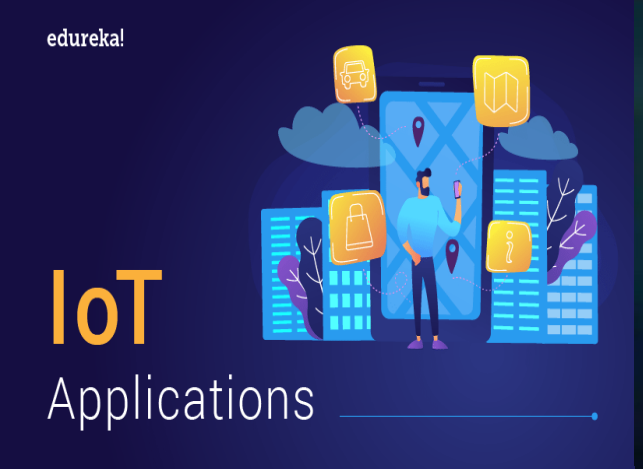


“Internet of Things Applications, Security Challenges, Attacks, Intrusion Detection, Future Visions: A Systematic Review”

Author: N. Mishra, and S. Pandya

Presenter:
Sushil Kumar Singh (수실 쿠마르 싱)
Supervisor: 박종혁 교수
2022.11.07
sushil.sngh001007@seoultech.ac.kr
Seoul National University of Science and Technology, Seoul, South Korea

Table of the Contents

Source: https://www.pinterest.co.kr/pin/512777107566262032/

Abstract

- Internet of Things (IoT) technology is prospering and entering every part of our lives, be it education, home, vehicles, or healthcare.
- With the increase in the number of connected devices, several challenges are also coming up with IoT technology: heterogeneity, scalability, quality of service, security requirements, and many more.
- Security is also the major issues in IoT applications as makes IoT vulnerable to security attacks, and give financial, reputational losses.
- In this paper, the authors survey of different security issues in IoT layers, focus on Distributed Denial of Service (DDoS) attacks.
- The presented review work compares Intrusion Detection and Prevention models for mitigating DDoS attacks and focuses on Intrusion Detection models.
- Furthermore, the classification of Intrusion Detection Systems, different anomaly detection techniques, different Intrusion Detection System models based on datasets, various machine learning and deep learning techniques for data pre-processing and malware detection.

1. Introduction

- Internet of Things (IoT) is an emerging field of collection and transfer of data without human intervention.
- It is referred to as a system of connected objects embedded with sensors, software, control systems.
- The healthcare sector has transformed with the introduction of IoT, be it wearables or telemedicine and remote monitoring of patient.
- IoT has been a game-changer in smart vehicles by introducing connected vehicles.
- Resource constraint, mobility, heterogeneity, scalability, data management, security, and privacy are the main challenges in IoT applications.

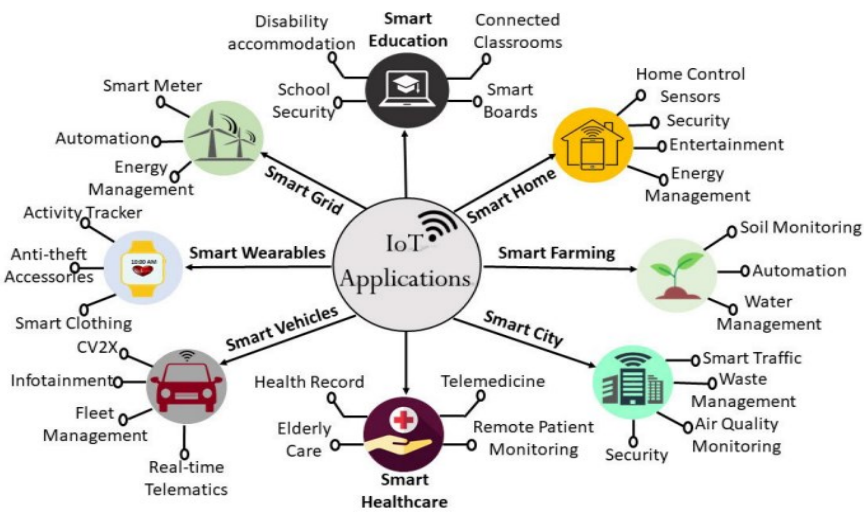

Fig. 1 : A representation of various IoT Applications

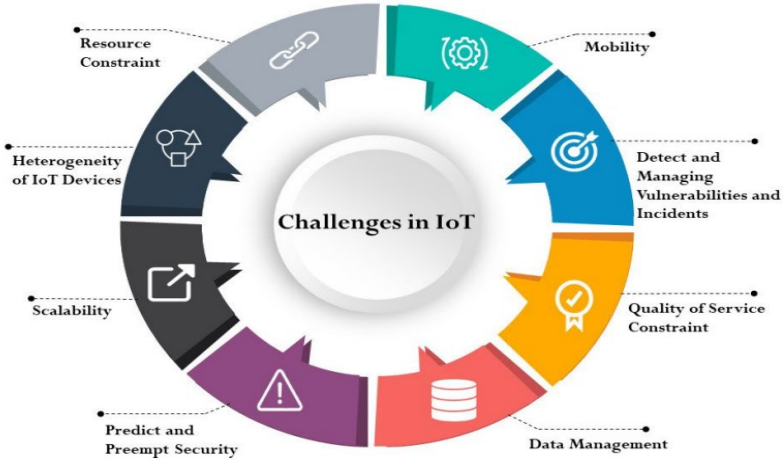

Fig. 2 : Challenges in IoT

1. Introduction

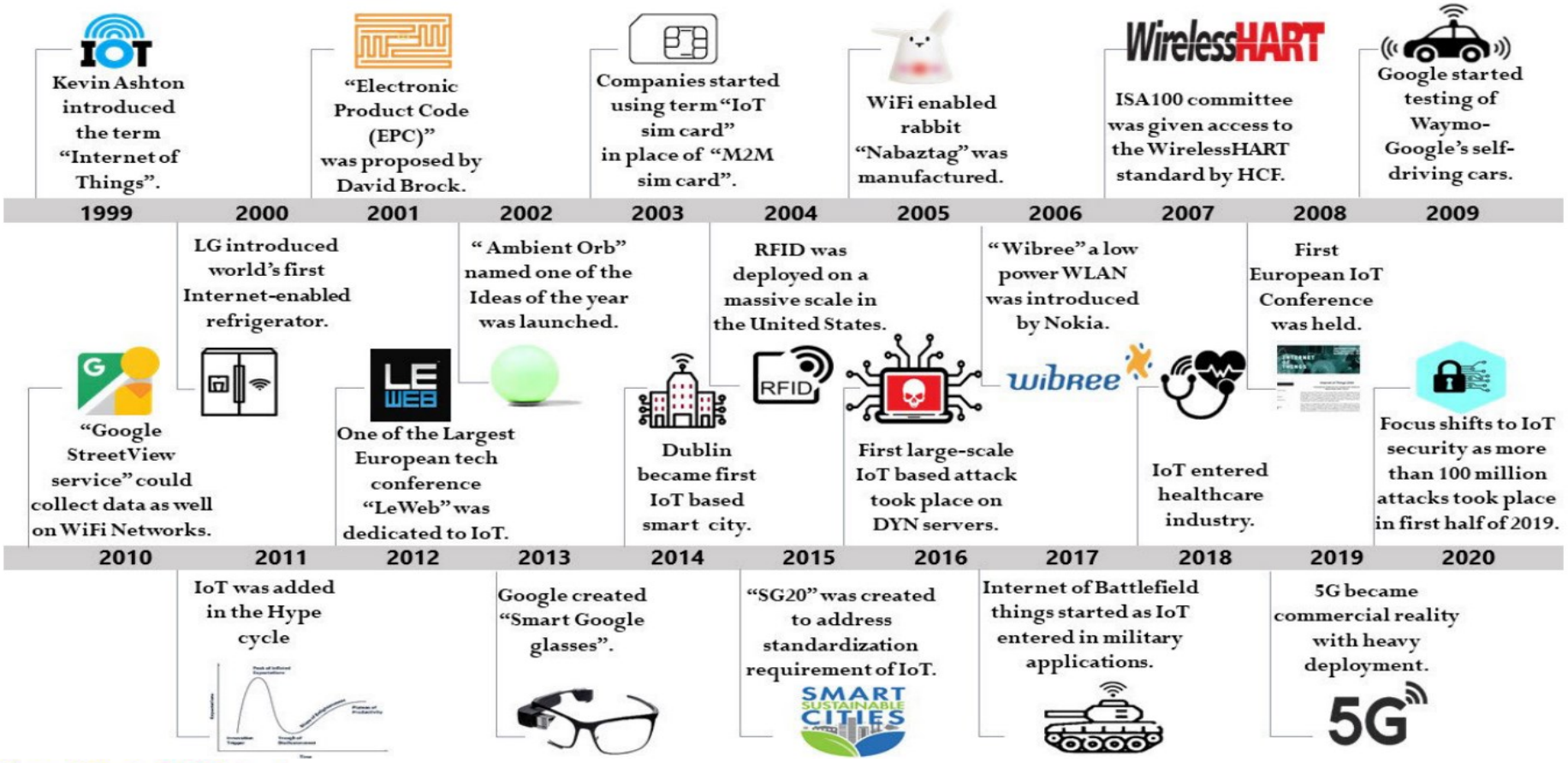

Fig. 3 : A chronological representation of the evolution of IoT technologies from 1999-2020

1. Introduction

TABLE 1. IoT research challenges.

Key Reference	Year	Research Challenges	Discussion
Pal *et al.* [13]	2018	Standardization and Regulatory Framework	Regulations and standardization are required for data ownership as data handling can even involve legal obligations in some cases like medical data.
Srivastava *et al.* [14]	2020	Security Requirements	Securing IoT devices becomes all the more difficult due to the range involved with different IoT devices and also varying issues of different IoT layers.
Pal *et al.* [13]	2018	Interoperability	Due to the diversity associated with heterogeneity involved in IoT, standard interfaces are significant for maintaining interoperability.
Ding *et al.* [15]	2020	Connectivity	Connectivity solutions are both licensed and unlicensed. Thus arises the need for a standard connectivity solution to address the decision-making issue of using specific connectivity solutions for IoT.

1. Introduction

TABLE 2. A detailed comparison of state-of-the-art surveys in the IoT security domain.

Authors	Year	Contribution	1	2	3	4	5	6
Yang *et al.* [16]	2017	The survey inspects four IoT security aspects: limitation of IoT devices and solutions, classification of IoT attacks, IoT authentication, and security attacks in different IoT layers.	✓	✓	×	×	×	×
Yu *et al.* [17]	2017	Edge computing and its use in IoT is thoroughly analyzed. Advantages and disadvantages associated with edge computing-based IoT are discussed.	✓	×	✓	×	×	×
Kouicem *et al.* [18]	2018	A top-down survey of IoT security solutions is conducted with focus on security solutions addressing resource constraints and scalability issues.	✓	✓	✓	×	×	×
Frustaci *et al.* [19]	2018	Different security issues and the availability of solutions for these issues are discussed in detail. Security issues raised due to communication protocols are also discussed.	✓	✓	✓	×	×	×
Noor *et al.* [20]	2019	New technologies related to IoT security, along with tools and simulators, are discussed in depth.	✓	×	×	×	×	×
Hassija *et al.* [21]	2019	A detailed review of security-related issues in IoT and discussion on emerging technologies for building a high trust level is presented.	✓	✓	✓	×	✓	×
Meneghello *et al.* [22]	2019	Security issues of different communication protocols and solutions are analyzed, particularly the weakness of commercial IoT solutions are discussed.	✓	×	✓	×	×	×
Srivastava *et al.* [14]	2020	Discussion on detection and defense against DDoS, Sybil, collusion attacks is presented along with different Intrusion Detection strategies.	✓	✓	✓	×	×	×
Anand *et al.* [23]	2020	Vulnerabilities associated with IoT in the backdrop of sustainable computing are analyzed, and a multifold study is presented, accompanied by a case study on smart agriculture.	✓	✓	✓	×	✓	×
The Proposed Survey	2021	Evolution of IoT, applications, and challenges associated with IoT, security issues in IoT are presented. Different types of DDoS attacks, their impacts, solutions, and anomaly detection are discussed in detail.	✓	✓	✓	✓	✓	✓

1. IoT Security issues, 2. DDoS attacks discussion, 3. Intrusion Detection System, 4. Database discussion for IDS, 5. Machine Learning Techniques for IDS, 6. Deep Learning Techniques for IDS

2. Security Issues in IoT Domain

Perception Layer: (Sensors, Actuators)

- Internet of Things (IoT) devices are increasingly growing in numbers, and lack of security in these devices has resulted in transforming IoT devices into a hotbed for malicious activities.
- Sensors are also known as nodes, and these are vulnerable to node capturing attacks where an attacker may either capture the node or replace it with a malicious node.
- Side channel attack based on laser, power consumption, and timing can occur in this layer.
- IoT devices are power constraint, and the attackers exploit this issue by draining the power source and causing Sleep deprivation.

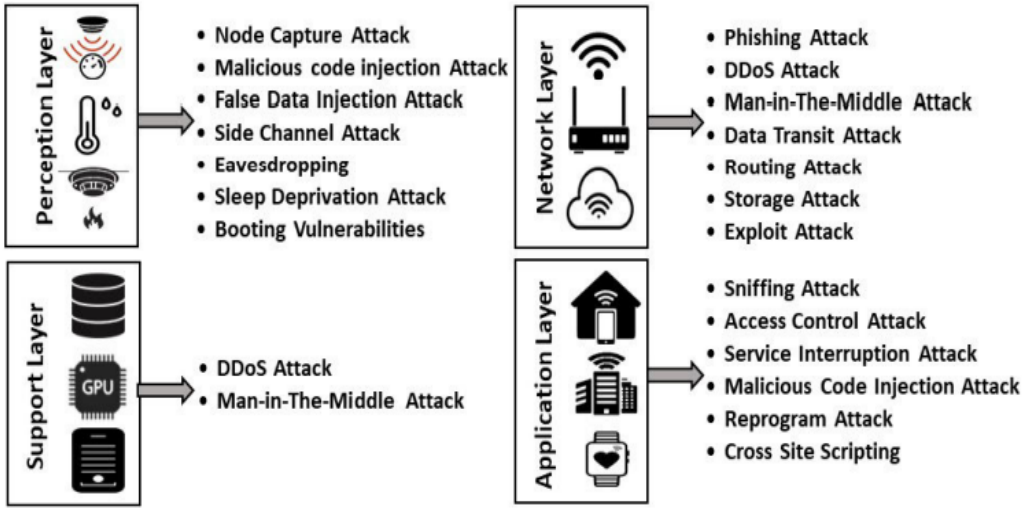

FIGURE 6. An illustrative representation of various security attacks in different IoT layers.

2. Security Issues in IoT Domain

Network Layer: (Communication)

- Phishing attack targets several IoT devices in an attempt to at least take control of a few of them.
- In a DDoS attack, an attacker tries to overwhelm the target by sending spoofed requests. IoT devices act as botnets in DDoS attacks and can create a massive flood of requests to deny the target further access to resources.
- Worm-hole, Sinkhole attacks are examples of Routing attacks in which the attacker tries to route the traffic to a different path by gaining access to nodes.

Support Layer: (Resource Allocation, Computing)

- In a Man-in-the-middle attack, the attacker takes control of the broker, thus controlling all the communication.
- The target of attack in the Support layer is usually to access data; therefore, database and cloud security are crucial in this layer.

Application Layer: (Smart home, healthcare,…)

- A service interruption attack is similar to a denial-of-service attack as it causes service disruption.
- Sniffing attack takes place with the help of sniffing tools where attacker sniffs network traffic data, and confidential data is compromised in this attack.

2. Security Issues in IoT Domain

A. Distributed Denial of Services (DDoS) Attacks

- Distributed Denial of Services (DDoS) is an amplified DoS attack. In a DDoS attack, requests are initiated from many sources, and hence it is named as distributed DoS. Due to this, it becomes challenging to mitigate DDoS attacks.
- TCP SYN Flood attack, Teardrop attack, Smurf attack, Ping of Death attack, Botnets are the types of DDoS attacks.
- DDoS attacks can also be classified as Reflection and Amplification attack.
- In a reflection attack, the size of the request and response is the same whereas, in an amplification attack, the size of the response is many times bigger than that of the request.

TABLE 6. A representation of the evolution of DDoS attack vectors.

Key Reference	Attack Vector with Year	Amplification Factor	Working Methodology
Hesselman *et al.* [43]	DNS (2013)	28-54	The functionality of open DNS resolvers is used to launch the attack.
Kawamura *et al.* [44]	NTP (2014)	556.9	The monlist command enabled NTP servers are abused to launch the attack.
Gondim *et al.* [45]	SSDP (2014)	30.8	UPnP protocol is exploited in SSDP reflection attacks.
Vasques *et al.* [46]	SNMP (2014)	6.3	Directly exposed servers with SNMP service are exploited under this attack.
Wisam *et al.* [47]	RIPv1 (2015)	131.2	Routers running RIPv1 with multiple routes are exploited in this attack.
Burch *et al.* [48]	CHARGEN (2015)	358.8	Internet-enabled devices running CHARGEN can be exploited to launch amplified attacks.
Noor *et al.* [49]	NetBIOS (2015)	3.8	Servers with open NetBIOS service are exploited for DDoS attacks.
Sieklik *et al.* [50]	TFTP (2016)	60	The protocol was intended for file transfers; its simple design omitted authentication capabilities and was exploited for the attack.
Choi *et al.* [51]	CLDAP (2017)	56-70	The gaming industry was the most affected by this attack. Servers with open UDP port 389 were targeted.
Agathe *et al.* [52]	Memcached (2018)	10000-51000	Attackers target Memcached servers with open TCP and UDP ports on 11211 to launch a DDoS attack.
Kondoro *et al.* [53]	CoAP (2018)	34	UDP garbage flood is created using IoT devices as amplifiers for launching this attack.
Malaimalavathani *et al.* [54]	WS Discovery (2019)	10-500	Being a UDP-based protocol, attackers use this to launch UDP flood attacks.
Bjarnason *et al.* [55]	ARMS (2019)	45	Operational management of Apple Remote Desktop (ARD) protocol running on UDP port 3283 was used to launch a DDoS amplification attack.

3. Intrusion Detection System

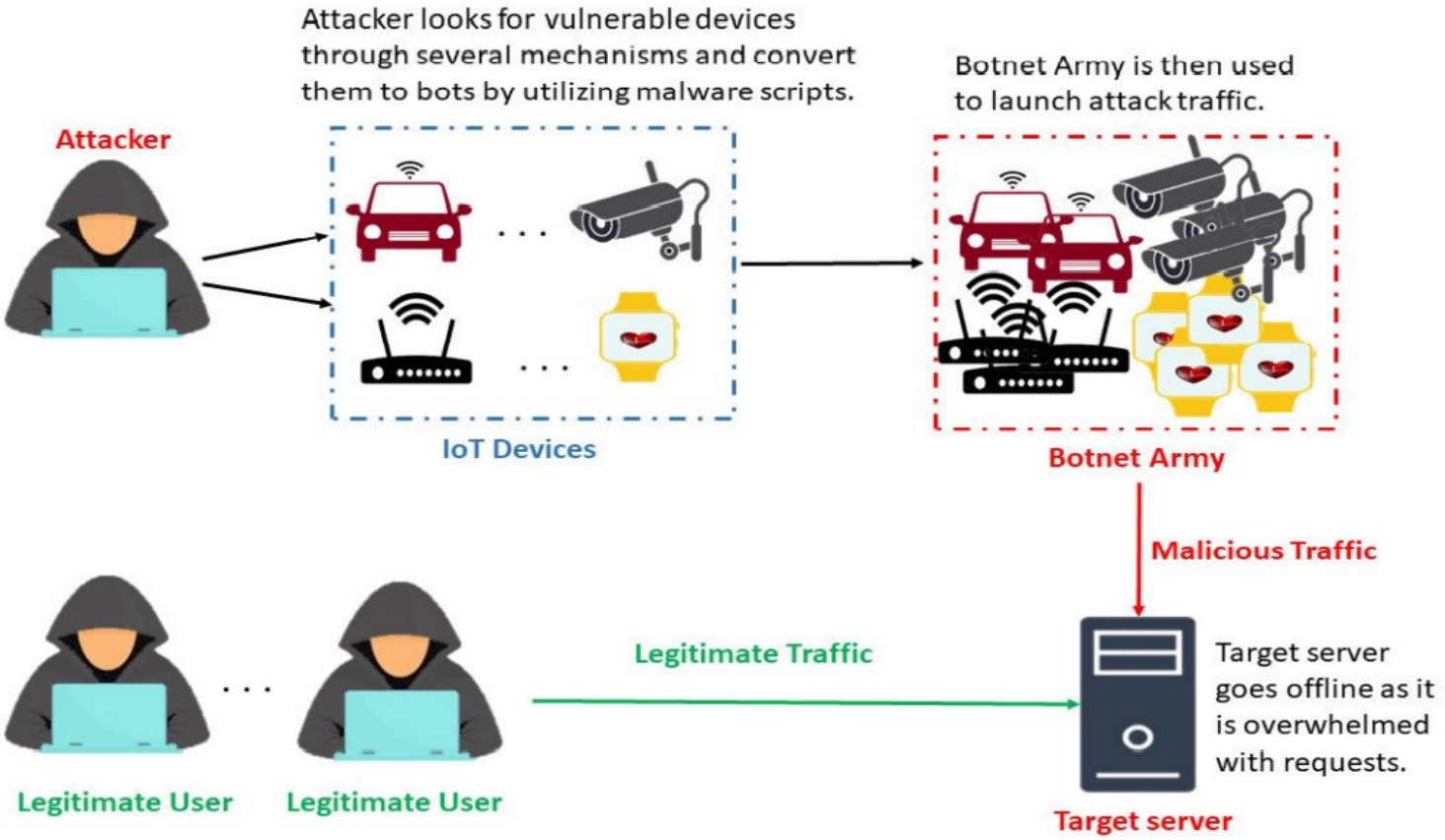

Fig. 8 : A Representation of attacker gaining access to IoT devices and launching DDoS Attacks

3. Intrusion Detection System

- Intrusion Detection System (IDS), Intrusion Prevention System (IPS) are the two solutions for mitigate the DDoS attacks.
- IDS is a precautionary measure where the system itself takes no action in case of intrusion; instead, an alarm is raised
- IPS is the punitive measure where an action is taken by the system in case of intrusion.
- In IPS, an issue arises in the case of false positives as legitimate users can also get blocked.
- Host-based IDS is specific to a system, detection of an inside intruder is strong, and it can very well assess the extent of the compromise, but it is expensive as one IDS is required per-host.

TABLE 8. Comparative analysis of IDS And IPS systems.

Intrusion Detection System (IDS)	Intrusion Prevention System (IPS)
IDS recognizes the threat and monitors the system.	IPS is a regulatory system as it takes monitors and defends the system.
Human intervention is required for action.	IPS takes action based on the rule set, and no human intervention is required.
IDS does not impact system performance.	IPS may slow down the system.
False alarm rate does not impact performance to the same extent as that for IPS.	False alarm rate is of high concern.
Legitimate users are not blocked as the system does not take action.	Legitimate traffic might be blocked due to false alarms.

- In Network-based IDS, the outside intrusion is very well detected, and it can protect all hosts, but there is too much traffic to analyze the DDoS.
- Hybrid IDS is flexible and provides more security as it combines features of both Host-based and Network-based IDS.

3. Intrusion Detection System

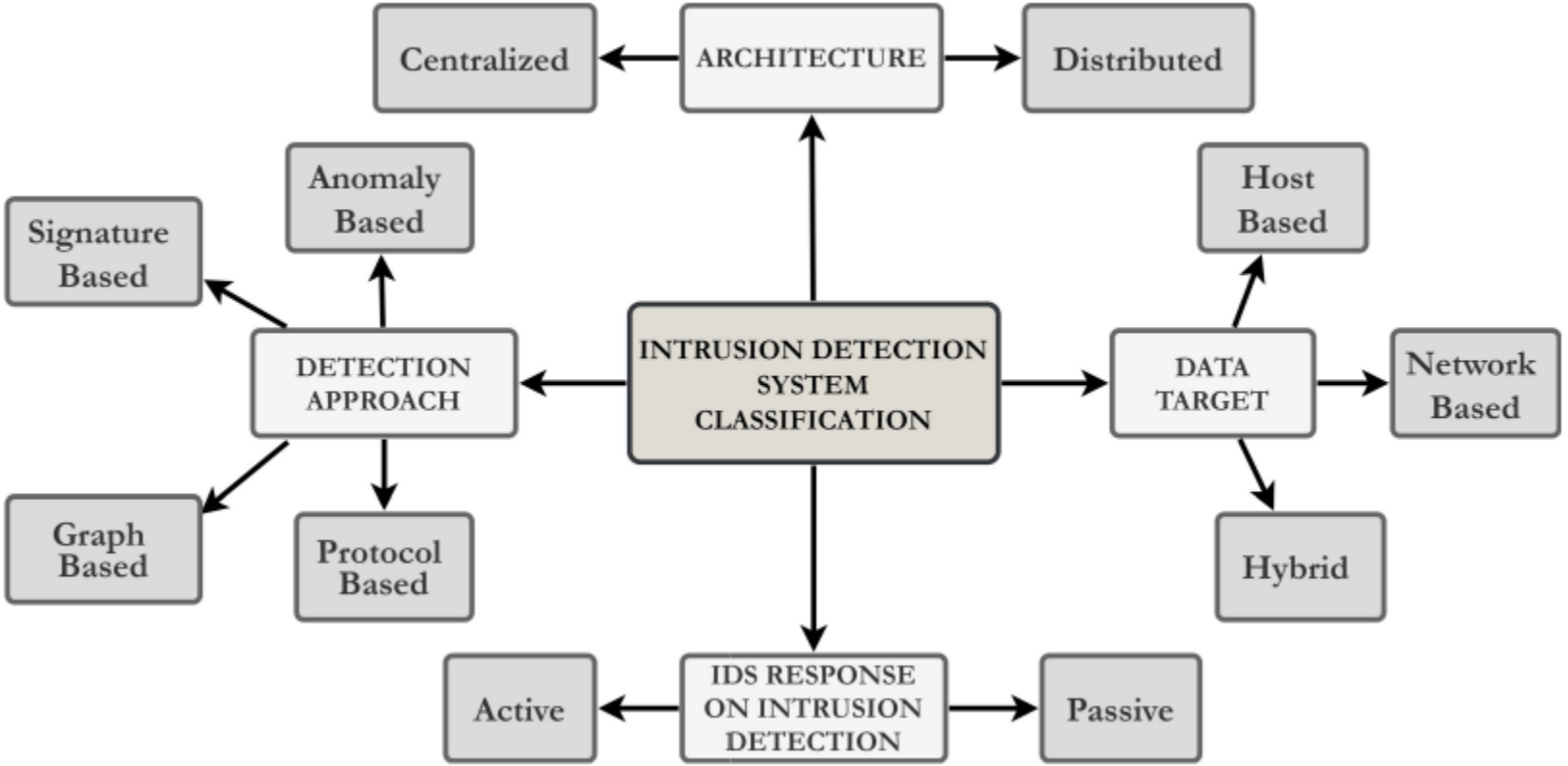

Fig. 9 : A graphical representation of the classification of various IDS techniques

3. Intrusion Detection System

A. Anomaly Detection Techniques

- Two major approaches being used for Malware Detection are Signature-based Detection and Anomaly-based Detection.
- Signature based detection technique is not successful in real-world applications (Botnets as Botnets keep on mutating and Bot signature also keeps on changing.
- Community Base Anomaly Detection where Bots are identified using Communication Graph.
- Bad Neighborhood is also one of the methods used in Spam and Phishing Detection; it is defined as a cluster of IP addresses that perform malicious activities over a certain period. Moura et al. [69] used this approach for IPv4 attacks and generated a blacklist of IPs.
- This approach is not entirely practical as DDoS attacks are widespread, and it is challenging to assign clusters for blacklisting.
- Another method is whitelisting IPs, as Yoon [76], where a VIP list is created assuming that VIPs will log in from a particular IP address, i.e., IP address not very dynamic for personal laptops.

3. Intrusion Detection System

TABLE 9. Comparative analysis of anomaly detection techniques.

Key Reference	Technique	Examples	Approach	Advantage (+) / Disadvantage (-)
Hai *et al.* [70]	Machine Learning Model	Neural Networks, Genetic Algorithms, Clustering, Classification, Outlier Detection	• Machine Learning models have two stages: Training and Testing. • Broadly it can be divided into two categories: Supervised Learning and Unsupervised Learning.	+ Can identify patterns quickly. + Wide application range. + Suitable for online datasets. + Can improve continuously. - Long training time is required. - Larger dataset is needed for better results.
Kalkan *et al.* [71]	Statistical Model	Univariate, Multivariate, Time Series, Markov Process	• The statistical model approach depends on mathematical calculations. • Model is created for normal behavior from the historical data of the user.	+ Model is simple. - Model depends heavily on statistical or mathematical modeling, thus affects accuracy.
Kim *et al.* [72]	Payload Based Model	Grained Model, N-gram analysis,	• The Payload-based approach model learns characteristics of the normal packet payload, and deviation is considered anomalous behavior.	+ Model works very well for known attacks. - Longer handling time because of computational overhead. - Lesser accuracy is achieved for new attacks.
Kumar *et al.* [73]	Rule-Based Model	Association Rule, Fuzzy Rule, Behavior Rule	• In a Rule-based model, rules are created from data traffic patterns. • If the rule is broken, it is considered anomalous behavior.	+ Model is simple. - Longer monitoring of traffic is required for rule creation. - High false-positive rate.
Choudhary *et al.* [74]	Protocol Based Model	Application protocol, Communication Protocol	• The protocol-based model works with monitoring protocols at different layers. • Computing techniques are used to identify anomalies associated with a particular protocol.	+ Detection accuracy is high for a particular type of attack. - Suitable for specific attack type so fails for other attacks.

3. Intrusion Detection System

		Actual: Positive	Actual: Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

A. Performance Metrics for IDS

1) ***Confusion Matrix:*** It is instinctive metrics for defining a classification model's correctness. There are two ways to reduce errors: reducing False Negatives and reducing False Positives. There is no set rule for the same, and it depends on the requirement.

2) ***Accuracy:*** Accuracy is defined as the number of correct predictions over total predictions. $Accuracy = \frac{TP+TN}{(TP+TN+FN+FP)}$

3) ***Precision:*** Precision is a measure to calculate the Machine Learning Model's accuracy in finding the number of actual positives out of total predicted positives. Precision $= \frac{TP}{(TP+FP)}$

4) ***Recall/ Sensitivity:*** Recall is a measure to calculate the Machine Learning Model's accuracy in finding the number of positives out of total actual positives. **Recall** $= \frac{\mathbf{TP}}{\mathbf{(TP+FN)}}$

5) ***F1 Score:*** It is calculated as a Harmonic Mean of precision and recall metrics to better evaluate model performance.

$$\text{F1 Score} = \frac{2 * Precision * Recall}{(Precision + Recall)}$$

3. Intrusion Detection System

		Actual: Positive	Actual: Negative
Predicted	Positive	True Positive (TP)	False Positive (FP)
	Negative	False Negative (FN)	True Negative (TN)

A. Performance Metrics for IDS

6) ***Specificity:*** Specificity is the opposite of Sensitivity (Recall), and it is a measure of False Positive Rate.

$$\text{Specificity} = \frac{TN}{(TN + FP)}$$

7) ***AUC-ROC Curve:*** Receiver Operating Characteristic (ROC) curve is a measure to determine the stability between precision and recall by a varying threshold. The Area Under Curve (AUC) represent the quality of the classification model.

$$\text{True Positive Rate (TPR)} = \frac{TP}{(TN + FN)}$$

$$\text{False}\mathit{Positive\ Rate}(FPR) = \frac{FP}{(FP + TN)}$$

It is a curve between TPR, i.e., Recall (Sensitivity), and FPR, i.e., (1 – Specificity). In general, AUC near to one represents a better classification model.

4. Review of Steps Involved in IDS

A. Data Sets in IDS

- Data can be collected in two ways in IDS: 1) Creating own datasets 2) Existing Datasets
- KDD-99 [77] is one of the oldest and extensive dataset, and despite it being highly imbalanced, it is used even now due to the lack of its alternatives.
- NSL-KDD [78] was created to remove the issues associated with KDD-99. It is one of the benchmark datasets used for Anomaly detections.
- The skewedness of KDD-99 and NSL-KDD is almost removed in UNSW-NB15 [79], [80], consisting of 49 features and 10 target classes, whereas KDD consists of 41 features and 5 target classes.
- For Botnets, CTU13 [81] having 13 scenarios; each of different Botnet samples is being used nowadays.
- ISOT is also one of the popular datasets [82], particularly for IoT Botnet attack databases.
- CICFlowmeter [96] is a java based tool used for extracting network features from raw network captures. It captures a set of 80 features and prepares a pcap or CSV file to be used for further analysis.

4. Review of Steps Involved in IDS

A. Data Sets in IDS

TABLE 10. A detailed review of various network-based datasets.

Datasets	Year	Dataset Publically Available	Nature of Traffic		Nature of Data	Labeled Dataset	Balanced Dataset	Network Type	Traffic Type
			Normal	Attack					
DARPA [83], [84]	1998-99	Yes	✓	✓	Packet, Logs	✓	×	Small Network	Emulated
KDD-99 [77]	1998-99	Yes	✓	✓	Other	✓	×	Small Network	Emulated
UNIBS [85]	2009	On Request	✓	×	Data Flows	×	×	University Network	Real
CDX [86]	2009	Yes	✓	✓	Packet	×	×	Small Network	Real
ISOT [82]	2010	Yes	✓	✓	Packet	✓	×	Small Network	Emulated
ISCX [87]	2012	Yes	✓	✓	Packet, Bidirectional Flows	✓	×	Small Network	Emulated
CTU-13 [81]	2013	Yes	✓	✓	Uni and Bidirectional Flows	✓	×	University Network	Real
Botnet [88]	2014	Yes	✓	✓	Packet	✓	×	Various Network	Synthetic
UNSW-NB15 [79]	2015	Yes	✓	✓	Packet, Other	✓	×	Small Network	Emulated
AWID [89]	2015	On Request	✓	✓	Other	✓	×	Small Network	Emulated
NDSec-1 [90]	2016	On Request	×	✓	Packet, Logs	✓	×	Small Network	Emulated
CICDOS [91]	2017	Yes	✓	✓	Packet, Bidirectional Flows	✓	×	Small Network	Emulated

4. Review of Steps Involved in IDS

A. Machine Learning Techniques in IDS

- Data pre-processing comprises several steps: adding missing values, normalizing data, removing unwanted features/outliers. Thus, Feature analysis and extraction is main for any machine learning model.
- For feature extraction, different optimization techniques are used by researchers: Principal Component Analysis (PCA), Genetic Algorithms (GA), and Boosting Algorithms.
- Botnet analysis has two major subdivisions, specifically Flow-based traffic analysis and Graph-based traffic analysis.
- These analyses differ mainly on the feature selection part as statistical features are selected for Flow-based analysis; otherwise, Graph-based features are chosen.
- A botnet can be Detected using Graph-based features, as done by Chowdhury et al. [101]. In this detection, efficiency was improved by removing inactive nodes, and detection methodology was given, where, by using only six nodes, Bots can be detected effectively.

4. Review of Steps Involved in IDS

A. Machine Learning Techniques in IDS

TABLE 11. A detailed review of various machine learning-based data pre-processing techniques.

Authors	Dataset	Technique	Discussion
Lin *et al.* [102]	KDD-99	Feature Engineering - Cluster Center and Nearest Neighbor Classifier – K-NN	This algorithm performs better than K-NN and SVM in terms of performance metrics and computational efficiency for testing and training time.
Bijalwan *et al.* [103]	ISCX	Feature Engineering – Dataset is segregated into normal and attack classes Classifier – Ensemble Classifier	The use of Ensemble Classifier provides better results than a single classifier.
Alejandre *et al.* [104]	1. ISOT 2. ISCX	Feature Engineering - Genetic Algorithm Classifier – C 4.5 algorithm	A genetic algorithm was used as an optimizer, and because of this higher detection rate is achieved.
Garg *et al.* [105]	1. NSL-KDD 2. Kyoto	Feature Engineering - 1. Horizontal Feature Selection - Infinite Feature Selection 2. Vertical Feature selection – Abridging Algorithm Classifier – SVM	This analysis helps understand the effect of feature selection and can be used to reduce execution time.
Chellammal *et al.* [106]	1. KDD-99 2. NSL-KDD 3. Koyoto 2006	Feature Engineering – Correlation Detection - Ensemble Learning	Data is partitioned by segregating majority and minority classes and creating multiple datasets by sampled data.
Devan *et al.* [107]	NSL-KDD	Feature Engineering - XGBoost Classifier – Deep Neural Network	Features are selected analytically, and therefore, results are also good, but the drawback is optimal learning rate is chosen from experience, not analytically.
Rajadurai *et al.*[108]	NSL-KDD	Feature Engineering - PCA Classifier – Deep Learning model	PCA retains significant features, thus giving better results. This work is suited for detecting known attacks.
Li *et al.* [109]	Two data subsets from VirusShare, Four data subsets from VXHeavens	Feature Engineering - Random forest algorithm, Feature grouping Classifier - RMSE is calculated using Autoencoder and classified using Kmeans	Because of the three-layer neural network structure, it is efficient and lightweight. Autoencoder technique can efficiently solve the sample imbalance problem.
Khare *et al.* [110]	NSL-KDD	Feature Engineering - Min-Max Normalization, 1-N Encoding, Spider Monkey Optimization Classifier – Deep Neural Network	The use of nature-inspired algorithms for dimensionality reduction reduces the issues of quantity and quality of high dimensional data.

4. Review of Steps Involved in IDS

A. Machine Learning Techniques in IDS

TABLE 12. A detailed review of various machine learning-based Malware detection techniques.

Authors	Dataset	Technique	Discussion
Meidan *et al.* [97]	Data is collected from nine IoT devices. Alexa Rank and GeoIP are used for enriching the dataset.	Feature Engineering – Twenty-three sets of features were extracted from five time windows. Classifiers -XGBoost, RandomForest, GBM	IoT and Non-IoT devices are classified using ML classifiers. This can be utilized for finding unauthorized IoT devices also.
Prokofiev *et al.* [124]	Data from 100 botnets is collected.	Feature Engineering - Logistic Regression Detection – Logistic Regression	Model is created to identify if a connection initiating device is running a bot.
Mazini *et al.* [125]	1. NSL-KDD 2.ISCXIDS2012	Feature Engineering - Artificial Bee Colony Evaluation – AdaBoost	The complexity of the model is less, and because of boosting algorithm, performance is also good.
Bezerra *et al.* [126]	Mirai, Bashlite, Hajime, Aidra, Tsunami, Dofloo are botnets used for attacks.	Feature Engineering – Scaling and Normalization Classifier - Elliptic Envelope, Isolation Forest, Local Outlier Factor, One-class Support Vector Machine	This work is mainly valuable when botnet details are end-to-end encrypted.
Khan *et al.* [127]	Five botnets were used for attacks; Wireshark was used to obtain CSV files.	Feature Engineering - Wrapper Method Evaluation – 10 fold cross-validation	Detection time is not mentioned, and the decision tree algorithm's depth is kept at eight, and the classification tree is set to 100 without explanation.
Djanie *et al.* [98]	The attack is launched using eight DoS attack tools. Wireshark is used to capture and display network traffic.	Feature Engineering – Manual Normalization Detection - SVM classifier	Snort IDS is used for testing, and a high detection rate is achieved.
Wang *et al.* [128]	Data is simulated using five new Botnets, namely Zues, Athena, Mirai, Ares, and Black Energy	Feature Engineering – Statistical and graph-based features are extracted. Evaluation - K means clustering, least-square technique, and Local Outlier Factor	In this study, a hybrid of both flow-based and graph-based detectors is used hence performs better than individual detectors.

4. Review of Steps Involved in IDS

B. Deep Learning Techniques in IDS

- Deep learning techniques are being extensively used for feature engineering because of their ability to learn high-dimensional features.
- Generative Adversarial Network (GAN) is one of the most common feature engineering techniques, specifically for their application in synthetic data creation and learning better about minority classes.
- Ferdowsi et al. [131] used GAN for feature engineering as well as detection. In this study, a distributed GAN is proposed to provide a fully distributed IDS for the IoT to detect anomalous behavior without reliance on any centralized controller.
- Lee et al. [139] by deploying GAN for feature engineering. Features engineering is achieved using the Flow Wasserstein GAN model and Attention GRU Model by Han et al. [140].
- An attention model is used to detect the payload-based attack.
- Yang et al. [141] used a Supervised adversarial Variational autoencoder for feature engineering. Regularization is achieved using Wasserstein GAN with gradient penalty.

4. Review of Steps Involved in IDS

B. Deep Learning Techniques in IDS

TABLE 13. A detailed review of various deep learning-based data pre-processing techniques.

Authors	Dataset	Technique	Discussion
Erfani *et al.* [112]	An experiment is done on six real and two synthetic datasets.	Feature Engineering - Deep Belief network Classifier - One-Class SVM	This Model is faster than a deep autoencoder with comparable results. The linear kernel can be used as this model is scalable.
Sun *et al.* [132]	Mimicking attack is generated using LSGAN and GAN.	Feature Engineering - LSGAN and GAN	LSGAN and GAN are compared, results obtained could not clearly establish the need to use LSGAN in place of GAN.
Ma *et al.* [133]	1. ISCX-IDS-2012 2. CIC-IDS-2017	Feature Engineering – 1. 1D CNN: sequence features 2. Deep Neural Network: statistical and environmental features Classifier - Neural Network	A hybrid solution is given for feature selection, and a Shallow neural layer is used for anomaly detection.
Huang *et al.* [134]	1. NSL-KDD 2. UNSW-NB15 3. CIC-IDS-2017	Feature Engineering - Imbalanced data filter and convolutional layers are added to GAN.	In this study, by conducting several experiments, it is observed that synthesized samples are necessary for better performance, especially the ones of minority classes.
Saraeian *et al.* [135]	1. NSL-KDD 2. ISCXIDS 2012	Feature Engineering and detection-Convolutional Neural Network	The deep learning techniques have stronger learning ability which is intuitive from achieved higher accuracy.
Merino *et al.* [136]	KDD-99	Feature Engineering - Generative Adversarial Network Classifier - Neural Network Binary Classifier	In this study, the quality of data generated using GAN is ensured by using a neural network for evaluating model.
Manimurugan *et al.* [137]	CICIDS 2017	Feature Engineering – Duplicating technique Classifier - Deep Belief Network	A greedy layer-wise training algorithm was used to train DBN one layer at a time. Minority samples were combined together to avoid getting misclassified as benign.
Kim *et al.* [138]	1. KDD-99 2. CSE-CIC-IDS2018	Feature Engineering – Numerical Samples are converted to RGB and grayscale images. Classifier – Convolutional Neural Network	Study shows that RGB images in both binary and multiclass classifications have higher accuracy than grayscale images.

4. Review of Steps Involved in IDS

B. Deep Learning Techniques in IDS

TABLE 14. A detailed review of various deep learning-based data Malware detection techniques.

Authors	Dataset	Technique	Discussion
Yousefi-Azar *et al.* [145]	Malware is collected from different sites like virustotal.	Feature Engineering - Hashing algorithm viz. tf -simhashing Classification - Novel extreme learning model	It extracts static features of any given binary file to distinguish malware from benign; hence, it helps mitigate the zero-day attack.
McDermott *et al.* [146]	A labeled dataset was created for four attack vectors (UDP, ACK, DNS, SYN Flood)of the Mirai botnet.	Feature Engineering - BLSTM-RNN with word Embedding Detection - Bidirectional LSTM-RNN	Dataset was generated in this study. Although results are promising but processing time is more, and for comparison, out of ten attack vectors, only four were considered.
Meidan *et al.* [97]	Mirai and Bashlite are used to infect devices. Data is collected using IoT devices; for sniffing, Wireshark is used.	Feature Engineering – Manual Normalization Classification - Deep Autoencoder	Back propagation is used, so as is the case for deep learning algorithms, the time taken for detection is more.
Pektaş *et al.* [147]	Six types of Botnets were used for the attack. Also, CTU 13 and ISOT were used as Benchmark datasets.	Feature Engineering - Graph structure is used to extract statistical-based network flow features. Detection - Convolutional and Recurrent neural network.	In this study, execution time is high even though higher configuration hardware is used.
Yeo *et al.* [148]	Nine different malware datasets from Stratosphere IPS are used.	Feature Engineering - Netmate Classification - CNN, MLP, SVM & RF	No analysis is given for the selection of parameters for models.
Ghasemi *et al.* [65]	KDD 99 and NSL KDD are used, and a new dataset is created based on five different labels using a Genetic Algorithm.	Feature Engineering - Genetic Algorithm Classification - Kernel Extreme Learning Machine	This model is trained for different behavior of all attacks individually; hence its performance is good.
Jahromi *et al.* [149]	VXHeaven, Kaggle, Windows ransomware, IoT malware, and a combined dataset of ransomware and IoT malware samples.	Feature Engineering – Not Required Classification - Novel extreme learning Machine model	Back propagation is avoided for training the network; thus, i is very fast compared to other approaches.
Qureshi *et al.* [150]	A subset of KDD dataset is used.	Feature Engineering - Pre-trained network on regression-related task is used for feature extraction. Detection - Novel Incremental SVM technique RS-ISVM	Oscillation problem of traditional SVM is reduced by retaining old samples, which are likely to become support vectors.
Kim *et al.* [151]	CTU-13	Feature Engineering - Manual sampling of features Detection - Recurrent Variational	At every time window, anomaly scores of every flow are calculated, which provides the degree of maliciousness of individual connections.

5. Security-based Challenges and Proposed Solutions

Challenge No. 1: Robust Machine Learning Model

- Robustness is defined as the property where results obtained in the training set are similar to that of the test set. A robust machine learning model is required for real-world applications.

Proposed Solution: Incremental Learning and Deep Learning can be used as a solution for achieving robustness. In Incremental Learning, the model keeps learning continuously, making it more robust. Deep learning techniques like the generation of adversarial data for checking the system's robustness are being used.

Challenge No. 2: Generalizability of Model

- Generalizability is defined by assessing the performance of a model on unseen test situations. Robustness and generalizability are usually not seen together to evaluate a model, whereas a robust, generalizable model should be the target to make a sustainable model.

Proposed Solution: Incremental and Transfer Learning can be used for mitigate above challenge. Incremental Learning is often used in image classification [116], target recognition [117], and used less in Intrusion Detection or information security.

5. Security-based Challenges and Proposed Solutions

Challenge No. 3: Real Time Analysis

- Real-time analysis is essential for any model to be adopted at the enterprise level. In malware classification, the challenge is in identifying patterns to distinguish between legitimate and malware traffic. In offline mode, machine learning models work on static datasets while the online stream of data is analyzed in online Learning.

Proposed Solution: Incremental Learning can be the solution for real-time analysis as the model can get updated according to newly added features. The related approach is used by Qureshi et al. [150] in case of support vectors, by retaining old samples which are likely to become support vector.

Challenge No. 4: Resource Constraints of IoT Devices.

- IoT devices are known to be constraint devices in terms of power, cost, and size. As for low-cost IoT devices, keeping all the security requirements is a major concern. Capacities of Deep Learning techniques could not be utilized because of these constraints.

Proposed Solution: A solution for this could be to use Deep Learning techniques with powerful hyperparameter tuning techniques. Although, Deep learning does not necessarily require feature engineering, using it makes the model lightweight. .

5. Security-based Challenges and Proposed Solutions

Challenge No. 5: Longer Training Time of IDS

- Most of the Intrusion Detection models suffer from longer training time, which affects the performance of the model to such an extent that sometimes compromise has to be made on overall system performance to reduce training time.

Proposed Solution: Transfer Learning is defined as the ability to use a pre-trained model for different yet similar work.

.

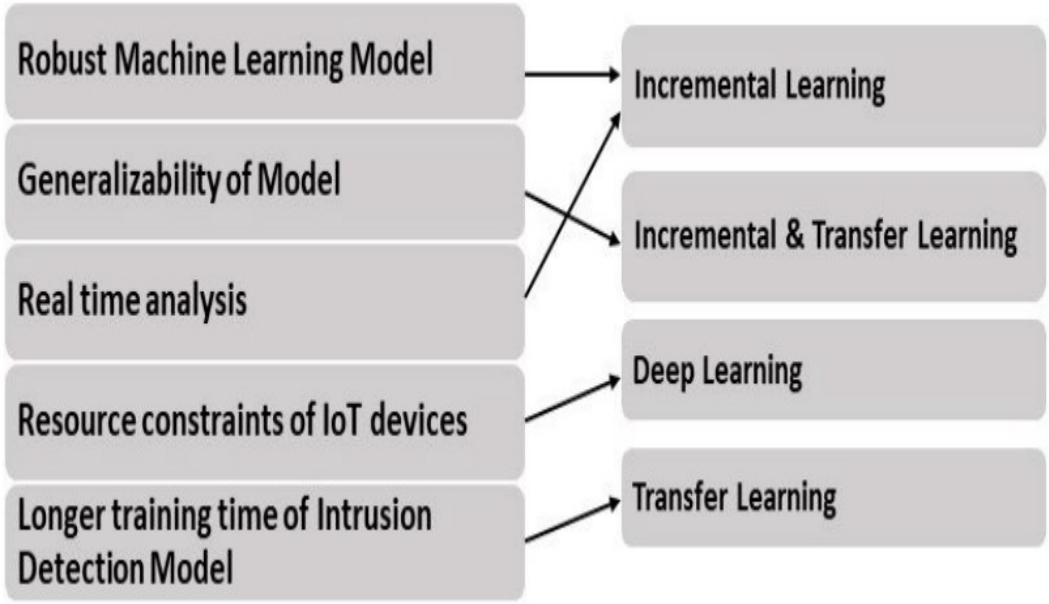

Fig. 12 : A logical mapping of comprehensive challenges, Research gaps and possible solutions.

6. Conclusion

- In this survey, two major solutions are found in the literature for preventing DDoS attacks, namely Intrusion Detection System (IDS) and Intrusion Prevention System (IPS).
- The Intrusion Detection System is analyzed in the undertaken review work, and various intrusion detection models have been evaluated.
- Furthermore, the authors have also discussed the classification of Intrusion Detection Systems, different anomaly detection techniques, various Intrusion Detection System models based on datasets, diverse machine learning, and deep learning techniques for data pre-processing and malware detection.

References

1. A. Pal, H. K. Rath, S. Shailendra, and A. Bhattacharyya, ‘‘IoT standardization: The road ahead,’’ in Proc. IntechOpen, 2018, pp. 53–74.
2. A. Srivastava, S. Gupta, M. Quamara, P. Chaudhary, and V. J. Aski, ‘‘Future IoT-enabled threats and vulnerabilities: State of the art, challenges, and future prospects,’’ Int. J. Commun. Syst., vol. 33, no. 12, pp. 1–40, 2020.
3. J. Ding, M. Nemati, C. Ranaweera, and J. Choi, ‘‘IoT connectivity technologies and applications: A survey,’’ IEEE Access, vol. 8, pp. 67646–67673, 2020.
4. Y. Yang, L. Wu, G. Yin, L. Li, and H. Zhao, ‘‘A survey on security an privacy issues in Internet-of-Things,’’ IEEE Internet Things J., vol. 4, no. 5, pp. 1250–1258, Oct. 2017.
5. W. Yu, F. Liang, X. He, W. G. Hatcher, C. Lu, J. Lin, and X. Yang ‘‘A survey on the edge computing for the Internet of Things,’’ IEEE Access, vol. 6, pp. 6900–6919, 2018.
6. D. E. Kouicem, A. Bouabdallah, and H. Lakhlef, ‘‘Internet of Things security: A top-down survey,’’ Comput. Netw., vol. 141, pp. 199–221, Aug. 2018.
7. M. Frustaci, P. Pace, G. Aloi, and G. Fortino, ‘‘Evaluating critical security issues of the IoT world: Present and future challenges,’’ IEEE Internet Things J., vol. 5, no. 4, pp. 2483–2495, Aug. 2018.
8. M. B. M. Noor and W. H. Hassan, ‘‘Current research on Internet of Things (IoT) security: A survey,’’ Comput. Netw., vol. 148, pp. 283–294, Jan. 2019.
9. C. Hesselman, M. Kaeo, L. Chapin, K. Claffy, M. Seiden, D. McPherson, D. Piscitello, A. McConachie, T. April, J. Latour, and R. Rasmussen,‘‘The DNS in IoT: Opportunities, risks, and challenges,’’ IEEE Internet Comput., vol. 24, no. 4, pp. 23–32, Jul. 2020.
10. T. Kawamura, M. Fukushi, Y. Hirano, Y. Fujita, and Y. Hamamoto, ‘‘An NTP-based detection module for DDoS attacks on IoT,’’ in Proc. IEEE Int. Conf. Consum. Electron.-Taiwan (ICCE-TW), Jun. 2017, pp. 15–16
11. J. J. C. Gondim, R. de Oliveira Albuquerque, and A. L. S. Orozco, ‘‘Mirror saturation in amplified reflection distributed denial of service: A case of study using SNMP, SSDP, NTP and DNS protocols,’’ Future Gener. Comput. Syst., vol. 108, pp. 68–81, Jul. 2020.
12. A. T. Vasques and J. J. C. Gondim, ‘‘Amplified reflection DDoS attacks over IoT mirrors: A saturation analysis,’’ in Proc. Workshop Commun. Netw. Power Syst. (WCNPS), Oct. 2019, pp. 1–6.

Thank you for your attention

sushil.sngh001007@seoultech.ac.kr
+82 10-6343-8556