK are sourced on processing and latency.

b. Classification
In classification stage, 3-Tier intrusion prevention is trained already with NSL-KDD and categorizing traffic using multi-layer classifier to identify anomalies. Classifier selection is owing to capability in recognizing abnormal characteristics as given below:

1. Classifiers for allocating class labels;
2. Approaches like k-NN and NB [28 – 40].
3. Similarity measure for sample instances to deal dataset imbalance
4. Bucketing approach to fasten classification task.

Features transformed are evaluated with correlation coefficient parameter. This metrics depicts that relation amongst features (variables) by allocating number in [-1, 1] interval, where ‘1’ is positive correlation, ‘0’ is linear correlation, ‘-1’ is negative correlation. Correlation co-efficient evaluation of features demonstrates that features transferred at two layers for DR module are independent, as $p = 0$. These metrics specify no specific dependency amongst features. Certainty based similarity in classification is sourced on distributed classes in training dataset to resolve dataset issues. Certainty-Factor (CF) is specified as number ranges in [-1, 1] and denotes certainty for provided samples.

4 Experimental Setup
In this investigation, experimentation was carried out using MATLAB 2016a that operates of personal computer (PC) and 8 GB RAM. Three tier (3T) is trained using both training sets and then measured for performance metrics like Accuracy, Sensitivity, specificity, recall, F-measure and so on. 3T prevention model is adapted for removing intrusion in IoT environment and to enhance its performance.
TABLE I: 3-TIER DETECTION RATIO

<table>
<thead>
<tr>
<th>Methods</th>
<th>Normal</th>
<th>Probe</th>
<th>DoS</th>
<th>U2R</th>
<th>R2L</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Tier</td>
<td>94</td>
<td>85</td>
<td>86</td>
<td>70</td>
<td>34</td>
</tr>
<tr>
<td>SVM</td>
<td>94</td>
<td>77</td>
<td>82</td>
<td>67</td>
<td>19</td>
</tr>
<tr>
<td>ESC-IDS</td>
<td>95</td>
<td>96</td>
<td>85</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td>Association rule</td>
<td>96</td>
<td>97</td>
<td>84</td>
<td>31</td>
<td>38</td>
</tr>
<tr>
<td>MLR</td>
<td>97</td>
<td>94</td>
<td>74</td>
<td>79</td>
<td>34</td>
</tr>
<tr>
<td>3-Tier</td>
<td>98</td>
<td>78</td>
<td>92</td>
<td>29</td>
<td>32</td>
</tr>
</tbody>
</table>

TABLE II: DETECTION RATE OF PROPOSED MODEL

<table>
<thead>
<tr>
<th>Methods</th>
<th>Detection rate</th>
<th>False Alarm Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>2 Tier</td>
<td>84</td>
<td>5.5</td>
</tr>
<tr>
<td>SVM</td>
<td>83</td>
<td>4.8</td>
</tr>
<tr>
<td>ESC-IDS</td>
<td>76</td>
<td>-</td>
</tr>
<tr>
<td>Association rule</td>
<td>80</td>
<td>-</td>
</tr>
<tr>
<td>MLR</td>
<td>69</td>
<td>-</td>
</tr>
<tr>
<td>3-Tier</td>
<td>89</td>
<td>8.9</td>
</tr>
</tbody>
</table>

With NSL-KDD, attacks in original KDD’99 are eliminated. Even though, there is some crisis in NSL-KDD, it will not influence any application of dataset in this investigation while validating the findings. Every record comprises of connection with 41 attributes (for instance, flag, service and network connection) that are labeled as normal or 24 attack classes (for example, R2L, U2R, DoS and probe). It comprises two training and testing set with various distributions. As test set comprises of 17 types that are not in training, it evaluates efficiency of 3-Tier intrusion prevention for both unknown and known attacks.
Table I and II specifies detection ration and false alarm rate of proposed three tier and existing approaches like SVM, Association rule, MLR and so on. Graphical modeling of proposed model is depicted in Fig 3 and Fig 4. Fig 5, Fig 6, Fig 7 and Fig 8 shows some performance metrics of proposed model in accuracy, specificity, precision and recall. Anticipated model shows better trade off while comparing with approaches like RF, SVM, SGB.

Table III shows execution time for performing performance metrics like accuracy. Execution time of anticipated model is 23.65 secs which is lesser than other models.

5 CONCLUSION

With extensive adaptation of IoT and services in Internet connected and data related environment, guaranteeing IoT security in its infrastructure is essential to ensure consistency and stability. An attack on IoT environment can be handled effectually. For instance, compromise with IoT services in various application environments can effortlessly leads to major cause or even life threatening conditions. In this investigation, model with 3-tier classification and DR was anticipated. 3-tier approach is modeled to recognize intrusion detection, specifically in recognising low featured attacks (for instance, R2L and U2R) that probably damage the system resources. The anticipated model outperforms the prevailing models with respect to detection rate of common attacks and low frequency attacks. As 3-Tier prevention with Linear Discriminant analysis and principal component analysis for extraction techniques, this approaches helps in accurately classifying various attack types and behavior of normal attacks. Future extension includes, investing the potential of non-parametric approaches like fuzzy clustering and reduction models to attain superior classification over R2L, U2R and other sort of attacks. Another direction of research is that the anticipated model can be extended to identify intrusions in IoT like application and network layers.

REFERENCES

[12] Le, “Flexible network based intrusion detection and prevention system on software defined networks,” in Int conf on Advanced Computing and Applications,


