Machine Learning Techniques for Securing IoT Environment

Amit Sagu, Nasib Singh Gill

Abstract* – *IoT (Internet of Thing) is becoming ubiquitous day by day and making dumb devices smarter by enabling them to transfer the information over the network. IoT not confined to homes or in utilities but can be found in array of fields. IoT is rapidly making the world smarter by connecting physical to digital world and it is estimated that by 2024 more than 20 billion devices are likely to be connected. It brings opportunity but also brings numerous kind of risks. The worry is how we to keep billions of devices secure and what to ensure the security of networks these run on. The present paper focused on all the issues concerning about securing IoT environment and how machine learning techniques may help to address these security issues. The paper also discusses the proposed approaches, parameters, characteristic of techniques and explores which technique could be more effective.

Keywords - IoT, security challenges, machine learning.

I. INTRODUCTION

IoT is network of connected devices each with a unique identifier that automatically collect and exchange data over a network. Connected devices use built in sensor to collect data and in some case act on it. The goal behind IoT is to have devices that self-report in real time, improving efficiency and bringing important information to the surface more quickly. The IoT promises to transform a wide range of fields. In medicine, for example connected device can help medical professional to monitor patients inside and outside of a hospital. According [1] the term "Internet of Things" is attributed to Kevin Asthon, who in 1999 article used to phrase to describe the role of RFID tags in making supply chain more efficient.

IoT Architecture

IoT architecture essentially comprises a number of elements i.e. cloud service, sensor and devices, network layers and end users. It's important to start including sensor in early stage of IoT architecture to get information that we need to process. A thing in IoT is an object equipped with sensor that gather data which will be transferred over network and actuator that allow things to act, for instance actuator help close circuit television camera(CCTV) to rotate in any direction to its pivot.

In next stage internet gateway work with Wi-Fi or cellular network and perform further processing. This stage is important to process the information collected from previous stage and compress it to the optimal size, this stage helps to make data digitized.

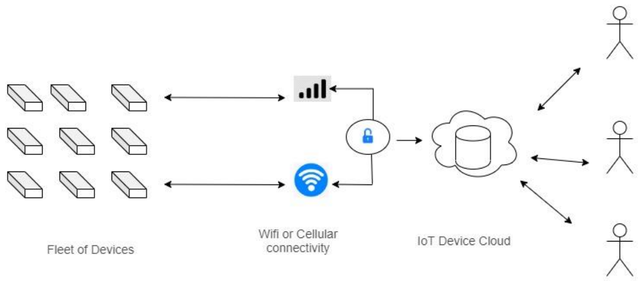

Fig. 1. IoT Architecture

In third stage data is processed in depth in data center. This stage require high end application along with skilled professional. Data might be gathered from other sources for execution. The last stage is end user with display unit to manage or monitor sensory data.

II. KEY CHALLENGES HINDERING IoT ADOPTION

IoT has made the rapid transition from abstract idea to reality. Despite trend of IoT there are still few key IoT challenges to be addressed. There are some common challenges faced by enterprises that cause quality issue and leading to project failure Compatibility & Interoperability – Sensor and networking are the integral component of IoT. But not every machine is equipped with advance sensor and networking capabilities to effectively communicate and share data. Also sensor of different power consumption capabilities and security standard inbuilt in legacy machine may not be capable to provide the same results.

Connectivity – As internet is still not everywhere at good performance. The quality of signal collected by sensor and to transmit over the network largely depend upon the local area network (LAN), metropolitan area network (MAN) or wide area network (WAN). The network has to be well connected through different technology to facilitate quick and quality communicate. Also the number of connected device is growing at a much higher rate than the network coverage, which create monitoring problem. Integration of things – Integration of IoT product with IoT platform is another challenge. For the successful implementation of IoT application we need to integrate various IoT connected products with right IoT platform. Lack of proper integration could lead to abnormalities in function and efficiency to deliver value to the customer.

Revised Manuscript Received on February 28, 2020.
***** Correspondence Author

Amit Sagu*, Research Scholar, Department of Computer Science & Applications, Maharshi Dayanand University, Rohtak, India. Email: saguamit98@gmail.com

Nasib Singh Gill, Professor, Department of Computer Science & Applications, Maharshi Dayanad University, Rohtak, India. Email: nasibsgill@gmail.com

Security – This is most significant barrier which limiting adoption of IoT. Increasing the number of connected devices increasing the opportunity to exploit security vulnerability.

Poorly designed devices which can expose user data to theft by leaving data stream inadequate protected and in some cases people's health and safety can be put at risk.

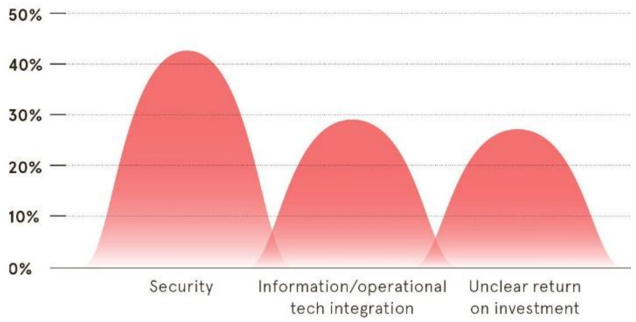

Fig. 2. Barrier in adoption of IoT (*Source: https://.raconteur.net*)

[2] Study shows that data security concern top the list of challenges that are slowing down IoT adoption for 70 percentage of the companies. In this paper we are going to concern about security aspect as fig. 2. Shows that more than 40 percentage of companies experienced security as the barrier [3].

Security Threats

IoT security threat are not merely theoretical. Attacker already have found the way to compromise many devices and networks. IoT may be introducing a lot of benefits to our modern life but it also has one major drawback i.e. security threat. IoT security risk could even more significant on user side where they often not aware of potential threats. IoT security threat can be used to steal critical data from people as well as organization. Attacker can exploit security vulnerability in IoT infrastructure to execute sophisticated cyber-attack. IoT security threat must be more concerning for user as they are unaware of their existence. Following are some most used security threats.

Denial of Service – A Denial of Service (DoS) attack deliberately tries to cause a capacity overload in the target study by sending multiple request. Attacker who implement DoS attacker don't have intention to steal data however it can be used to slow down or disable service. Even for seconds of halt of security service can be huge threat.

Home Invasion – Perhaps one of the scariest things nowadays. IoT devices are used in large number at home and offices which has given rise to home automation. The security of these IoT devices is a huge matter of concern as it can expose device internet protocol (IP) address point to residential address. Even if we are using IoT device in home security system then there is a possibility that they might compromised.

Untrustworthy Communication – There are many IoT devices which send messages to the network without any encryption. This is one of the biggest IoT security challenge which exist out there. Many encryption technology is being used, one the emerging is lightweight encryption. It is very suitable for IoT device as these devices are resource constrained in respect of battery power, complexity handling power. But it is very complex and not easy to implement lightweight encryption also it causes low performance of IoT network.

Man in the middle attack – In this attack, attacker breach into the communication channel between two systems in attempt to intercept the message from them. Attacker gain control over their communication and send illegitimate message to participate system. Such attack used to hack IoT devices in home or organization.

Physical Attack – Most of the time sensor are deployed at remote location connected with network and administration has no control on their physical aspect. Physical attack target the hardware of IoT system and include breaches at sensor layer. Attacker can alter node to gain control over node/device in an IoT environment and use that to extract the information. Also attacker can physically damage IoT devices to disrupt the availability of service.

Botnet – A botnet is a network that combine various system together to remotely take control over a victim system and distributed malware. Attacker then control botnet using command and control server to steal confidential data. One of the most known botnet is Mirai botnet which has showed that how dangerous IoT security threat can be. Mirai botnet has infected an estimated 2.5 million devices including router, smart camera and printers.

IoT Security Solution

We have discussed many security threat for IoT which can limiting the adoption of infrastructure. As far there are many solution which can effective as security aspect. We can achieve secure IoT infrastructure by securing the network virtually, but it is difficult to ensure security physically since many device or sensor deployed in remote location. Lightweight encryption is emerging trend nowadays in which data is encrypted before send in the network. Despite encryption algorithm are lightweight but yet it is very complex to implement as IoT devices are resource constrained in their battery power consumption, complexity handling power. There is another alternate which is emerging nowadays i.e. Machine Learning (ML). Since machine learning is knowing for automation and prediction it can be great tool to counter the security challenges or attacks. Machine learning algorithms can be used to predict or identify the attacks so that it can be prevented. Further we will see number of machine learning algorithms and techniques which could help to achieve secured IoT infrastructure.

III. MACHINE LEARNING IN IOT

Machine learning is the science of getting computer to learn and act like human do and improve their learning over time in autonomous fashion. Machine learning algorithm use statistics to find pattern in huge amount of data and learn from it, then make a determination or prediction about something.

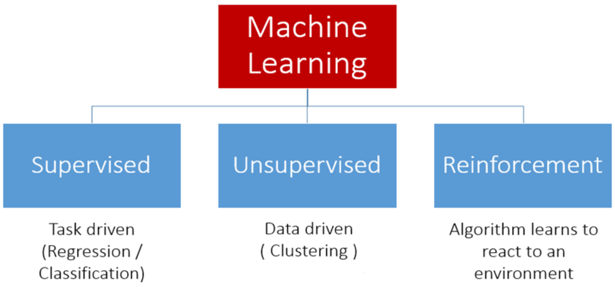
Machine Learning
Supervised
Unsupervised
Reinforcement
Task driven (Regression / Classification)
Data driven (Clustering)
Algorithm learns to react to an environment

Fig. 3. Machine Learning Fields (*source:* *https://mc.ai/understanding-machine-learningas-6-jars/*)

Machine learning can be classified into three type as in Fig. 3. In supervised learning the database on which we train our model is labelled. There is clear and distinct mapping of input and output (1), where x variable is input and y variable is output. Based on the input example the model is able to get trained.

$$f(x) \to y \qquad (1)$$

In unsupervised learning there are no labelled data and algorithm identifies the pattern within database and learn from them. The unsupervised algorithm groups the data into various cluster based on their density or characteristics.

Reinforcement learning is emerging and most popular type of machine learning algorithm, the aim of this algorithm is to reach the goal in dynamic environment. It reaches the goal based on reward and penalty provided to it by system.

Machine learning has seen a significant rises in popularity across a very broad range of application in recent year. Using of machine learning in security of IoT is emerging trend and it is becoming new alternate which is giving us good results. We will see here what algorithm and techniques can be used or being used to secure IoT infrastructure especially for home IoT environment.

Machine Learning Algorithms in IoT Security

There are many algorithms which can be effectively used to secure IoT infrastructure. Supervised learning works better when we know the environment variable i.e. output to correspond to every input. We use unsupervised learning where we do not about the output, mainly this learning is used to categorize the characteristics. On other hand reinforcement learning is different from two learning, in this learning software agent learns from their own positive or negative experience.

Table- I: Machine Learning Categories

Supervised	Unsupervised	Reinforcement
Input vs. Output pair	Input only	Input & feedback
Learning phase vs. acting phase	Learning phase vs. acting phase	Learning and acting simultaneously
Learn by training	Learn by experience	Learn by explore environment and trial & error

In Fig. 4, Support vector machine (SVM) is one of widely supervised learning algorithm. It can be used for both classification and regression challenge. However it is mostly used in classification problem. In this algorithm we plot each data item as point in n-dimension space with value of each feature being the value of a particular coordinator, then we perform classification by finding the hyper plane that differentiate the two classes.

Random forest is another supervised learning algorithm. There is direct relationship between numbers of trees in forest, the number of decision tree, the more accurate result will be. Decision tree is a basic element of this algorithm which is also a decision support tool. A tree like graph is used to show possible consequence. The main advantage is that this algorithm can be used for both classification and regression task.

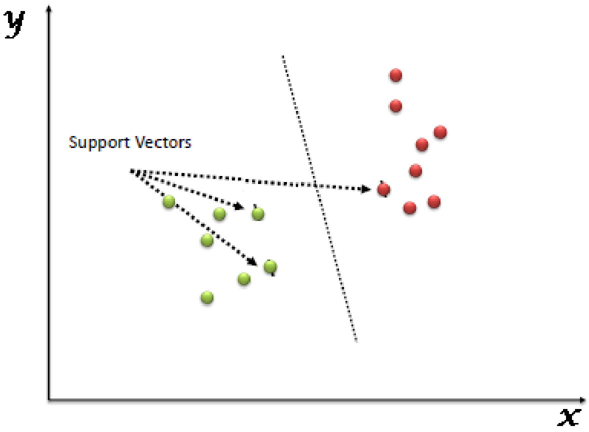

Fig. 4. Support Vector Machine (*Hyper plane differentiating two classes*)

Naive Bayes classifier is a straightforward and powerful algorithm for the classification task. It gives great results when we use it for textual data analysis, such as Natural Language Processing. Naive Bayes Classifiers based on the Bayes' Theorem, which is based on conditional probability means an event A will happen, given that another event B has already happened. The theorem (2) allows a hypothesis A to be updated each time new evidence B is introduced.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \qquad (2)$$

Where P is denoted as probability, P (A | B) is the probability of event A occurring given that B has occurred already. P (B | A) is the probability of the event B occurring given that A has occurred. P (A) the probability of event B occurring and P (B) probability of event A occurring.

Table- II: Machine Learning Algorithm

Sr.	Machine Learning algorithm	Data processing task
1	Support Vector Machine	Regression/Classification
2	Random Forest	Regression/Classification
3	Linear Regression	Regression
4	Naïve Bayes	Classification
5	K-means	Clustering
6	K- nearest neighbor	Classification
7	Classification and regression tree	Regression/classification

IV. RELATED RESEARCH WORK

Divergent work has been done or being done by researcher community to secure IoT infrastructure.

Hyo-Sik Ham & et al. (2014) proposed an Android malware-detection mechanism using machine learning algorithms for reliable IoT services. The Linear Support Vector Machine (SVM) method is used. Malware is detected based on the collected data by monitoring resources in an Android environment. Paper is suggested not to use behavior based detection since behavior-based detection increases the usage of a smartphone's battery and memory [4].

Mehdi Nobakht & et al. (2016) propose an intrusion detection and mitigation framework, called IoT-IDM (Intrusion Detection & Mitigation), to provide a network-level protection for smart devices deployed in home environment. The main contribution in framework are advent of SDN technology, OpenFlow protocol, machine learning technique for detection attack pattern and Java module Floodlight for implementation. They used host based intrusion and detection instead of network based system. In order to demonstrate the applicability of framework they selected Philips Hue bulb which accept command from user via HTTP protocol [5].

Janice Canedo & Anthony Skjellum (2017) presented a concept to secure IoT edge device within gateway using ANN machine learning technique. R programming tool is used to create ANN. While making test bed Ardunio Uno device are used to emulate edge devices which connect to WiFi chip and temperature sensor and Raspberry pi model 3 to implement gateway which is credit card size micro controller and low power consumption, technical specification detailed . The neural network used five layer with three hidden layer and input to network model are sensor value, device ID and time stamp as features [6].

Suman Sankar Bhunia & et al. (2017) proposed SDN based secure IoT framework called SoftThings used with machine learning algorithm. The core idea is based on SDN controller which dynamically control the traffic flow also separate the control plane and data plane. Support Vector Machine (SVM) and SVM: non-linear machine learning algorithm is used which classifies the attack in normal and abnormal traffic. The main component of framework are IoT devices, SDN switch, SDN controller and Master SDN controller. SDN controller consist three main module i.e learning module which analyze the flow pattern, classification module for classify the network traffic and flow management module to control the network flow [7].

Deyban Perez & et al. (2017) presented seven multiple hybrid model to detect intrusion in network, since single algorithm strategy model shows high rate of false alarms. The strategy is based on two approaches first is use of supervised learning, which in the context of detection of 10 known attack, second is to use unsupervised learning for the detection unknown and new attacks. The hybrid model increase the high hit rate compare to single algorithm model. The algorithm used are Neural Network (NN) and Support Vector Machine (SVM) for supervised learning, K-mean for unsupervised learning. As feature selection technique Principal Component Analysis (PCA) and Gradually Feature Reduction (GFR) are used [8].

Miettinen & at el. (2017) presented IoT sentinel which is capable to identify new introduced device in network. Authors does so by controlling the network traffic flow of vulnerable devices. IoT sentinel restrict the communication so that adversary is not able to connect vulnerable device to exploit. They use major component i.e Security Gateway and IoT security service provider (IoTSS). Security gateway is SDN [10] based to monitor the profile of individual device and send fingerprint to IoTSS. IoTSS is used here to assess the vulnerability of device. IoTSS uses machine learning classification to check whether fingerprint match or not of individual device. For mitigation strategy they implement concept of network isolation, traffic filtering and user notification. For device identification device fingerprint is used, which is observation of passive network history log i.e source address, time stamps, propagation time etc. It help to extract the features which is further used in machine learning classifier [9]. Prachi Shukla (2017) present three new Intrusion Detection Systems (IDSs) for IoT i.e K-means clustering unsupervised learning based IDS, decision tree based supervised IDS, and a hybrid two stage IDS that combines K-means and decision tree learning approaches. All the three IDS are centralized and scalable approaches. The K-means approach achieves 70-93% detection rate for varying sizes of random IoT networks. Decision tree based IDS achieves 71-80% detection rate and the hybrid approach attains 71-75% detection rate for the same network size [10]. Ravi Kumar & at el. (2017) proposed a framework to detect android malware application based on permission asking by them using machine learning techniques. Since android allow many other open source such as torrent, Google play store and direct download make it more prone to attack. Authors classified malware after extracting the permission database stored in file Androidmenifest.XML. Different machine learning techniques viz. Naïve Bayes, J48, Random forest, Multiclass classifier and multilayer perceptron applied on sample data and performance evaluation is done by confusion matrix. It shows Multi-layer perceptron computation complexity is very poor among all machine learning techniques [11]. Perez & at el. (2017) proposed Multiple hybrid machine learning model for intrusion detection in computer network. They used supervised learning for known attack, and for unknown attack unsupervised is used. Here two terms are used i.e. Multiple and Hybrid, hybrid refers more than one category of machine learning is used and multiple more than one algorithm is used of either supervised or unsupervised learning. Supervised learning SVM and NN are used, in unsupervised learning K-mean used. In their implementation, model uses NN and SVM in first level where known attack are identified and K-mean in the second level where unknown attack are identified. NSL-KDD dataset are used as testing and training dataset. DoS, Normal, Probing, R2L and U2R are the supervised output label, Normal & Attack are two class for unsupervised. PCA and GFR are used as feature selection technique [8].

V. RESULT AND DISCUSSION

Table- III: Comparative Analysis

Pub. Year	Title of Paper	Tech. used	Research gaps
2014	Linear SVM-Based Android Malware Detection for Reliable IoT Services [4].	SVM, ANN	Efficient and lightweight implementation of the SVM algorithm that can be embedded to a smartphone for real-time detection.
2016	A Host-based Intrusion Detection and Mitigation Framework for Smart Home IoT using OpenFlow [5].	SDN, Openflow, FloodLight	The framework is work only for single and selected host in the home IoT environment that do not add a lot of overhead on SDN controller.
2017	Machine-learning Classifiers for Security in Connected Medical Devices [12].	Decision tree classfier, K-mean, SVM	The approach fail if the attacker is insider and more familiar with pattern of data access from medical device.
2017	Using Machine Learning to Secure IoT Systems [6].	ANN, Multilayer perceptron	The data set was so small to test. Large data set could be effective for intrusion detection & Need to further research with a larger scale system and increased data collection.
2017	Dynamic Attack Detection and Mitigation in IoT using SDN [7].	SDN, SVM:Linear, SVM: non-linear	No hardware development is given.
2017	IOT SENTINEL: Automated Device-Type Identification for Security Enforcement in IoT [9].	SDN, Machine learning classifier	Results were not able to comprehensively investigate the impact of updates since only a few device offered update during experiment period.
2019	IoT Security Viewer System Using Machine Learning [13].	Random forest, ARP	The proposed system is applicable only in same network segment.

Since lot of work has been done in context of IoT security and this paper attempts to explore several research gaps which can either be filled or provides a future scope of work in a more efficient manner. After formulating the problem statement, the paper comes to conclude many solutions to secure IoT environment. The proposed solution intrusion detection system (IDS) would work between home IoT environment and external network. Intrusion detection system will not let pass any traffic without scanning their characteristics like destination address, source address, packet size, packet interval time etc. It uses machine learning algorithm in intrusion detection system to find any type of abnormality in the incoming traffic. In proposed techniques, model can be trained against any type of attack based on their symptoms. For instance to shield IoT environment from DoS attack which is very common attack for IoT, then model has to understand the traffic symptom of attack which is, in DoS attack packet sending interval time is very less, almost all the packets have same size in their payload. So based on these symptom machine learning algorithm can detect abnormal traffic and can take action like halting the traffic or alert the user about threat. In Fig. 5, intrusion detection system operates as middle box between IoT environment and internet. In proposed technique intrusion detection system is machine learning trained model against different number of attacks.

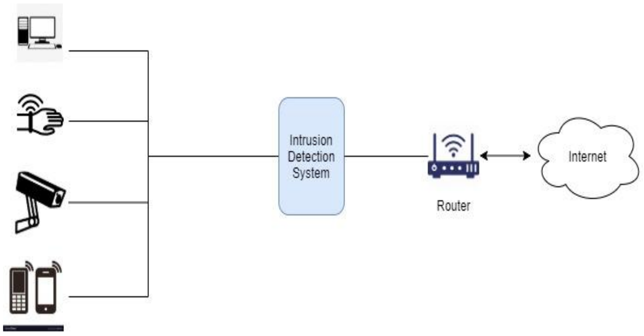

Fig. 5. Proposed Approach

VI. CONCLUSION

In present paper, it can be seen that increasing number of attacks and security challenge might be barrier to adopting IoT universally. Paper discern about solution and conclude that machine learning is emerging as alternate approach. This paper compared analysis of many recent work and came to infer that there are some research gap which can be a future reference. Proposed approach uses different type machine learning algorithms to detect the attack in IoT environment. The model supposed to use between home network and external network which filters all the incoming traffic from outside source.

REFRENCES

1. K. Ashton, "That 'Internet of Things' Thing," RFID Jouranl, 2009.
2. M. Milojevic, "Digital Industrial Transformation with the Internet of Things," Accenture Digital, 2017.
3. S. CHARARA, "IoT FOR BUSINESS," 2018.
4. H.-H. K. M.-S. K. M.-J. C. Hyo-Sik Ham, "Linear SVM-Based Android Malware Detection for Reliable IoT Services," Hindawi Publishing Corporation, 2014.
5. V. S. R. B. Mehdi Nobakht, "A Host-based Intrusion Detection and Mitigation Framework for Smart Home IoT using OpenFlow," in 11th International Conference on Availability, Reliability and Security (ARES), 2016.
6. A. S. Janice Ca˜nedo, "Using Machine Learning to Secure IoT Systems," in 14th Annual Conference on Privacy, Security and Trust (PST), 2016.
7. M. G. Suman Sankar Bhunia, "Dynamic Attack Detection and Mitigation in IoT using SDN," in 27th International Telecommunication Networks and Applications Conference (ITNAC), 2017.
8. P. Deyben, "Intrusion detection in computer networks using hybrid machine learning techniques," in XLIII Latin American Computer Conference (CLEI), IEEE, 2017.
9. T. F. A.-R. S. S. M. N. A. I. H. S. T. Markus Miettinen, "IOT SENTINEL Demo: Automated Device-Type Identification for Security Enforcement in IoT," in IEEE 37th International Conference on Distributed Computing Systems, 2017.
10. P. Shukla, "ML-IDS: A Machine Learning Approach to Detect Wormhole Attacks in Internet of Things," in Intelligent Systems Conference, 2017.

11. K. P. S. R. Ravi Kiran, "Android mobile security by detecting and classification of malware based on permissions using machine learning algorithms," in International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud), IEEE, 2017.
12. G. T. Sida Gao, "Machine-learning Classifiers for Security in Connected Medical Devices," in 26th International Conference on Computer Communication and Networks (ICCCN), 2017.
13. H. S. A. K. Yuya Kunugi, "IoT Security Viewer System Using Machine Learning," Springer Nature Switzerland, p. 1071–1081, 2019.

AUTHORS PROFILE

Mr. Amit sagu, has passed M.sc in 2016 in Computer Science and Applications from Department of Computer Science & Applications, Kurukshetra University Kurukhsetra, India. He has also worked as Assistant Professor at DAV Centenary College, Faridabad, India. He is currently pursuing Ph.D. in Computer Science at M. D. University, Rohtak. His research interests include IoT, Machine Learning, Big Data Analytics and Data Mining.

Dr. Nasib Singh Gill, is at present senior most Professor of Department of Computer Science & Applications, M. D. University, Rohtak, India and is working in the Department since 1990. He earned his Doctorate in Computer Science in the year 1996 and carried out his Post-Doctoral research at Brunel University, West London during 2001-2002. He is a recipient of Commonwealth Fellowship Award of British Government for the Year 2001. Besides, he also has earned his MBA degree. He has published more than 245 research papers in reputed National & International Journals, Conference Proceedings, Bulletins, Edited Books, and Newspapers. He has authored seven books. He is a Senior Member of IACSIT as well as a fellow of several professional bodies including IETE and CSI. He has been serving as Editorial Board Member, Guest Editor, Reviewer of International/National Journals and a Member of Technical Committee of several International/National Conferences. He has guided so far 9 Ph.D. scholars as well as guiding about 7 more scholars presently in the areas – IoT, Machine Learning, Information and Network Security, Computer Networks, Measurement of Component-based Systems, Complexity of Software Systems, Decision Trees, Component-based Testing, Data mining & Data warehousing, and NLP.