Deep Learning Approaches for IoT Security in the Big Data Era

K. S. Sunitha Krishnan and Sabu M. Thampi

He who would search for pearls must dive below

John Dryden

1 Introduction

The confluence of innovative technologies in wireless communications led to the evolution of the Internet of Things (IoT). According to recent studies, this cartel of things entrenched with electronic components, software, sensors, actuators coupled with the Internet, will increase to 50 billion by 2020. The giant stride in the number of IoT devices makes them the major genesis of data. IoT is triggering a massive influx of big data. To reap out the maximum efficacy of IoT, the massive amount of data is harnessed and converted to actionable insights utilizing the big data analytics. This makes the Internet of Things more intelligent than mere monitoring devices. Big data and IoT works well conjointly to offer analysis and insights. With the conjunction of the Internet of things, big data analytics shift the computing paradigm to the edges for real-time decision making.

A vast amount of data can be seen in various arenas like oil exploration, health care, social media information, power management etc. The organization and interpretation of these data are very useful in business at all levels. This data is unlayered and unstructured which cannot be used in machine learning algorithms which use

K. S. Sunitha Krishnan
Indian Institute of Information Technology and Management – Kerala (IIITM-K),
Kazhakkoottam, Kerala, India

Cochin University of Science and Technology, Kochi, Kerala, India
e-mail: sunithakrishnan.res17@iiitmk.ac.in

S. M. Thampi (✉)
Indian Institute of Information Technology and Management – Kerala (IIITM-K),
Trivandrum, Kerala, India
e-mail: sabu.thampi@iiitmk.ac.in

Z. Md. Fadlullah, A.-S. Khan Pathan (eds.), *Combating Security Challenges in the Age of Big Data*, Advanced Sciences and Technologies for Security Applications, https://doi.org/10.1007/978-3-030-35642-2_6

supervised instructions. Deep learning can avoid this drawback since they excel in label-less unsupervised learning especially when it comes to prediction and pattern recognition. Deep Learning algorithms create a layered, and hierarchical architecture of learning and representation of data. They have the ability to recognize the latent features and translate them into useful insights in no time. In this chapter, we are trying to identify, review and analyze the state of art deep learning approaches which contribute to the perpetuation of security in the Internet of Things (IoT). We mainly concentrate on the deep learning techniques aiding in the process of authentication feature extraction and detection of threatening invasions and malware. The chapter concludes with the discussion on the challenges faced while developing the algorithms for IoT networks which are suitable for diverse application scenarios and also provides a glimpse to the future perspectives.

1.1 Big Data and Internet of Things

The concept of big data which is characterized by the three Vs, volume, velocity, and variety is a paradigm which has received wide acceptance in the digital world in the last decade. Having more information, big data gave the opportunity to tackle the problems using completely different approach the use of social networks, online services and the development of open source frameworks expanded the possibility of big data. The advent of Cloud computing which offered scalability even broadened the opportunities of big data. IoT was another promising partner to take the big data to a higher level. The amount of data created and stored took a giant leap in this period with the emergence of the Internet of Things (IoT). The Fig. 1 shows the giant stride in the number of IoT devices used, according to the survey done by the connectivist.com. Techniques in big data analytics have the ability to handle the massive amount of continuous stream of data generated by these devices. Figure 2 picturizes the data flow in the process of insight creation from the data collected by the IoT devices.

1.2 Security in Internet of Things

In spite of the considerable benefits of IoT, it comes along with major security problems which need to be addressed. The global connectivity brings along the innate vulnerabilities and security risks. The vulnerabilities present in the IoT devices acts as the entrance for the adversaries to flare up various attacks in the IoT network [1]. The relationship between the vulnerabilities is exploited by the adversaries to invade into these networks. Therefore the advantages of IoT cannot be reaped out to its maximum without the multiple layers of security which safeguards the interconnected systems and devices. The increased rate of diversities, integrated with the wide scale of IoT systems, amplified the security threats of the current communication

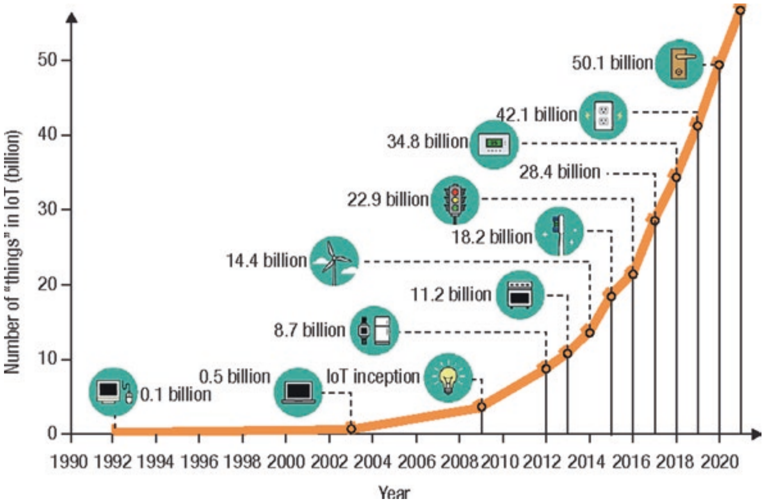


Fig. 1 IoT growth over years

scenario, which is being increasingly used to let interact humans, machines, and robots, in any combination.

IoT infrastructure is engrossed in the continuous transmission of data across the network of things for the achievement of specific goals. In such a symbiotic environment the security requirements authentication, authorization, access control, trust, integrity, confidentiality, privacy, secure middleware and trust etc. needs much importance [2] as shown in Fig. 3. Traditional security measures have to be tailored to the limited resource structure and the ad-hoc nature of the IoT network or new security solutions should be introduced to satisfy these requirements. Additionally, the scalability issue of the structure has to be addressed since the infrastructure is more dynamic in nature. These security issues have to be handled with a high degree of adaptability. Effective security mechanisms are to be deployed befitting to the limited functionality and constraint environment of the IoT systems. This necessitates the techniques for device-level security in communication and network monitoring. Traditional security solutions in conjunction with built-in security in the devices are needed to achieve dynamic detection, prevention, isolation and countermeasures against successful breaches.

Security of Internet of Things spans over all layers of the IoT Infrastructure i.e. Perception layer, network layer, transport layer and application layer [3]. The perception layer consists of various hardware nodes or sensors entrusted with the job of acquisition of various parameters from the residing environment and the network which is responsible for the transmission of the collected data to other nodes. Transportation layer which is an association of heterogeneous networks provides pervasive access atmosphere for the for perception layer, understands the information gathered, handles the transmission of data. The application layer consists of the application support layer as well as the IoT application layer. The support layer is

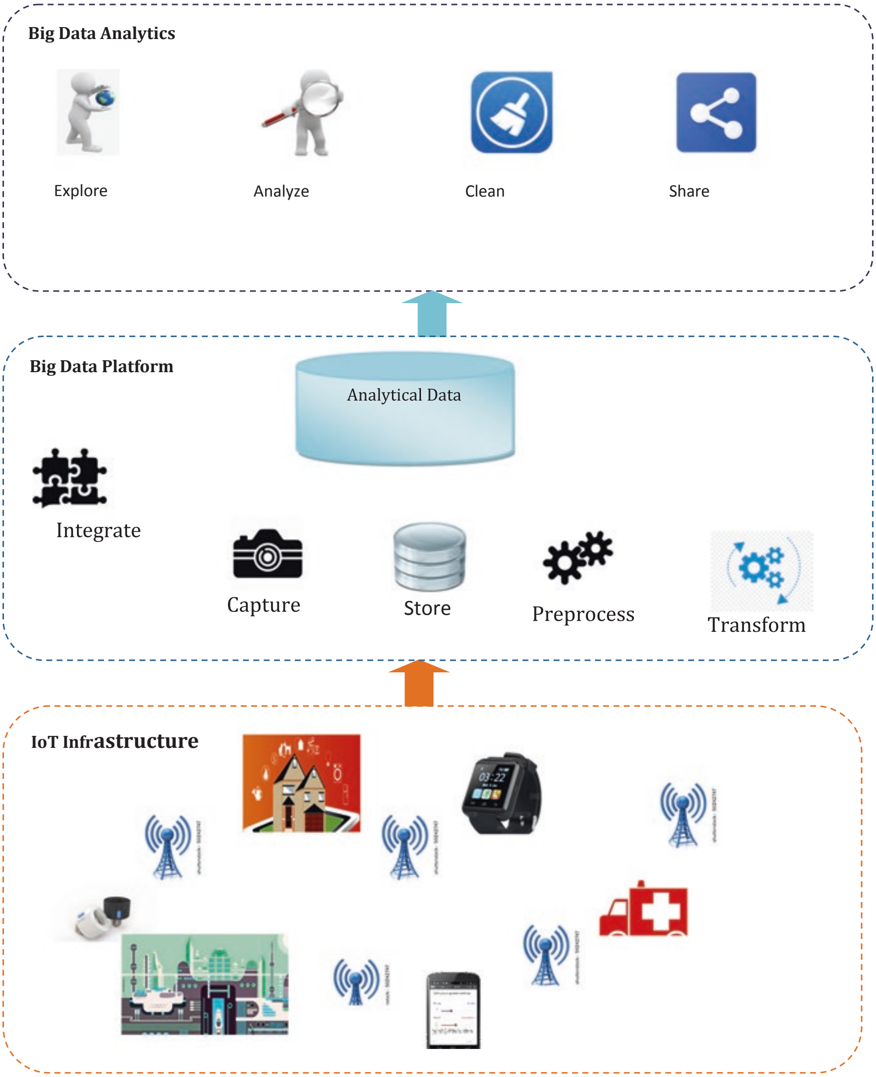


Fig. 2 The flow of Big data in Internet of Things

entrusted with the job of supporting all kind of business services, realizing intelligent computation and the allocation of resources in screening, selecting, producing and processing data. They should recognize malicious and benign data. The IoT application layer includes integrated or specific business applications. The issues in the technologies used in each layer contribute to the security threats of the layer. The summary of the issues in each layer is shown in the Table 1.

The security solutions for these problems cannot be realized by a specific quick fix in a single layer. Therefore, solutions that support cross layer usage are needed

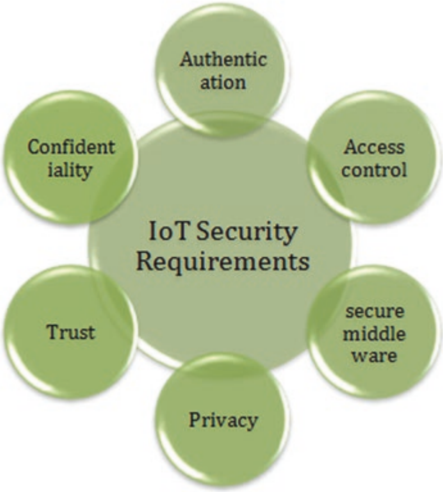


Fig. 3 IoT Security Requirements

Table 1 Summary of security issues in the IoT layers

Sl no	IoT Layer	Security Issues
1	**Perception layer**	No uniform encoding standard for RFID
		Multiple RFID tags send data simultaneously
		Lack of privacy protection
		Lack of trust management systems
		Data confidentiality
		Data authenticity
		Data integrity
		Issues due to heterogeneous integration
2	**Transportation layer**	Access network related issues
		Data security
		Phishing attacks
		Eavesdropping and interference
		Illegal node access
		DDos/dos attacks
		User information leakage
3	**Application layer**	Service interruption and attack issue
		Insecure data
		Issues related to access control
		DDoS attack

to be designed for addressing the issues in IoT infrastructure. The resource constraints in the nodes of a sensor network and multihop communications in open wireless channel make the security of sensor networks even more heavy challenge. Due to the explosive demand of IoT devices and their applications nowadays, the aspect of security demands high priority.

1.3 Overview of Deep Learning Techniques

Deep Learning has been a major focus in data science due to its capability to handle the enormous amount of data tactfully. The key benefit of deep learning in big data is that they can learn from a massive amount of unsupervised data or raw data which is uncategorized. Deep learning in its initial phase was proved to be successful in feature learning tasks. Deep Learning acquires the features itself, which enables the learning process to be more accurate and helps in the creation of better models. Feature extraction using deep learning techniques annex nonlinearity to the data analysis and make the discriminative tasks closely to heuristics. They fit perfect in the IoT paradigm which involves a large amount of data and complex relationships between different parameters, for solving intuitive problems. Nowadays, the potential for deep learning is utilized for classification tasks like intrusion detection, malware analysis, authentication etc.

Deep learning has earned success since it needs very little engineering by hand utilizing large amount of data. According to the authors of [4] a deep-learning architecture is a "multilayer stack of simple modules, all (or most) of which are subject to learning, and many of which compute non-linear input–output mappings". Each node in the stack converts the input to increase both the selectivity and the invariance of the representation. With multiple non-linear layers a system can implement extremely intricate functions of its inputs that are simultaneously sensitive to minute and insensitive to large irrelevant variations.

1.3.1 Evolution of Deep Learning

Deep learning finds its roots in neural networks which were formulated by Walter Pitts and Warren McCulloch in 1943. This mathematical model mimicked the working of neurons, the cells in the human brain which helps them in the thought process and decision making. The 50s and 60s saw the development of machine learning programs and the groundwork of deep learning was put in by Frank Rosenblatt in 1957 with the idea of perceptrons. In 1960 the control theory was introduced by Henry J Kelly, which laid the basics for the development of backpropagation model. The creation of Neocognitron, an ANN mainly used for pattern recognition tasks was created by Kunihiko Fukushima. This model which was used for handwritten character and pattern recognition tasks, recommender systems etc. influenced Hubel and Wiesel which resulted in the formulation of a variant of multilayer perceptrons which needs a minimal amount of preprocessing, Convolutional Neural Networks in 1979. Subsequently, Recurrent Neural Networks which work well for sequential data was introduced in 1980 but gained popularity after the advent of GPUs because of its computational complexities. Later, with the significant progress of backpropagation in the 70s, Yann LeCun combined Convolutional Neural Networks with back propagation in 1989. Long short-term memory a framework of recurrent neural networks was developed in 1997, by Sepp Hochreiter and Juergen Schmidhuber, which works well for sequential data.

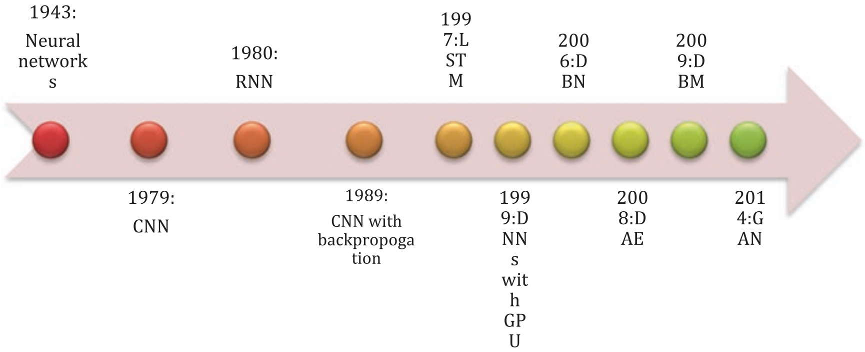


Fig. 4 Evolution of deep learning

With the rise of fast computing GPUs in the late nineties, deep learning took a new dimension. With GPUs faster processing with the images, the computational speed increased by a thousand times. Deep Belief Networks, which is widely used for dimensionality reduction for unsupervised training and as a classifier for supervised training, were introduced by Hinton in 2006. 2008 saw the emergence of Denoising Auto encoders which is trained to build up the data from the input containing noise. Deep Boltzmann Machine, in which the output of one BM is cascaded to multiple BMs, was introduced in 2009 by the Hinton. Recent advancement in deep learning is the introduction of Generative Adversarial Networks (GAN) which comprises two networks competing for each other to learn the data and get smarter. It is considered the most interesting idea in the last ten years of machine learning. Figure 4 shows the time line of the various mile stones in the development of deep learning techniques. Deep learning acts as a central axis where the processing of Big Data and the evolution of Artificial Intelligence, revolve around. Deep Learning is still in its adolescence and needs many innovative ideas to be incorporated.

1.3.2 Deep Learning Architectures

Deep learning has become one of the hot topics of research in the area of artificial intelligence. We present various deep learning architectures and their brief descriptions. Figure 5 shows the broad classification of deep learning algorithms.

Unsupervised(Generative) Algorithms

Unsupervised (generative) algorithms make the most of unlabelled data for training. They learn the likelihood of a given input to be in a class label and are assigned to the label to which it has the highest probability. Following sections explain various types of unsupervised mechanisms.

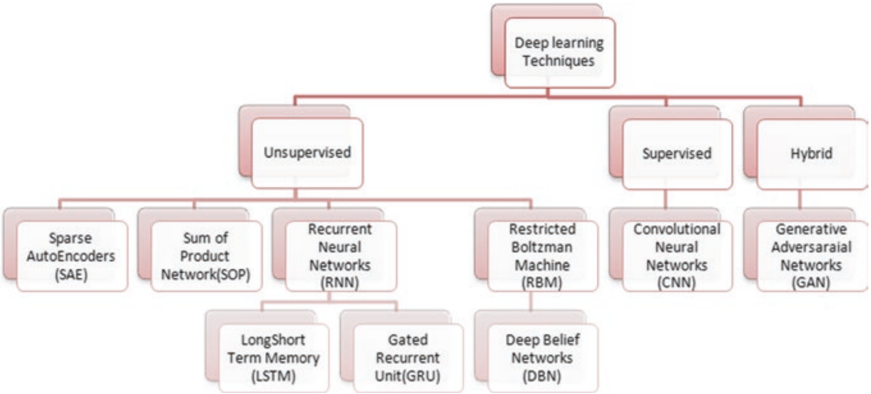

Deep learning Techniques
Unsupervised
Supervised
Hybrid
Sparse AutoEncoders (SAE)
Sum of Product Network(SOP)
Recurrent Neural Networks (RNN)
Restricted Boltzman Machine (RBM)
Convolutional Neural Networks (CNN)
Generative Adversaraial Networks (GAN)
LongShort Term Memory (LSTM)
Gated Recurrent Unit(GRU)
Deep Belief Networks (DBN)

Fig. 5 Classification of Deep Learning algorithms

Sparse Auto Encoders

Auto encoders are neural networks which apply back propagation and try an approximation to the identity function such that the output x^ ≈ x, where x is the input. The identity function is a trivial function applying several constraints like limiting the number of hidden units on the network, discovering inherent structure about the data. Auto encoders have an encoding stage and a decoding stage [5]. In the encoding stage, the input x is converted to the hidden layer h using the encoding function h.

$$h = f\left(W\left(1\right)x + b\left(1\right)\right).$$

Then the hidden representation h is recreated to the original input in the decoding stage.

$$y = g\left(W\left(2\right)h + b\left(2\right)\right).$$

Stacked (Sparse) auto encoders can be considered as a deep learning model which is constructed by stacking multiple auto encoders (as shown in Fig. 6) which uses layer-wise unsupervised pre-training. Pre-training in Auto Encoders is to train a single auto encoder using a single hidden layer. Each Auto encoder is trained separately before cascading it [6]. The number of nodes in hidden layers of the auto encoders will be lesser than that in the input layer which represents a new reduced feature set. The data is then reconstructed after complicated computations and these new transformed features are formed at different depths in the network. Denoising Auto encoders are a variant of auto encoders that is trained to build up the data from the input containing noise.

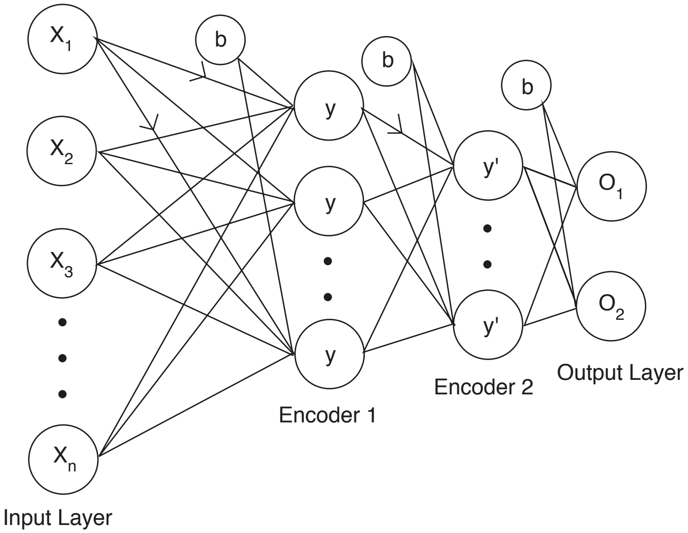
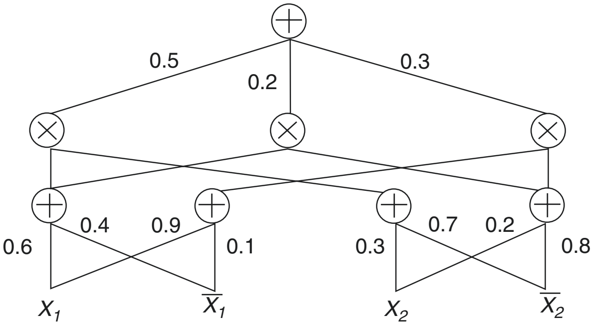


Fig. 6 Sparse Auto Encoder with two hidden layers and two class labels

Fig. 7 SPN implementing a naïve Bayes mixture model



Sum of Products Network

Sum of Product network(SPN) is a deep probabilistic model representing a tractable probability distribution [7]. They can incorporate features into an expressive model without requiring approximate inference. It is a rooted directed acyclic graph whose leaves are the variables and whose internal nodes are sums and products [8]. The sum nodes provide mixture models, while the product nodes express the feature hierarchy. Figure 7 shows an example of an SPN implementing a naive Bayes mixture model with three components and two variables. SPNs have achieved remarkable results on numerous datasets.

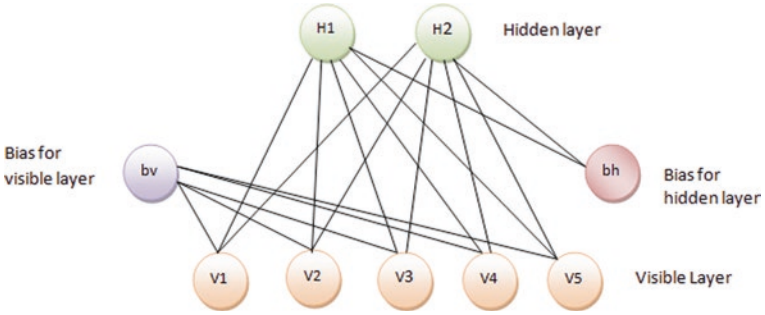


Fig. 8 RBM architecture

Restricted Boltzmann Machine

Restricted Boltzmann Machine (RBM), a network of stochastic neurons is a part of the family of energy-based models. At the same time, it is a probabilistic model too. They have the easiest architecture with two layers, visible layer and the hidden layer, and bias for each layer. Figure 8 shows the schematic representation of RBM. The hidden layer takes part in the process of transformation in the system maintaining the impervious to the observations. The neurons in the machine are in binary state 0 or 1 in a particular point of time. The state refers to the values of neurons in the visible and hidden layers. Conditional Probability is calculated for each node at each state P(h|v) to calculate the value of each unit in the hidden layer and then uses the conditional probability P(v|h) to calculate the value of each unit in the visible layer. This is repeated until convergence.

Deep Belief Networks

Deep Belief networks were constructed by Hinton by stacking various Restricted Boltzmann Machines creating a generative model consisting of many layers by greedily training each layer (from lowest to highest) as an RBM using the previous layer's activations as inputs (Fig. 9). The RBM in each layer exchanges the information with both the former and subsequent layers. Each layer is made up of a set of binary or real valued units. The heap of RBMs has a final Softmax layer which makes it a classifier that groups the unlabeled data in an unsupervised manner. Other than the initial and the final layer in Deep belief networks every layer serves as hidden layers to the nodes comes prior to them and as input (visible) to the nodes that come later [9].

Recurrent Neural Network

Recurrent Neural Networks are conventional sequential learning models that are effective in the processing of sequential information. They are called recurrent networks since they carry out the same job for every input independent of the prior

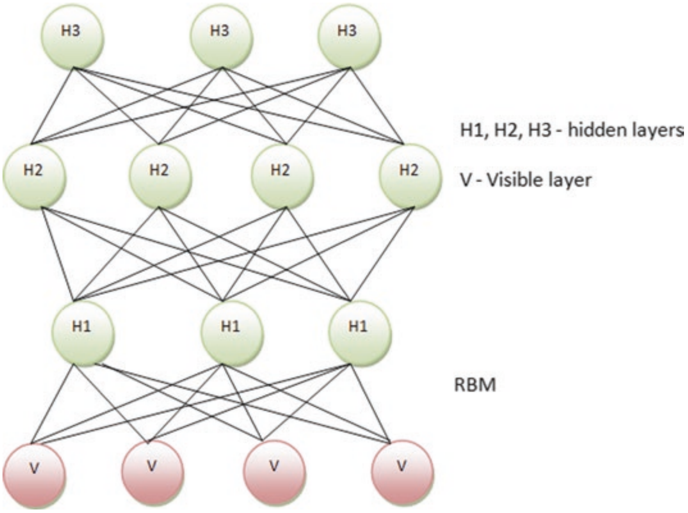


Fig. 9 DBN architecture

computations. They learn the features for the data by keeping the former inputs in the memory. A directed cycle is brought in to create the connections between neurons, as shown in Fig. 10 [10]. The deepness of the network will be as large as the length of the input data sequence. RNN has been found more beneficial in modeling the sequential data.

The input units: $\{x_0, x_1,\ldots.. ,x_t , x_{t+1},\ldots\ldots\ldots \}$
The output units: $\{y_0 , y_1,\ldots..,y_t , y_{t+1},\ldots.\}$
The hidden units: $\{H_0, H_1,\ldots\ldots,H_t, H_{t+1},\ldots.. \}$.

At the time step t, the recurrent neural network takes the current sample x_t and the previous hidden representation H_{t-1} as input to obtain the current hidden representation

$Ht = f(x_t , H_{t-1}), f$ is the encoder function

Several RNNs can be piled together to get a deep learning model. RNNs and its variants have displayed impressive performance in the domains like speech recognition, natural language processing etc. where there exists dependency among the input data.

Long Short term Memory

Recurrent neural networks capture random length dependencies of the input data but fail to acquire long-term dependencies because of the vanishing gradient. This drawback is surpassed by the model long short term memory model introduced by Hochreiter and Schmidhuber by preserving the error forbidding the gradient explosion. LSTM is a variant of RNN with four neural networks in a single layer. The main feature of LSTM is the presence of the state cell on the top of every layer,

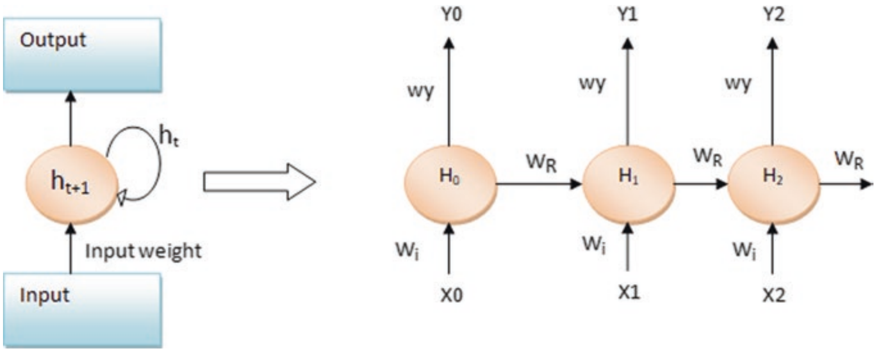


Fig. 10 RNN Architecture: Unfolded (right)

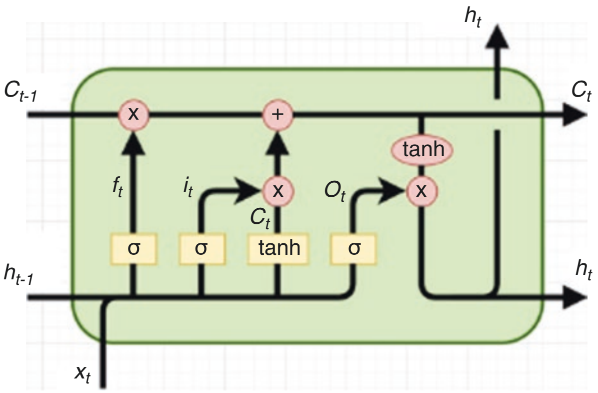


Fig. 11 Architecture of LSTM

which is responsible for the transmission of information from the former layer to the next layer. The gates in the LSTM accounts for the management of the information to be passed or dropped. To control the flow the information input gate, forget gate and output gates are used as shown in Fig. 11.

Gated Recurrent Units

Gated recurrent Unit is a less complex model of LSTM model decreasing the number of gates in the architecture. The GRU combines the "forget gate" and "input gate" in an LSTM to form an "update gate" and merges the hidden state and cell state, which led to the formation of a much simpler architecture of the model as shown in Fig. 12.

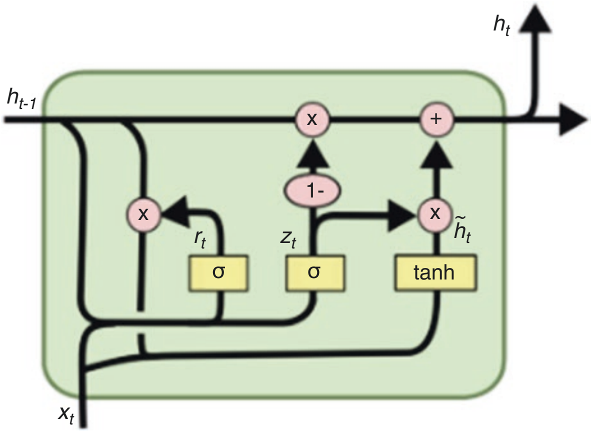


Fig. 12 Architecture of Gated Recurrent Unit

Supervised Learning

The main aim of supervised learning or discriminative model is to distinguish some parts of data for pattern classification with labeled data. Convolutional Neural Networks is the discriminative model among the deep learning models.

Convolutional Neural Networks

Convolutional neural networks are the deep learning model used extensively for feature learning and image classification. The main reason for the drastic boom in deep learning was the use of Convolutional networks in image recognition. They are used to categorize the images, group them by the similarity and do object recognition within the images. These algorithms had the ability to identify faces, persons, street signs and many other variations of perceptible data. Analogous to the other traditional neural networks, its structure is influenced by the neurons in animal and human brains. It mimics the visual cortex in a cat's brain containing a complex sequence of cells.

The time delay networks were the key influence on the origin of CNN. The reduction in the computation in TDNNS is due to the fact that the weights are been shared in a temporal dimension. The matrix multiplication in the conventional neural networks was replaced by convolutions in Convolutional neural networks. Thus the complexity of the network was reduced with the reduction of a number of weights. The feature extraction process in the traditional learning algorithms refrains in these networks thereby the images can be directly fed into the networks as raw input. So minimal preprocessing is done in the case of CNN model. Spatial relationships are utilized to reduce the number of parameters in the network, and leveraging the standard back propagation algorithms the performance is improved. Multilayer networks can be trained by CNN utilizing gradient descent to learn complex, high dimensional non linear mappings from large collections of data. Three basic concepts, local receptive fields, shared weights, and pooling is used by these networks. AlphaGo by Google is one example of the successfully implemented using CNN.

CNN is composed of a number of Convolutional layers succeeded by pooling layers and fully connected layers(similar to Perceptrons) as final layers as shown in Fig. 13. The input is three dimensional, p x p x q where p denotes the height and width of the input, q refers to the depth of the channel. There exist several filters in each layer of size m x m x n where m is smaller than the input image but n can be lesser or equal to q. Filters convolve with input and share the parameters, weight, and bias to create the feature maps of size p x m x 1. CNN calculates the dot product with the weights and its inputs as shown below

$$h^k = f\left(W^k \, X \, x + b^k\right)$$

But the created inputs are small regions of the real input volume. Overfitting is controlled by decreasing the parameters in the network by down sampling the feature map. A small contiguous region of the filter size is selected and the pooling operation is done on the region. Pooling might be max pooling or average pooling. Similar to the traditional neural networks the final stage layers are fully connected layers. They produce a high level abstraction of the data utilizing the prior low level and mid-level features. The final layer produces the probability of an instance to in a specific class or the classification scores.

For the classification of images as in the case of Fig. 13, the raw pixels will be the input to the Convolutional neural network. CNN learns to discover the edges from these raw pixels in the first layer. It utilizes these edges to identify simple shapes in the next layer. The successive layers will be capable of learning higher level features like facial shapes, buildings etc. utilizing the simple shapes from the previous layer.

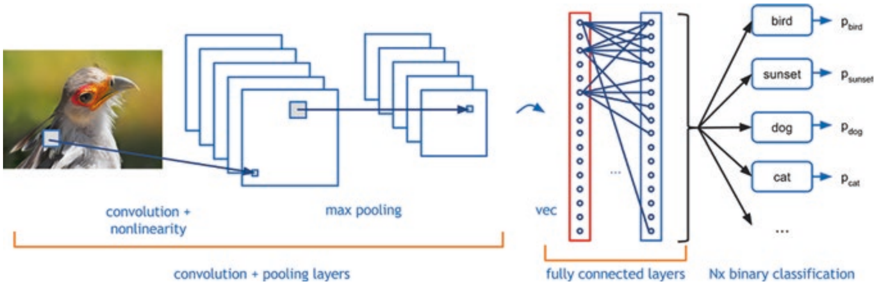


Fig. 13 Architecture of CNN

Hybrid Learning

Hybrid architecture integrates the advantages of supervised and unsupervised learning. They try to cluster data as well as identify the data. Generative adversarial network is an example of hybrid learning technique.

Generative Adversarial Networks

An innovative framework which trains both supervised and unsupervised simultaneously was put forward by Goodfellow in 2014 [11]. It consists of a two models generative G and discriminative D as shown in Fig. 14 where G captures the distribution of the data p_g in the real data t and D model differentiate the original input data and the data from the model G i.e. p_m. In every iteration, the generative model is opposed against its adversary, a discriminative model which tries to identify whether the given sample is generated by the model or the original data. Generative Model G generates more realistic data to fool and complicate discriminator model D tries to identify the genuine ones. Tug of war among these models helps them improve their techniques to identify the genuine one from the fake one. This two-player game is conclusively proved with Value function V(G,D).

$$\min_{G} \max_{D} V(G,D) = E_{t-pdata}[logd(t) + E_{m-pm(m)}\left[\log\left(1 - D\left(G(m)\right)\right)\right]$$

Where

D(t): the probability that t came from the data
p_{data}: distribution of the real-world data.

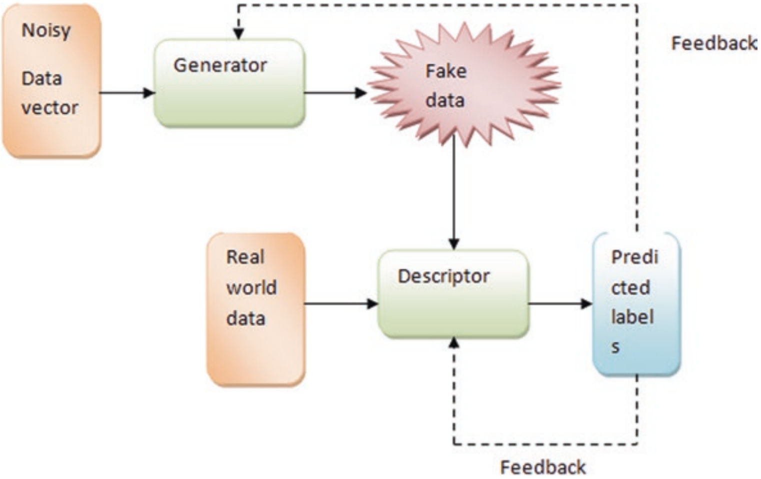


Fig. 14 Architecture of GAN

The model reaches the equilibrium when both reach the point where none of them can be improved i.e.

$p_g = p_{data}$. This means that the discriminator can no longer identify between the two distributions.

2 Pertinence of deep learning in IoT security

Deep learning techniques which gained remarkable achievements in the area of computer vision, automatic speech recognition, pattern recognition etc., have now been used extensively for the sustainment of security in IoT [12]. Figure 15 shows the taxonomy of the application of deep learning techniques for IoT security. They are classified as the approaches used for authentication, intrusion detection, feature selection and malware detection.

2.1 Deep Learning for Authentication

The focus of the deep learning techniques while applied in authentication is on identity assurance rather than fraud detection. These techniques have been used for the authentication of the users as well as the IoT devices. Various deep learning approaches used for the authentication process of users and IoT devices are summarized in Table 2.

2.1.1 User Authentication

A user authentication framework was proposed by Lee et al., extracting features based on users' interaction with the touchscreen, which used deep belief networks to classify the users [13]. They extracted stroke based features and session-based features for authentication. A modified DBN with two hidden layers was used for

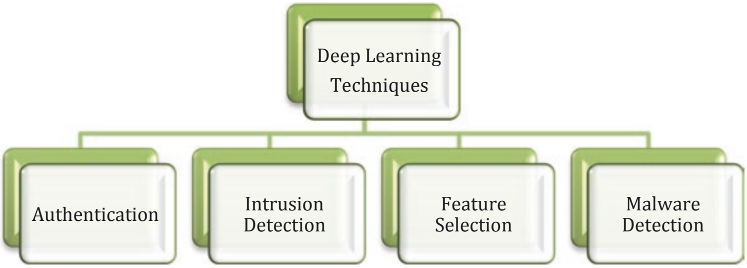


Fig. 15 Applications of deep learning in IoT security

Table 2 Deep Learning Techniques for Authentication

Sl No	Deep Learning Techniques for Authentication	Inferences
1	User authentication using stacked auto encoders [15]	Authentication of users from the physiological and behavioral features from channel state information (CSI) measurements of WiFi signals.
		Resilient to user spoofing attacks.
2	User authentication using Deep Belief Networks [13]	Collects user interaction features, stroke based features and session based features to authenticate the user.
		DBNs including the dropouts are used thereby avoiding overfitting on small training sets.
3	User authentication and identification using Deep Belief Networks [14]	Provide very less EER in identifying and authenticating a user using keystroke timing features.
4	Authentication of IoT devices using LSTM [17, 18]	Preserve long term dependencies in sequential data which are suited for wireless signals.
5.	Signal Authentication using LSTM [17]	LSTM reduces complexity and latency of the attack detection compared to other security methods
		Authenticates the signal, extracting the stochastic features from IoT signal and watermarking these features inside the original signal.
		Allows the cloud to detect sophisticated eavesdropping
		Attacks, since the attacker will not be able to extract the watermarked information

the classification. DBN produced impressive results compared to other examined methods with an identification rate of 81.5% and a median EER of 9.93%. Deep Belief Network was used for authentication using another modality, keystroke dynamics (a behavior-based unique timing patterns in an individual's typing rhythm) by Saket et al. [14]. They considered the identification of a user as a binary classification problem and used the keystroke features like the hold time, key down- key downtime and key up-key downtime to authenticate a person. This network model, DeepSecure had three hidden layers of 100, 400 and 100 dimensions layers. The considerable number of hidden layers introduces sparsity which can secure the inter-feature relations. This eliminates the need for manual feature engineering and eventually bringing forth a model which is more robust and less prone to over fitting on this key-stroke recognition problem when compared to a simpler 1 hidden layer model. Another deep learning based user authentication Scheme (as shown in Fig. 16) was proposed by the authors of [15], in which representative features were extracted from channel state information (CSI) measurements of Wi-Fi signals, to accurately identify an individual user. The system performs activity recognition and human authentication by building a three-layer deep neural network (DNN) model based on AutoEncoder. Unlike other authentication schemes based on high dimension feature sets and linear classification models (e.g., SVM), non-linear physical and biometric abstractions learned by DNN model are computation efficient and are robust to small-scale input variations. The stated network roughly identifies the

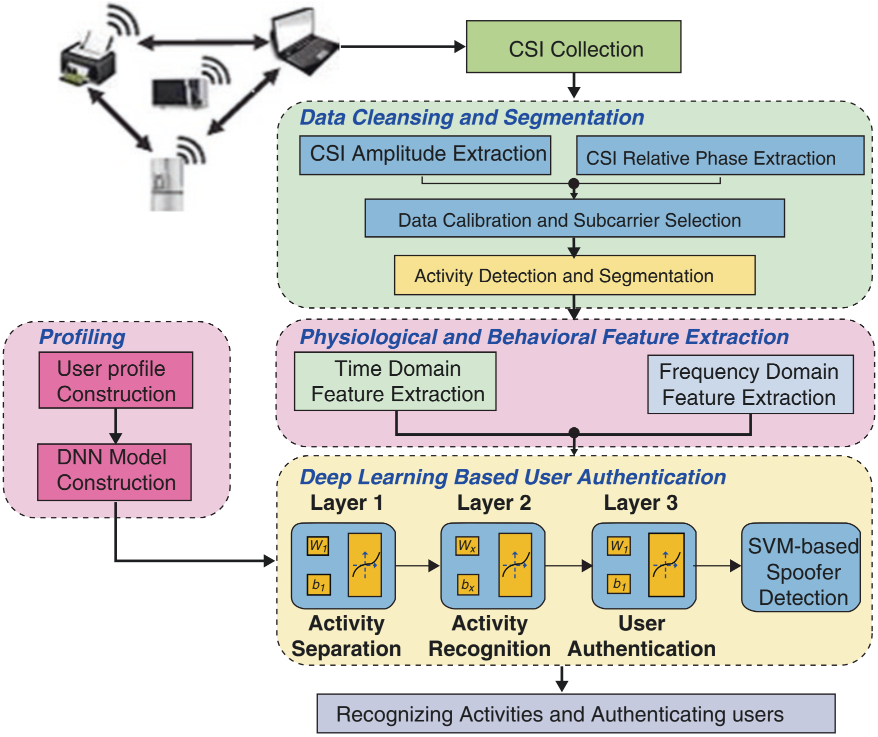

CSI Collection
Data Cleansing and Segmentation
CSI Amplitude Extraction
CSI Relative Phase Extraction
Data Calibration and Subcarrier Selection
Activity Detection and Segmentation
Profiling
User profile Construction
DNN Model Construction
Physiological and Behavioral Feature Extraction
Time Domain Feature Extraction
Frequency Domain Feature Extraction
Deep Learning Based User Authentication
Layer 1
Layer 2
Layer 3
W_1
b_1
W_x
b_x
W_1
b_1
SVM-based Spoofer Detection
Activity Separation
Activity Recognition
User Authentication
Recognizing Activities and Authenticating users

Fig. 16 Overview of deep learning based user authentication formulated by Shi et al., 2017

activity type in the first layer and subsequently the activity details in the second layer. The third layer recognizes each individual user with a softmax function. The integration of the SVM model with the DNN, immune the system against spoofing attack.

2.1.2 Device Authentication

The efficacy of the model LSTM was exploited by Rajshekar et al. to learn the hardware imperfections of the low powered radio device and identify the features which makes them unique [16]. LSTM capitalize on the temporal correlation between the I/Q streams of wireless signals to identify legitimate nodes from high power adversaries that transmit identical modulation, coding, and even data, given that these adversaries inadvertently introduce their own distinct imperfections. The technique was examined with LoRa transmitters and much higher software radio Adversaries and found buoyant to noise, multi-path, and signal attenuation. This approach was strengthened by Ferdowsi et al. by integrating game theory to the framework [17].

The framework was designed to allow the cloud to authenticate the signals and disclose the presence of any adversary who may change the devices' output signal. LSTM extracts the stochastic features like spectral flatness, skewness, kurtosis, and central moments from IoT signal and watermark these features inside the original signal. Since enormous amount of computational resources is required for the authentication, the cloud cannot authenticate all transmitted signals from the IoT devices simultaneously. Predicting the vulnerability of IoT devices is considered as a non-cooperative game between the cloud and the attacker, considering the constraint of the resources in these devices. The cloud optimally chooses the device to be authenticated with the help of Nash equilibrium. 30% reduction in the number of compromised devices was observed using this approach and improved the protection of the system in massive IoT scenario.

2.2 Deep Learning for Intrusion Detection

On top of the secure foundation built by the cryptographic techniques and the secure protocols Intrusion Detection Systems (IDS) act as the first layer of defense in the arena of IoT Security. The ability to recognize, the patterns of typical attacks and abnormal activity patterns, makes IDS primary choice which can be deployed over all levels. IDS monitors, recognize the patterns of typical attacks and abnormal activity patterns and reports to the security management system. Deep learning which has been considered as a breakthrough in the arena of Artificial Intelligence has raised the potential of intrusion detection to achieve high detection rate and low false alarm rate. They utilize the network traffic data to identify the intrusions. Tables 3 and 4 summarizes the deep learning approaches used for the classification tasks in intrusion detection systems.

Most of the literature has utilized the KDDCUP99, NSL–KDD and UNSW-NB15 datasets to substantiate their proposed techniques. KDDCUP99 is a collection of raw TCP dump data which contains 41 attributes and a label assigned to each instance as either attack type or as normal. There are 22 attacks in the training data out of the 39 attacks in the test data. The attack types are categorized into 4 groups: DOS: Denial of service – e.g. syn flooding, Probing- Surveillance and other probing, e.g. port scanning, U2R: unauthorized access to local superuser (root) privileges, e.g. buffer overflow attacks, R2L: unauthorized access from a remote machine, e.g. password guessing. NSL-KDD is a sophisticated version of KDDCUP99 having similar attack types. UNSW-NB15 data set is also raw traffic dataset which contains 9 attack groups- Backdoor, Analysis, DoS, Exploits, Fuzzers, Generic, Reconnaissance, Shellcode and Worms. There are 49 features in the dataset along with the class label. The deep learning architectures act on these raw traffic data to categorize benign traffic and attacks.

Table 3 Deep Learning approaches for Intrusion detection

Sl No	Approaches for Intrusion detection	Inferences
1	Deep neural networks for Intrusion detection [21] [19]	Effective attack detection
2	Distributed attack detection scheme using DNN([20]	Better performance than the centralized model
		Collaborative sharing of learning parameters avoids overfitting of local parameters which helps in achieving better performance
3	Deep neural networks for intrusion detection in invehicle security [23]	A real-time response to the attack with a improved detection ratio in controller area network (CAN) bus.
4	Ensemble Algorithm using DNN for Intrusion Detection [24]	DNN along with spectral clustering algorithms is used.
		Better performance than shallow counterparts
5	Intrusion detection using Auto Encoders [25]	Self taught learning
		Works on unlabelled network traffic data collected.
6	Stacked auto encoders for Intrusion detection [6].	IDS designed for different layers uses layer specific features.
		Lightweight IDS achieve comparable detection rate as the ordinary IDS.
7	Auto Encoders for traffic Identification [26]	Deep structures of works better than shallow counterparts.
8	Intrusion detection using Deep Belief Network [27]	Improved classification rate for known and unknown attacks with minimum number of false alarm rate.
		Achieved higher accuracy with the training done on smaller amount training data
9	Deep belief Network for Intrusion detection [28]	Uses four hidden layer RBMs.
		Efficient use of very large sets of unlabeled data and can be pre-trained in completely unsupervised learning.
		Limited labeled data is used to fine-tune DBN for a classification task
10	Restricted Boltzmann machine for Intrusion detection [29]	Combines the expressive power of unsupervised models with good classification accuracy capabilities to infer part of its knowledge from incomplete training data.
		Not restricted to a prior knowledge base,
		It can enable the detection of any type of unknown anomalous events
		Effective in coping with the zero-day attacks.

2.2.1 Deep Neural Networks

Far-reaching researches have been done in the area of detection of intrusions in the cyberspace. A deep learning method was used by Yavuz et al. to identify the routing attacks in the Internet of Things [19]. They used highly scalable, deep-learning based attack detection methodology for detection of IoT routing attacks with high accuracy and precision for continuous monitoring. Another attempt using the same

Table 4 Recurrent Neural Networks for Intrusion detection

Sl No	Approaches for Intrusion detection	Inferences
1	LSTM-RNN for intrusion detection [35]	Find an optimal hyper-parameter for LSTM-RNN and confirm the detection rate and false alarm rate
2	Reduced size recurrent neural networks for Intrusion detection [32]	a reduced-size structure of RNN is used, based on the four group of input features
		Improved classification rates
3	Recurrent neural networks for Intrusion detection [33]	Fully connected model has stronger modeling ability and higher detection rate than the reduced-size RNN model
4	Deep recurrent neural Network paradigm for intrusion detection [10]	Uses a descriptive model of deep recurrent neural Network (RNNs)
		Works with low false alarm with new unseen threats
		Bi-directional techniques can neutralize sequence dependencies via considering forward and backward order of request sequences.
5	LSTM-RNN for DDoS attack detection [36]	Resolve the vanishing gradient problems
		Keep details of attacks learnt from training process and make detection decisions based on this stored information on gated cell
6	LSTM based ensemble method for intrusion detection [37]	A language based model for intrusion detection
		Learns the semantic meaning and interactions of each system call
		Needs significant smaller training overhead since no database is used for the storage of patterns
7	MS-LSTM for anomaly detection [40]	A multiscale LSTM is used assuming the internet flow as a multi-dimensional time sequence and learns the traffic pattern from historical features in a sliding time window.
8	Traffic classifier using CNN-RNN [41]	RNN combined with a convolutional neural network is used to provide best detection results.
		Robust and gives excellent F1 detection scores under a highly unbalanced dataset
9	Gated recurrent unit for Intrusion detection [39]	Bi-directional GRU and multi-layer GRU is used for intrusion detection
10	CNN –RNN for Intrusion detection [42]	CNN with first layer and variants of RNN are used as subsequent layers.
		Remarkable performance than other classifiers.

approach was done by Diro et al. [20] for identifying the intrusions in the IoT network. They utilized the self taught and data compression capabilities of deep learning techniques to discern attacks from the benign traffic. They proved that the deep model outperformed the existing methods available for detecting the attacks. Impressive detection rate was observed for the experiments with the stated approach [21]. An ensemble model which combines spectral clustering with deep neural networks to detect the attack types was proposed by Ma et al. [22]. The Clusters capture the network features and break down them into k subsets to learn more knowledge and patterns from analogous clusters. Deep neural networks help in acquiring highly abstract features from these subsets. The model is proficient in classifying the sparse attack cases and increases the security in real security systems. The optimization of the weight parameters and the thresholds of each DNN layer remain as a limitation of the work. DNN has been applied to secure the in vehicular network inspecting the CAN network packets. Experimental results demonstrate that the stated method demonstrates a superior performance in terms of the detection rate.

2.2.2 Auto Encoders

Auto encoders are widely used for dimensionality reduction and data denoising nowadays. But attempts were done for the classification tasks too. Niyaz et al. used self-taught learning, based on sparse auto-encoder and soft-max regression, to develop a Network Intrusion Detection Systems [25]. Auto encoders were used to learn the features from the dataset. The learned features were applied to the labeled test dataset for classification. They used the n-fold cross-validation technique for the evaluation of performance and obtained a reasonable result. Aminanto et al and Wang et al. have examined the applicability of autoencoders as classifiers in network traffic data [6, 26].

2.2.3 Restricted Boltzmann Machines

The capability of Restricted Boltzmann machines to identify the latent factors in the data is exploited to find the anomalies in the security domain. The abnormal behavior of the network depends upon several factors and these factors are captured easily by Boltzmann machines and classify them as benign traffic or attack traffic. Fiore et al. utilized RBM which belongs to the family of energy based models to find anomalies in the network in a semi-supervised manner [30]. The generative power and the classification accuracy of DRBM make them efficient to extract the inherent aspects of the benign traffic. Since they are not confined to any prior knowledge base, they can be used for the detection of anomalous behavior. The performance of such IDS is enhanced by combing RBM with SVM by Bo Dong et al. [31].

2.2.4 Deep Belief Networks

Being the most influential deep neural networks, DBN is used for classification while associating the class labels with the feature vectors [27]. DBN utilize a very large set of unlabeled data and make use of unsupervised learning for pretraining. A limited number of labeled data can be used for the process of fine-tuning the model for classification. Gao et al. have proved that the deep belief networks perform better than the SVM and traditional neural networks [28].

2.2.5 RNN

RNN are powerful for modeling sequences since they have cyclic connections. Sheikhan et al. proposed a reduced-size structure of RNN, based on the four group of input features. They showed remarkable classification rates [32]. However, the nodes of layers are partially connected, the reduced RNNs do not show the ability of deep learning to model high-dimensional features, and the authors do not study the performance of the model in the binary classification. Chuan et al. proposed a three layer RNN architecture with 41 features to model a deep approach for intrusion detection [33]. They proved that fully connected model has stronger modeling ability and higher detection rate than the reduced-size RNN model and is superior to other classification methods in both binary and multiclass classification.Lopez et al. used a combination of RNN with CNN to classify the network traffic [34].

Jihyum etal proposed LSTM- a variant of RNN based model for finding the intrusions [35]. They find an optimal hyper-parameter for LSTM-RNN and confirm the detection rate and false alarm rate. Two variations of RNN, bi-directional RNN, Long Short Term Memory (LSTM) and bi-directional (LSTM) was used by Elsherif to develop solution that detects anomaly inside a sequence of user's requests [10]. He used a descriptive model of deep Recurrent Neural Network (RNNs) and works with low false alarm with new unseen threats. He proved that bi-directional techniques can neutralize sequence dependencies via considering forward and backward order of request sequences. The problem of vanishing gradient is resolved by using LSTM for IDS by Bediako et al. [36]. It keeps the details of attacks learnt from training process and make detection decisions based on this stored. A language based ensemble model for intrusion detection was proposed by Kim et al. which learns the semantic meaning and interactions of each system call [37]. It needed significant smaller training overhead since no database is used for the storage of patterns. Cheng et al. used a multiscale LSTM assuming the Internet flow as a multidimensional time sequence and learns the traffic pattern from historical features in a sliding time window [38]. Gated Recurrent Unit has been used to detect the attacks by the authors of [39]. They used bi-directional GRU and multi-layer GRU for intrusion detection.

2.2.6 Convolutional Neural Networks

Convolutional Neural Networks have been widely used in the field of computer vision since they have proved its efficacy in working with the images. A small amount of work in the area of intrusion detection is available in the cyber security paradigm using CNN. The capability of CNN to excerpt high-level feature representations that portrays the abstract form of low-level feature sets of network traffic is exploited to distinguish benign and malignant traffic. The authors of [42] assessed the efficacy of CNN and the integration of sequential data modeling techniques for the classification of benign and malignant network connections. They used CNN as the first layer with a recurrent neural network and its variant as subsequent layers.

They claim that deep learning based approaches such as CNN and RNN, LSTM, GRU are suitable at modeling network traffic as a sequence of TCP/IP packets in comparison to other conventional machine learning classifiers. Lopez et al. used a combination of RNN with CNN to classify the network traffic [34].

2.3 Deep Learning for Feature Selection

Feature Selection is a major process that influences the performance of a specific model. The absence of manual feature manipulation is one of the important advantages of deep learning. Deep learning has been used for feature selection for an intrusion detection system by the authors of [6] to prove that the reduced input features are sufficient to achieve comparable detection rate as the whole features. They provide good feature representation of the unlabelled data collected from the network. Wang et al. have used stacked auto encoders to learn an efficient, compressed representation for a set of data for the identification of anomalies in TCP flow data [26]. The three-layered architecture transformed the raw data very efficiently with some computations in an unsupervised manner which makes them the prime choice for feature extraction. Auto encoders can restore data based on less information loss and error. Li et al. used this approach to prepare the data for malicious Code Detection, converting high-dimensional data into low dimensional codes with the nonlinear mapping and extracted the main features of the data [24]. Tobiyama et al. utilized the efficacy of RNNs with LSTM units to construct behavioral language model for the extraction of features from the process behavior of the terminal [43]. The model consists of an input layer a normal hidden layer, two hidden LSTM layers, and an output layer. Dropout for non-recurrent connection is used in the training phase. The features processed using trained RNN are then converted to feature images. The information of previous inputs is accumulated in the last hidden layer. Some sort of regularity will be found in the extracted features if the RNN is trained well. The model could classify malware processes with more preciseness by using a larger amount of data. Pascanu et al. experimented to learn the language of malware for the detection of unknown threats [44]. Bidirectional recurrent models were trained to predict next API call and use the hidden state that encodes the past event history as the fixed-length feature vectors. This is given to a separate classifier for the classification process. Max-Pooling is used over the values of the hidden units in the time since the hidden units may learn to specialize in detecting different and potentially reordered temporal patterns.

The efficacy of feature extraction using deep learning is utilized in the paradigm of anomaly detection in gas turbine combustors, where Stacked denoising auto encoders are used for the learning the features from the sensor readings of exhaust gas turbine combustors [45]. The features captured with deep learning approach perform better in acquiring the relationship between all sensor measurements and the latent behavior of the combustor compared to manual feature engineering. The learned features were fed into a neural network for identifying the anomaly in the measurements. This increased the performance of the anomaly detection system considerable. Also, the use of SDAE makes the system more immune to the noise in the input. The same characteristics of SDAE made Yao Wang et al. use them for the identification of malicious JavaScript code in web pages on the Internet. The use of feature engineering using deep learning without human intervention increased the detection accuracy of the classifier remarkably [46]. Salama et al. used DBN to reduce the dimension of feature sets making them appropriate for intrusion detection. The hybrid intelligent system combining the advantages of deep belief network and support vector machine. The reduced data output is improved by the use of DBN along with back-propagation. The model has the BP-DBN structure composed of 2 RBM layers. The data is reduced from 41 to 13 features by the first RBM layer and from 13 features to 5 output features by the second RBM layer on NSL-KDD data. DBN gives better performance than the other reduction methods (Table 5).

Table 5 Deep learning approaches for feature extraction

Sl No	Approaches for feature extraction	Inferences
1.	Feature extraction using stacked denoising auto encoders [46]	Extract more abstract features of JavaScript code
		Yields high classification accuracy compared to its shallow counterparts
2.	Auto Encoder for Dimensionality Reduction [47]	Space mapping ability of AutoEncoder's is utilized for reducing dimensionality of the data thereby abstracting the main characteristics.
		Restore data based on less information loss and error.
3.	Feature extraction and selection using auto encoders [26]	Reduce the manual work since the model is trained automatically once inputs of the model and stopping criterion of the iteration are determined.
4.	Feature extraction using stacked denoising auto encoders [45]	Features are explicitly learned without class labels
5.	Feature learning using Recurrent Neural Networks [48]	Language of malware is learned for the detection of unknown threats.
		Bidirectional recurrent models are used.
		Max-pooling is used over the values of the hidden units in time since the hidden units may learn to specialize in detecting different and potentially reordered temporal patterns.
6.	Recurrent Neural Networks for feature Extraction [43]	RNN is used to the extract features of the process behavior in a terminal.
		Trained features are converted to a feature image which is sent to classifier to be labeled malignant or benign.
		Regularity will be found in the extracted features RNN is trained well
7.	Deep belief networks for feature reduction [49].	Features were reduced considerably, from 41 features to 5 features in NSL KDD data
		DBN gives better performance than the other reduction methods.

2.4 Deep Learning for Malware Detection

Malware detection has emerged in the past years due to the rise in the threat caused by malware to large organizations. The major approaches for malware detection are static analysis and dynamic analysis [50]. The malware file or the group of files is evaluated precisely in the binary form or unpacked in the static analysis while, the binary files are executed, and the actions are reported in dynamic analysis. Dynamic detection is less exposed to obfuscation can offer direct observation of malware action, and makes it difficult to recycle existing malware. Static analysis, is exposed to obfuscation, and need no special set up for the data collection, but they are cooperative with the deep learning. Deep models perform efficiently in terms of number of fitting parameters than shallow networks. Table 6 summarizes the major works that uses deep learning for malware detection.

Saxe et al. proposed a malware detection approach based on deep neural networks (DNN) which achieves high a detection rate of 95% and a low false positive rate of 0.1% on an empirical dataset of over 400,000 software binaries [50]. This approach requires simple computation to perform feature extraction and it can achieve good accuracy. Even though the approach gives remarkable results, the performance collapse significantly in the time split validation since relying on syntactic features. Lie et al. has proposed a model which adapt to the environment to obtain remarkable detection of malicious code. They used DBN as a classifier for several times deep learning detecting malicious code. The detection accuracy was improved as the number of iterations was increased. The use of multiple deep learning shows better performance than surface learning model. An ensemble of deep feed forward networks and deep recurrent neural networks was used by Jung et al. for the detection of zero day flash malware detection [51]. Based on process behavior in possible infected terminals, Tobiyama et al. utilized the efficacy of CNN to annotate the

Table 6 Deep Learning Approaches for Malware detection

Sl No	Approaches for malware detection	Inferences
1.	Deep neural networks for malware detection [50].	Requires simple computation to perform feature extraction and it can achieve good accuracy Performance decays significantly in the time split validation since relying on syntactic features
2.	Malicious Code Detection using Deep Belief Networks [47]	Increase in the number of iterations in the DBN, increases the performance
3.	Malware Detection with CNN-RNN using Process Behavior [43]	Based on process behavior in possible infected terminals. CNN classify the malware process from the features extracted by RNN
4.	Visualized Malware Classification Based-on Convolutional Neural Network [52]	Malware features are converted to images and these features are fed into CNN for classification

behavior as malware or benign. Another deep learning technique RNN was used to convert the features collected to feature images such that they can be fed into CNN for the classification. Better performance was obtained for the work done by the authors of [52] by using the same approach.

3 Challenges and the Road Ahead

3.1 Challenges in Applying Deep Learning in IoT Security

The security issues in the Internet of Things are application specific, so are their solutions. With heterogeneous application contexts and various security requirements, they demand application-specific solutions. The network has to be equipped with technologies which can adapt to real time changes during the production or to foresee and refrain from events that might annihilate various operations. The IoT arena demands cross-layer security architecture since a quick fix solution is not applicable. Lightweight solutions that meet the specific requirements are to be designed for the specific application. The resource constraints and the limited computational capabilities of edge nodes are the major challenges for developing deep learning solutions in IoT.

3.2 Future Perspectives

3.2.1 Resource Constraint Deep Learning for Edge Computing

Adopting Artificial Intelligence and machine learning to the security of IoT, leveraging the efficiency of deep learning, reaps the reward of enhanced security in the system. Deep learning contributes to a feasible solution for the security scenario in the IoT networks to prevent the intrusions before any harm is done to the whole system. Deep-learning based algorithms surpass the explicit hand-made feature extraction methods amassed with traditional classifiers and can achieve equivalent accuracies for both noise-free and noisy data. The absence of manual feature manipulation, unsupervised pre-training and compression capabilities makes the application of deep learning beneficial for the resource constrained networks. Despite the distinguished performance of deep learning techniques, due to the increased computational complexity, there is a high demand in the designing of light weight versions of these techniques to make them resource friendly in the IoT scenario. Resource constrained solutions can be embedded in the device. The data processing needs can be contented "at the edge," where the data is collected, or where the user performs certain actions. Including additional capability of intelligence to process the data at the edge reduces the overhead of transmission of large chunks of data in real time. Furthermore, it reduces the response time to events by forbidding the transit of data to and fro the cloud for the computational purpose. Adding decision making capability closer to the devices contributes to the overall performance of the systems.

3.2.2 Adversarial Deep Learning

Adversarial deep learning has caught attention recently since they have evolved as a serious threat to the machine learning systems. It can be considered as a rendezvous of machine learning systems and cyber security systems. ADL which was concentrated in computer vision has disseminated to other domains too. ADL refers to the alteration in the original data to confuse the machine learning model and force them to misclassify the data. DL systems have to deal with mainly two types of attacks, Evasion, and poisoning. In evasion, the attacker alters the inherent behavior of the data to stay anonymous, and poisoning means the training data itself is altered. Robust solutions against the AML should be used along with intrusion detection systems to make the approaches impervious. Adversary samples are made with the help of evolutionary algorithms, Fast Gradient Sign method (FGSM) and Jacobian-based Saliency Map Attack (JSMA). The variations done are hard to be sensed by the humans. Although solutions like distillation, incorporation of the adversarial component in the loss function, training with adversarial samples first to reduce the effect, etc. have been proposed, there exists a large realm for the researchers interested in ADL to work on.

4 Conclusion

The revolution of connectivity that brewed around the world in the past three decades gave rise to the third wave in the development of internet, Internet of Things which became an inevitable part of human life. Big data analytics harness the massive amount of data generated by Internet of Things and convert to well-analyzed data which is extremely valuable in today's world. To discover the sophisticated latent features abstract deep learning techniques are used. This abstraction ability and capability to handle the enormous amount of data tactfully, makes it a major focus in data science. In this chapter we have tried to limelight different deep learning approaches utilized in the area of cyber security. It gives a broad analysis of the deep learning techniques for feature extraction and classification tasks like intrusion detection, malware analysis, authentication etc. This chapter provides a sketch of the state of the art deep learning techniques, challenges faced and pointers to the future research direction. With Internet of things as the senses, big data as the powering force and the deep learning as central processing pivot, we can realize a smart connected world in future.

References

1. George G, Thampi SM (2018) A graph-based security framework for securing industrial IoT networks from vulnerability exploitations. IEEE Access 6(September):43586–43601
2. Sicari S, Rizzardi A, Grieco LA, Coen-porisini A (2015) Security, privacy and trust in internet of things: the road ahead. Comput Netw 76:146–164
3. Jing Q, Vasilakos AV, Wan J, Lu J, Qiu D (2014) Security of the Internet of Things: perspectives and challenges. Wirel Netw 20(8):2481–2501
4. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
5. Kim, Kwangju, Muhamad Erza Aminanto, and Harry Chandra Tanuwidjaja.(2018) Network Intrusion Detection Using Deep Learning: A Feature Learning Approach. Springer.
6. Aminanto, M. E., & Kim, K. (2016) Deep learning-based feature selection for intrusion detection system in transport layer. In Proceedings of the Summer Conference of Korea Information Security Society (CISC-S'16), pp 535–538, 2016
7. Rooshenas A, Lowd D (2014) Learning sum-product networks with direct and indirect variable interactions. Proc 31st Int Conf Mach Learn 32:710–718
8. Poon, H., & Domingos, P. (2011, November). Sum-product networks: A new deep architecture. In 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) (pp. 689–690). IEEE.
9. Kim, J. W. Classification with Deep Belief Networks." https://www.ki.tu-berlin.de/fileadmin/fg135/publikationen/Hebbo_2013_CDB.pdf
10. A. Elsherif(2018) "Automatic intrusion detection system using deep recurrent neural network paradigm," Journal of Information Security and Cybercrimes Research (JISCR), vol. 1, no. 1, 2018.
11. J. P.-A. Ian J. Goodfellow, D.-F. , Mehdi Mirza, Bing Xu, S. Ozair†, and Y. B., Aaron Courville, "Generative Adversarial Nets," *arXhiv*, vol. 155, no. 4, pp. 270–275, 2013
12. Saleema, A., & Thampi, S. M. (2018) Voice Biometrics: The Promising Future of Authentication in the Internet of Things. In Handbook of Research on Cloud and Fog Computing Infrastructures for Data Science (pp. 360–389). IGI Global.
13. Lee YS et al (2016) Touch based active user authentication using deep belief networks and random forests. Proc 6th Int Conf Inf Commun Manag ICICM 2016:304–308
14. Maheshwary S, Ganguly S, Pudi V (2017) Deep secure: a fast and simple neural network based approach for user authentication and identification via keystroke dynamics. IWAISe First Int Work Artif Intell Secur 2017:59
15. Shi C, Liu J, Liu H, Chen Y (2017) Smart user authentication through actuation of daily activities leveraging WiFi-enabled IoT. In: Proceedings of the 18th ACM international symposium on mobile Ad Hoc networking and computing – Mobihoc'17, pp 1–10
16. Das R, Gadre A, Zhang S, Kumar S, Moura JMF (2018) A deep learning approach to IoT authentication. In: IEEE international conference communication, vol. 2018–May
17. A. Ferdowsi and W. Saad (2018) Deep learning for signal authentication and security in massive Internet of Things systems, pp 1–30
18. Rajasegarar S, Leckie C, Palaniswami M (2008) Anomaly detection in wireless sensor networks. IEEE Wirel Commun
19. Yavuz, F. Y. (2018) Deep learning in cybersecurity for internet of things (Doctoral dissertation).
20. Diro AA, Chilamkurti N (2018) Distributed attack detection scheme using deep learning approach for internet of things. Futur Gener Comput Syst 82:761–768
21. Kim J, Shin N, Jo SY, Kim SH (2017) Method of Intrusion detection using deep neural network. Int Conf Big Data Smart Comput:313–316
22. Ma T, Wang F, Cheng J, Yu Y, Chen X (2016) A hybrid spectral clustering and deep neural network ensemble algorithm for intrusion detection in sensor networks. Sensors 16(10):1701
23. Kang M, Kang J (2016) Neural network for in-vehicle network security. PLOS One 11:1–17
24. Li, Y., Ma, R., & Jiao, R. (2015) A hybrid malicious code detection method based on deep learning. International Journal of Security and Its Applications, 9(5), 205–216.

25. Niyaz Q, Sun W, Javaid AY, Alam M (2015) A deep learning approach for network intrusion detection system. In: Proceedings of 9th EAI international conference Bio-inspired Information and Communication Technologies
26. Wang Z (2015) The applications of deep learning on traffic identification. Black Hat, Washington, DC
27. Alom MZ, Bontupalli V, Taha TM (2015) Intrusion detection using deep belief networks. In: 2015 National Aerospace & Electronics Conference, pp 339–344
28. Gao N, Gao L, Gao Q, Wang H (2015) An intrusion detection model based on deep belief networks. In: Proceedings – 2014 2nd international conference on advanced Cloud Big Data, CBD 2014, pp 247–252
29. Fiore U, Palmieri F, Castiglione A, De Santis A (2013) Neurocomputing Network anomaly detection with the restricted Boltzmann machine. Neurocomputing:1–11
30. Fiore U, Palmieri F, Castiglione A, De Santis A (2013) Network anomaly detection with the restricted Boltzmann machine. Neurocomputing 122:13–23
31. Dong B, Wang X (2016) Comparison deep learning method to traditional methods using for network intrusion detection. In: 8th IEEE international conference on communication software networks, pp 581–585
32. Sheikhan M, Jadidi Z, Farrokhi A (2012) Intrusion detection using reduced-size RNN based on feature grouping. Neural Comput Appl 21(6):1185–1190
33. Chuan-long Y, Yue-fei Z, Jin-long F, Xin-zheng H (2017) A deep learning approach for Intrusion detection using recurrent neural networks. IEEE Access 5:1–1
34. Lopez-Martin M, Carro B, Sanchez-Esguevillas A, Lloret J (2017) Network traffic classifier with convolutional and recurrent neural networks for internet of things. IEEE Access 5:18042–18050
35. Kim J, Kim J, Thu HLT, Kim H (2016) Long short term memory recurrent neural network classifier for intrusion detection. In: 2016 international conference on platform technology and service, no. February, pp 1–5
36. Bediako PK (2017) Long short-term memory recurrent neural network for detecting DDoS flooding attacks within TensorFlow Implementation framework
37. Kim G, Yi H, Lee J, Paek Y, Yoon S (2017) LSTM-based system-call language modeling and ensemble method for host-based intrusion detection. pp 1–12
38. Cheng M, Li Q, Lv J, Liu W, Wang J (2018) Multi-scale LSTM model for BGP anomaly classification. IEEE Trans Serv Comput, no NetworkML:1–6
39. Putchala MK (2017) Deep learning approach for Intrusion Detection System (IDS) in the Internet of Things (IoT) network using Gated Recurrent neural networks (GRU) p 63
40. Cheng M, Xu Q, Lv J, Liu W, Li Q, Wang J (2016) MS-LSTM: a multi-scale LSTM model for BGP anomaly detection, no. NetworkML, pp 1–6
41. Lopez-martin M, Member S, Carro B (2017) Network traffic classifier with convolutional and recurrent neural networks for Internet of Things. IEE Access 5
42. Vinayakumar R, Kp S, Poornachandran P (2017) Applying convolutional neural Network for Network Intrusion detection, pp 1222–1228
43. Tobiyama S, Yamaguchi Y, Shimada H, Ikuse T, Yagi T (2016) Malware detection with deep neural network using process behavior. In: 2016 IEEE 40th annual computer software and applications conference, pp 577–582
44. R. Pascanu, M. Marinescu, and A. Thomas (2015) Malware classification with recurrent networks. In: IEEE international conference on Acoustics, Speech and Signal Processing – Proceedings, v2015-August, pp 1916–1920
45. Yan W, Yu L (2015) On accurate and reliable anomaly detection for gas turbine combustors : a deep learning approach. PHM Conf:1–8
46. W. C. and P. W. Yao Wang∗ (2016) A deep learning approach for detecting malicious JavaScript code. Secur Commun NETWORKS Secur Comm Networks 2016, 9(22):1520–1534
47. Li Y, Ma R, Jiao R (2015) A hybrid malicious code detection method based on deep learning. Int J Secur Its Appl 9(5):205–216

48. Pascanu R, Stokes JW, Sanossian H, Marinescu M, Thomas A (2015) Malware classification with recurrent networks. In: ICASSP, IEEE internation conference Acoustics speech signal process. – Proceedings, vol. 2015–August, pp 1916–1920
49. Salama MA, Eid HF, Ramadan RA, Darwish A. 103_Hybrid Intelligent Intrusion Detection Scheme.pdf, pp 1–11
50. Saxe J, Berlin K (2015) Deep neural network based malware detection using two dimensional binary program features
51. Jung W, Kim S (2015) Poster: deep learning for zero-day flash malware detection. Proc IEEE Symp Secur Priv:2–3
52. Seok S, Kim H (2016) Visualized malware classification based-on convolutional neural network. J Korea Inst Inf Secur Cryptol 26(1):197–208

K. S. Sunitha Krishnan received her M.Tech degree in Computer Science, with specialization in Software Engineering, from the Department of Computer Science, Cochin University of Science and Technology in 2011 and B.Tech degree in Computer Science and Engineering from Mahatma Gandhi University in 2004. Her research interests include Machine Learning, Cybersecurity, Intrusion Detection, data analytics and Internet of Things (IoT).

Sabu M. Thampi is a Professor at the Indian Institute of Information Technology and Management, Kerala (IIITM-K), Trivandrum, India. He has completed his Ph.D in computer engineering from the National Institute of Technology, Karnataka. His research interests include network security, security informatics, bio-inspired computing, video surveillance, cloud security, secure information sharing, secure localization, and distributed computing. He has authored and edited few books published by reputed international publishers and published papers in academic journals and international and national proceedings. He is currently serving as Editor for Journal of Network and Computer Applications (JNCA), Elsevier and Journal of Applied Soft Computing, Elsevier; and Associate Editor for IEEE Access and International Journal of Embedded Systems, Inderscience, UK; and reviewer for several reputed international journals. He is a Senior Member of IEEE and member of IEEE Communications Society, IEEE SMCS, and ACM.