MACHINE LEARNING BASED CYBER SECURITY OF THREATS DETECTION USING INTRUSION SYSTEM

[1]Kancheti Sri Harshitha, B.Tech scholar, Hindu College of Engineering and Technology, Guntur, A.P, India.
[2]Kaza Mohan Krishna, B.Tech scholar, Narasaraopeta Institute of technology, Narasaraopet, A.P, India.

ABSTARCT: Cyber-security has recently resulted in significant changes in operations and technology, and data science is also developing in the context of computing. It is crucial to draw patterns or insights about security incidents from cyber-security data and create models based on the essential data in order to automate and make security systems useful. Security issues increase along with increased internet usage. Due to system security problems, malware degrades system performance and affects data privacy. Attacks can be detected and reported using intrusion detection systems (IDS). It is determined by an Intrusion Detection System (IDS) whether network traffic behaviour is typical, unusual, or suggestive of a specific type of attack. Machine Learning is being used more and more in cyber security. Making the process of detecting malware more realistic, scalable, and efficient than current methodologies is the main objective of applying Machine Learning to cyber security. As a result, Machine Learning-based intrusion system security against threats is presented in this study. This system use the Support Vector Machine (SVM) classifier to detect threats in a highly accurate and efficient manner. Accuracy, sensitivity, and specificity will be used as metrics to assess the performance of the system being presented.

KEYWORDS: Cyber-security, Intrusion detection System, Machine Learning

I. INTRODUCTION

In the modern world of computers, the majority of gadgets are interconnected through the Internet of Things (IoT). These gadgets communicate and exchange data using the Internet, an unsecured (open) communication method. Generally speaking, this information is private (for example, health information, banking information, insurance information, other financial information, and social security numbers).

The threat actors, such as online attackers (hackers), are constantly looking for opportunities to manipulate things (for instance, by conducting attacks like replay, man-in-the-middle, impersonation, credential guessing, computing session keys, injecting malware, and data alteration) [1]. Cyber security refers to methods and strategies that prevent unauthorized people from damaging, attacking, or accessing programmes, networks, computers, and data [2].

Cyber-security is divided into many different categories and includes a wide range of applications, including business and mobile computing. (1) Network security, application security, information security, and operational security are all required in order to protect devices and software from cyber risks or threats. Information security (ii) focuses primarily on the security and confidentiality of the data in question; operational security (iii) refers to the procedures for handling and protecting data resources. Attackers and vulnerabilities in cyber security are widespread and unavoidable [3].

Nowadays, political and commercial entities are increasingly using sophisticated cyber-warfare to harm, interfere with, or restrict computer networks informative content. Network protocol architecture must be resistant to attack by a strong attacker in control of some network nodes [4]. There are several security problems due to the rising use of the Internet. Several technologies, including firewalls, data encryption, and user authentication, are utilized to close this security gap. These security measures guard against various attacks. However, deep packet analysis is not possible with current security methods. As a result, they will never achieve the desired level of attack detection [5].

A firewall, antivirus programme, or intrusion detection system are examples of conventional cyber-security solutions used in network and computer security systems. Systems like Intrusion Prevention System (IPS) and Intrusion Detection System (IDS) were created to fill in the gaps left by these security protocols. These systems Machine Learning, Deep Learning, and Artificial Intelligence-based algorithms enable them to analyze data more thoroughly than other security systems. IDS systems are only used for intrusion detection and analysis, while IPS systems act as both intrusion detection and prevention measures. A group of cyber-security-based technologies known as Intrusion Detect Systems (IDSs) were first created to identify defects and exploits in a target host. An IDS's only function is to find threats.

Consequently, there is a need for more adaptable, strong, flexible, and capable cyber defence systems that can identify numerous threats in real time. Artificial intelligence (AI) technology use has grown recently and is becoming increasingly important for identifying and preventing cyber threats [6]. Data Science (DS), a key component of "Artificial Intelligence" (AI), is driving the change in cyber-security, which has recently undergone enormous technological and operational developments in the context of computers. Plays an important role in drawing insights from data.
Machine Learning has the potential to significantly change the cyber security environment, and data science is bringing in a new scientific approach [7]. As a result, it concentrates more on providing a thorough explanation of various Machine Learning technology types, as well as their interactions and applications in the context of cyber security. To efficiently uncover insights or patterns of security incidents, a variety of Machine Learning approaches, including feature projection, data clustering, classification and association analysis, or deep learning methods based on neural networks, can be applied.

This approach offers machine learning-based cyber security that utilizes intrusion detection systems to find threats. The rest of the text is divided into different sections: Section II describes the relevant work. Section III presents the given ML-based intrusion system-based threat detection system. Section IV provides an analysis of the suggested technique. This process is finally completed in Section V.

II. LITERATURE SURVEY

Zohre Nasiri Zarandi,Iman Sharifi et. al. [8] Detect and identify cyber dangers in cyber-physical systems using Machine Learning techniques. The suggested method uses a deep Neural Network structure in the detection phase, which should inform the system to the attack's presence at its very beginning. According to experimental study, Deep Learning algorithms can detect threats more accurately than conventional methods, and they can also make cyber security simpler, more proactive, less expensive, and more effective.

Una-May O’Reilly, Jamal Toutouh, Marcos Pertierra, Daniel Prado Sanchez, Dennis Garcia, Anthony Erb Luogo, Jonathan Kelly, Erik Hemberg et. al. [9] Adversarial Genetic Programming (GP) for Cyber Security is presented. GP is a growingly significant application area. Through Genetic Programming (GP), adversarial genetic programming for cyber security has been expanded to replicate and research the dynamics of cyber attackers' behaviour and interactions. The study of network security fighting is supported by a framework called RIVALS. By computationally modelling and recreating the dynamics of cyber networks under attack, it seeks to describe such dynamics.

Arunan Sivanathan, Hassan Habibi Gharakheili, and Vijay Sivaraman et. al. [10] Utilizing Machine Learning and programmable telemetry, IoT provides controlled cyber-security. This solution uses programmable flow-based telemetry, flexible data-driven models, and machine learning to control IoT devices depending on network activity. This eliminates the need to choose between the performance and expense of the solution, and it demonstrates how a monitoring system may be used to detect behavioural changes (firmware upgrade or cyber attacks).

Pramita Sree Muhuri, Prosenjit Chatterjee, Xiaohong Yuan, Kaushik Roy and Albert Esterline et. al. [11] Use recurrent long-term memory neural networks to classify network threats (LSTM-RNN). In this article, created a new intrusion detection strategy by combining Recurrent Neural Network (RNN), Genetic Algorithm (GA), and LSTM to pick the best feature to categorize the Network Security Laboratory-Knowledge Discovery (NSL-KDD) dataset into a database. The NSL-KDD dataset was used to assess the classifier's performance. The NSL-KDD dataset was divided into binary (normal and aberrant) and multiclass datasets using LSTM-RNN.

Ployphan Sornsuwit and Saichon Jaiyen et. al. [12] The detection of cyber security threats using a new hybrid Machine Learning approach based on adaptive validation is given. To integrate multiple categories, adaptive weights are used to merge irrelevant features, and correlation-based feature selection is utilized to eliminate irrelevant characteristics. The UNB-CICT dataset or network traffic was used in the experiment to gauge the effectiveness of the suggested approach. The findings demonstrate that this strategy is highly effective in identifying all forms of attacks.

Anastasia Gurina and Vladimir Eliseev et. al. [13] An strategy based on anomalies is described to identify various types of network attacks. The detection of abnormal threats is made simple using a new method. It depends on being able to interpret the web server's dynamic response as it processes the request. To detect deviations in dynamic response, an auto-encoder is used. The case study of the MyBB web server is presented. Using this technique, numerous SQL injection and flooding attacks have been created and successfully stopped.

Charles Feng, Shuning Wu, Ningwei Liu et. al. [14] Supplying the Cyber Security Operations Center with a Machine Learning platform that is user-centric. For enterprise cyber-security operations centres, develop a framework for user-centric Machine Learning. Explains how typical SOC data sources operate and how to leverage and modify them to create powerful Machine Learning algorithms.

Sanjay Kumar, Ari Viinikainen, Timo Hamalainen et. al. [15] For network-based intrusion detection systems, a Machine Learning categorization model is offered. In this study, various supervised Machine Learning classifiers are developed utilising labelled network traffic feature examples from both malicious and benign applications. Due to the prevalence of mobile malware worldwide and the popularity of Android among users, this research focuses on Android-based malware. According to the evaluation's findings, the model used to have an accuracy of up to 99.4% when it came to identifying both known and unidentified threats. For the purpose of detecting sophisticated threats and minimising false positives, this ML model can also be combined with conventional intrusion detection systems.

III. THREATS DETECTION USING INTRUSION SYSTEM

This method uses an intrusion system to illustrate machine learning-based cyber security that can identify threats. Fig. 1 shows the workflow of the system that is being discussed. The crucial step that connects security flaws in cyber infrastructure to the relevant data-driven remedial stages in this structure is the gathering of important cyber-security data. The nature and quantity of cyber data determine the viability and effectiveness of addressing the targeted security issue.

The specific security issue and enterprise project determine the overall procedures for gathering and handling security data from diverse data sources. In a network architecture, a security system can utilize a variety of security data to give specialized security services, including IDS logs, firewall logs, network traffic data, system data, packet data, and honeypot data, among others. For instance, data analysis using information on IP addresses and their online activities can reveal if a specific IP is dangerous or not.

The intrusion detection system is provided with basic security data that hasn't been processed from a variety of sources. "Software, device, or application that monitors a computer system or network for malicious activity or policy deviations" is an Intrusion Detection System (IDS). When it comes to identifying and analyzing attacks and abnormalities, an IDS system works similarly to a signature- or anomaly-based system. An intrusion detection system that uses signatures matches network input to attack patterns stored in a database. In order to identify abnormalities in network traffic, anomaly-based detection first learns patterns in regular network traffic and then identifies deviations from those patterns as anomalies.

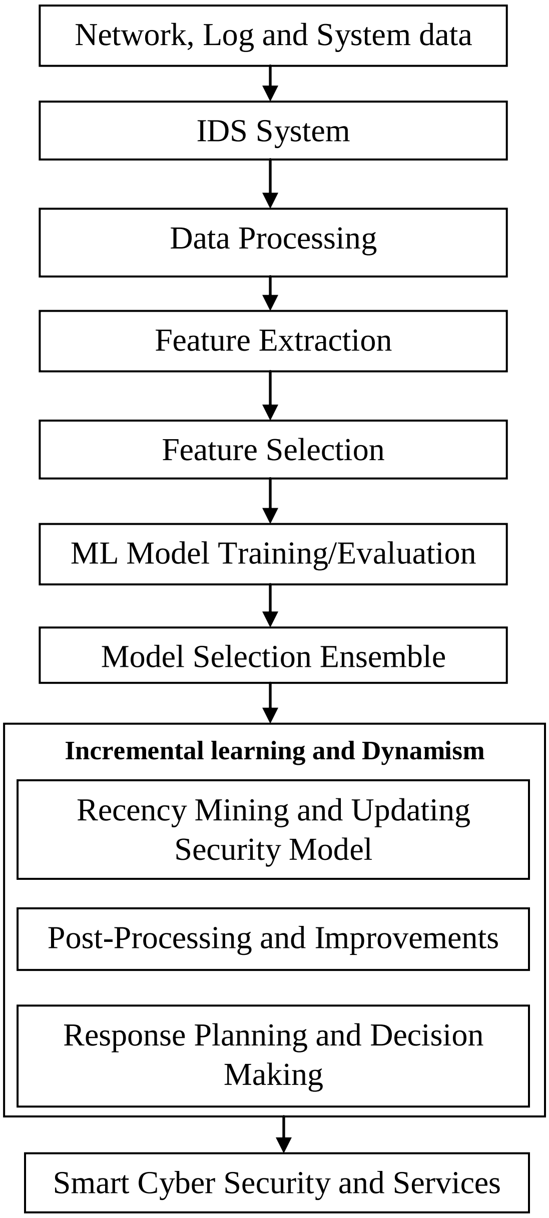
Network, Log and System data
IDS System
Data Processing
Feature Extraction
Feature Selection
ML Model Training/Evaluation
Model Selection Ensemble
Incremental learning and Dynamism
Recency Mining and Updating Security Model
Post-Processing and Improvements
Response Planning and Decision Making
Smart Cyber Security and Services

Fig. 1: MACHINE LEARNING BASED CYBER SECURITY OF THREAT DETECTION USING INSTRUSION SYSTEM

In order to identify dynamic trends, automatically build data-driven models, profile usual behaviour, and detect anomalies at every step, an anomaly-based intrusion detection system first looks at user activity and network traffic. By examining security data from numerous important sites throughout a network or system, IDS finds solutions to issues.

Not all gathered information is useful for cyber-security modelling. As a result, the remaining network sniffer data must be cleaned up of any unnecessary information. Your data might also be noisy, missing or corrupted, or contain features of various shapes and sizes. Data-driven models, which develop a function that maps inputs to outputs based on instances of input-output combinations, need high-quality data to achieve high accuracy. As a result, procedures may be needed to manage missing or faulty values and clean up the data. The secured data may also have other characteristics or traits, including being continuous, discrete, or symbolic. Must preprocess the data and properties to convert it to the target type in addition to being aware of these data kinds, as well as the permitted characteristics and operations. Normalization, transformation, and sorting are examples of data preprocessing techniques that can be helpful for structuring data.

With a starting set of metrics, feature extraction creates useful, non-redundant (feature) values to speed up the training and generalization processes and, in some situations, improve human interpretations. Dimensionality reduction is a component of feature extraction. The process of feature extraction includes using less resources to describe huge amounts of data. By using only these data and removing noise from the data, feature selection is a technique for lowering the number of input variables into this model. automatic selection of the ideal characteristics for a Machine Learning model according on the nature of the challenge. Feature selection is a technique used in machine learning to improve accuracy. By focusing on the most crucial variables and eliminating redundant and irrelevant data, it also improves the algorithm's prediction ability.

The accessibility of cyber security data supports Machine Learning in this field. The basis for developing Machine Learning algorithms for cyber security is datasets, which are collections of records holding information in the form of various characteristics or features and related facts. It is crucial to understand the nature of cyber-security data, which includes various cyber events and essential elements. ML is a crucial stage in the process of gaining information and insights from data by utilizing cyber-security data science.

Machine learning is a technique used to protect data in the cloud by identifying suspect user behaviour. It continuously learns by examining data to find patterns that can be used to identify insider threats, anticipate where "bad neighbourhoods" are online, and detect malware in encrypted communication. This method use SVM to detect threats and raise the proposed system's threat detection accuracy in the cloud by flagging risky user behaviour. Support Vector Machine (SVM), a method for supervised learning, examines the data and identifies patterns. SVM, or supervised Machine Learning, is a technique that can do classification, regression, and outlier detection. A linear SVM classifier works by drawing a straight line between two classes.

The KDD'99 Cup dataset, which has 41 features required to evaluate ML models, is utilized to train the ML model. The most used data set for evaluating anomaly detection techniques is KDD CUP99. Threats are divided into four categories: Remote-to-Local (R2L), DoS, probing, and User-to-Remote (U2R).

To complete the resulting security model, further dynamics and information may be included as necessary. Using semantics or other forms of domain knowledge, this module helps cyber-security applications better correlate attacks. By building a security model based on the most recent data, the most recent security model is responsible for updating the security model with greater performance. In terms of processing and output, dynamic updates to the most recent model are more effective than reprocessing all security data. The resulting cyber-security model may become more intelligent and dynamic as a result. Finally, the reaction planning and decision making module is in charge of providing automated and intelligent services by making decisions based on information acquired and taking the appropriate actions to prevent cyber attacks on the system. Services may differ depending on the individual needs for specific security challenges.

This framework is frequently used in data science for cyber security to extract relevant information from security data and create intelligent systems for cyber-security, intrusion detection, access control management, anomaly and fraud detection, or denial of service or denial of service attacks.

IV. RESULT ANALYSIS

This analysis uses an intrusion system to detect threats utilising Machine Learning-based cyber security. The result analysis of presented system is demonstrated here. The purpose of an intrusion detection system is to find data anomalies. The SVM classifier is trained on the KDD' 99 Cup dataset. Threats and attacks are found using the SVM classifier. Using True Positive (TP), True Negative (TN), False Positive (FP), and False Negative, which are defined as follows, the performance of ML-based cyber security is evaluated for threat detection.

TP: If a instance is detected correctly as positive and actually it is positive.

TN: If an instance is detected as negative and actually it is negative.

FP: if an instance is detected as positive but actually it is negative.

FN: if an instance is detected as negative but actually positive.

Sensitivity: It is sometimes referred to as Real Positive Rate (TPR), and it is described as the proportion of true positive cases to actual positive instances.

$$Sensitivity = \frac{TP}{TP + FN} \times 100 \ (1)$$

Accuracy: It is described as the proportion of accurate occurrences to all instances, and it is given as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \times 100 \ \ (2)$$

Specificity: The ratio of classified true negative instances to real negative instances (i.e., FP plus TN) is what it is known as, and it is stated as

$$Specificity = \frac{TN}{TN + FP} \times 100 \ \ (3)$$

This value is also known as the test's selectivity. This corresponds to a 1 False Positive Rate.

The table 1 shows the performance metrics evaluation of ML based cyber security of threat detection using Intrusion system. The performance of presented system is compared with different ML classifiers.

Table 1: PERFORMANCE METRICS EVALUATION

ML Classifiers	Accuracy (%)	Sensitivity (%)	Specificity (%)
Decision Tree (DT)	84.5	82.3	83.5
Naïve Bayes (NB)	91.34	90.67	92.2
SVM based Intrusion system	97.65	96.87	97.62

The Fig. 2 shows the sensitivity and specificity comparison between presented and other ML classifiers.

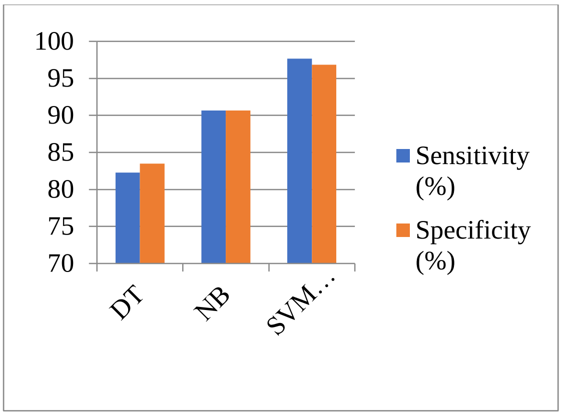

Fig. 2: COMPARATIVE GRAPH FOR SENSITIVITY AND SPECIFICITY

The different ML classifiers are represented on the x-axis, while the % values for sensitivity and specificity are shown on the y-axis. Compared to DT and NB presented SVM classifiers has high specificity and sensitivity. The Fig. 3 shows the accuracy comparative graph.

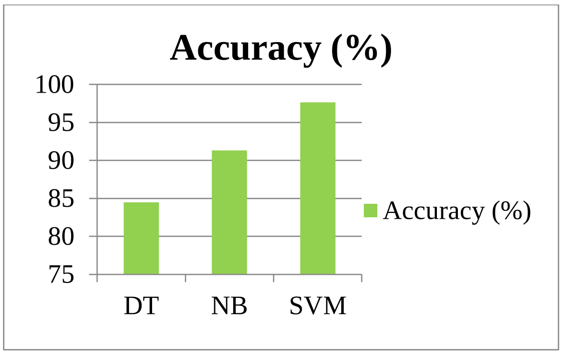

Fig. 3: ACCURACY COMPARATIVE GRAPH

Hence ML based cyber security of threat detection using intrusion system has high accuracy, sensitivity and specificity than other ML classifiers. This system effectively detected the different types of threats (R2L, DoS, U2R and probing) and abnormalities.

V. CONCLUSION

In this analysis, Machine Learning based cyber security of threat detection using Intrusion system is presented. This method recognises the four main threat types: Remote-to-Local (R2L), DoS, probing, and user-to-remote (U2R). The applied data is examined for irregularities using the intrusion detection system. The Machine Learning model for threats detection is trained using the KDDCUP 99 dataset. The Support Vector Machine classifier is utilized in order to accurately and effectively identify threats and problems. The accuracy, sensitivity, and specificity of the system's performance are evaluated. SVM classifier performance is compared to that of Decision Tree and Naive Bayes classifiers. The findings demonstrate that the proposed approach has superior performance in terms of detection accuracy, specificity, and sensitivity. This method has effectively and properly detected the various threats.

VI. REFERENCES

[1] Mohmmed Alrowaily "Investigation of Machine Learning Algorithms for Intrusion Detection System in Cybersecurity", University of South

Florida Digital Commons @ University of South Florida, March 2020.

[2] Iqbal H. Sarker, A. S. M. Kayes, Shahriar Badsha, Hamed Alqahtani, Paul Watters and Alex Ng "Cybersecurity data science: an overview from machine learning perspective" Journal of Big data, (2020) 7:41 https://doi.org/10.1186/s40537-020-00318-5, Swinburne University of Technology, Melbourne, VIC 3122, Australia.

[3] Pramita Sree Muhuri, Prosenjit Chatterjee, Xiaohong Yuan, Kaushik Roy and Albert Esterline, "Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to Classify Network Attacks", Information 2020, 11, 243; doi:10.3390/info11050243

[4] Thanh Cong Truong, Quoc Bao Diep and Ivan Zelinka, "ArtificialIntelligence in the Cyber Domain: Offense and Defense", Symmetry 2020, 12, 410; doi:10.3390/sym12030410

[5] Anuj Puri and Sumit Ray, "Interpretable Machine Learning Using Switched Linear Models For Security of Cyber-Physical Systems", 2020 Integrated Communications Navigation and Surveillance Conference (ICNS), 978-1-7281-7270-5/20, 2020 IEEE

[6] R. Devakunchari, Sourabh, Prakhar Malik "A Study of Cyber Security using Machine Learning Techniques" International Journal of Innovative Technology and Exploring Engineering (IJITEE) ISSN: 2278-3075,Volume-8, Issue-7C2, May 2019.

[7] Valliammal N. and Barani Shaju "Deep learning algorithm based cyber-attack detection in cyber-physical systems-a survey" International Journal of Advanced Technology and Engineering Exploration, Vol 5(49) ISSN (Print): 2394-5443 ISSN (Online): 2394-7454 http://dx.doi.org/10.19101/IJATEE.2018.547030

[8] Zohre Nasiri Zarandi,Iman Sharifi, "Detection and Identification of Cyber-Attacks in Cyber-Physical Systems Based on Machine Learning Methods", 2020 11th International Conference on Information and Knowledge Technology (IKT) December 22-23, 2020; Shahid Beheshti University - Tehran, Iran, DOI: 10.1109/IKT51791.2020.9345627

[9] Una-May O'Reilly, Jamal Toutouh, Marcos Pertierra, Daniel Prado Sanchez, Dennis Garcia, Anthony Erb Luogo, Jonathan Kelly, Erik Hemberg, "Adversarial genetic programming for cyber security: a rising application domain where GP matters", Genetic Programming and Evolvable Machines, 2020, Springer, doi:10.1007/s10710-020-09389-y

[10] Arunan Sivanathan, Hassan Habibi Gharakheili, and Vijay Sivaraman, "Managing IoT Cyber-Security using Programmable Telemetry and Machine Learning", IEEE Transactions On Network And Service Management, Volume: 17, Issue: 1, March 2020, DOI: 10.1109/TNSM.2020.2971213

[11] Pramita Sree Muhuri, Prosenjit Chatterjee, Xiaohong Yuan, Kaushik Roy and Albert Esterline, "Using a Long Short-Term Memory Recurrent Neural Network (LSTM-RNN) to Classify Network Attacks", Information 2020, 11, 243; doi:10.3390/info11050243

[12] Ployphan Sornsuwit and Saichon Jaiyen, "A New Hybrid Machine Learning for Cybersecurity Threat Detection Based on Adaptive Boosting", APPLIED ARTIFICIAL INTELLIGENCE 2019, VOL. 33, NO. 5, 462–482, doi:10.1080/08839514.2019.1582861

[13] Anastasia Gurina and Vladimir Eliseev, "Anomaly-Based Method for Detecting Multiple Classes of Network

Attacks", Information 2019, 10, 84; doi:10.3390/info10030084

[14] Charles Feng, Shuning Wu, Ningwei Liu, "A User-Centric Machine Learning Framework for Cyber Security Operations Center", 2017 IEEE International Conference on Intelligence and Security Informatics (ISI), DOI: 10.1109/ISI.2017.8004902

[15] Sanjay Kumar, Ari Viinikainen, Timo Hamalainen, "Machine learning classification model for Network based Intrusion Detection System", 2016 11th International Conference for Internet Technology and Secured Transactions (ICITST),**DOI:** 10.1109/ICITST.2016.785 6705