An Enhanced Intrusion Detection And Secure Mobile Cloud Computing System Using Automated Ensemble Machine Learning Classification

Haftom Gebregziabher[1], B.Ravindra Babu[2], V.Sidda Reddy[3]
[1,2]Research Scholar, [3]Associate Professor
[1]Dept of ICT, [2]Dept of CSE, [3]Dept of IT
[1]Ethiopian Technical University, Addis Ababa, Ethiopia
[2]JNTUH, Hyderabad, India
[3]CVR College of Engineering, Hyderabad, India

Abstract

Mobile Cloud Computing (MCC) is fastest growing technology era and has been presented as one of the most efficient techniques for hosting and delivering such mobile services over the internet. In this fast-growing world security and data privacy in such environment are the most challenging problem. In this scenario Machine Leaning (ML) brings an important role to detect all the attacks and providing a data privacy and confidentiality levels. In this paper, an enhanced intrusion detection and secure mobile cloud computing system using machine automated ensemble machine learning techniques is proposed. The proposed system uses an ensemble learning method for the attack detection by taking best solution from Naïve bayes, adaptive boost and part classifiers and the Training dataset Filtration Key Nearest Neighbour (TsF-KNN) classifier is used as automated data classification to classifies the data file as confidential and non-confidential based on attributes. Then the symmetric encryption methods are used to assuring privacy and confidentiality levels for the data storage. The results show that proposed system offer enhanced detection rate and improved data storage security and privacy compared to existing methods.

Key Words: Machine Leaning (ML), ensemble learning method, TsF-KNN classifier and secure data storage.

I. INTRODUCTION

Distributed computing is the most arising innovation that is quickly being embraced by the IT (Information Technology) business because of its exceptional highlights of versatility, high-accessibility, proficient asset usage, and decreased expenses. It offers omnipresent, advantageous, request based admittance to a common gathering of configurable registering assets (like stockpiling, organization, administrations, applications and workers) that can be immediately provisioned and delivered with least administration exertion or specialist cooperation [1]. Cloud offers its types of assistance to clients through the Internet, in an unexpected way. Notwithstanding, because of some intrinsic blemishes of hypervisors, joined with open and disseminated design of cloud, interlopers may get unapproved admittance to Virtual Machines (VMs) subsequently putting the clients' basic data in danger. These interruptions could increment in intricacy to frame diverse enormous scope and appropriated assaults, for example Circulated Denial of Service (CDoS) assaults, ICMP flood assault, SYN flood assault, application-level flood assaults [2]. To distinguish such pernicious exercises, it is smarter to send an organization or a host-based Intrusion Detection System (IDS) that can identify interruptions and alarm the framework director preceding any harm to cloud assets. IDS is an effective method to recognize interruptions in cloud since it gives an extra line of guard when contrasted with firewalls which is additionally used to get an organization from vindictive clients of outside world [3].

ML calculations are utilized to tackle security issues and oversee information all the more productively. ML is the utilization of man-made awareness that empowers structures to typically take in and improve really without being explicitly modified. The methodology toward learning begins with discernments or information, like models, direct arrangement or heading, to divert for structures in data and pick better choices later regarding the matter to the models that are given [4]. The goal of this examination is to execute an improved interruption location framework decide the secrecy-based class of the information into a document through AI (Artificial Intelligence) calculation and lessen the information encryption and unscrambling measure by applying information encryption to just for basic secret information. This proposed system will build cell phone effectiveness and diminishes the capacity intricacy and pursuing time encryption and decoding of the portable information. After all ML calculations a safer and successful model that contributes versatile information mystery alongside uprightness in portable environment.

II. ML FOR CLOUD SECURITY AND ITS TYPES

2.1 ML and Cloud Security

Specific effort is used to implement the logical examination of calculations and measurable models for computer system in ML. These approaches will express headings, contingent upon models and acceptance in ML. computerized reasoning is the subset of ML. Hence, ML is used in the future which is significant in the cloud. Now, let us discuss the usage of ML in security of distributed computing. cloud security is improved by proposed methodology which gives accurate identification [5].

2.2 Types of ML Algorithms

1. Supervised learning: This learning in ML will contribute the data yield sets procedure using yield subject. Preparation models are prompted based on the capacity for naming data. data science plays very important role in managed learning. A data limited for initiating the ML assignment to involve many getting ready models.

(a) Supervised Neural Network:

Yield of the information is known in supervised neural network. Real yield is compared with predicted yield in neural system. In the neural system parameters are changed based on the address. Feed-forward neural system is implemented in administered neural system.

(b) K-Nearest Neighbor (K-NN):

Characterization and regression issues are solved by the administered ML calculation technique. Genuine number is obtained in the output for regression issue [6].

(c) Support Vector Machine (SVM):

Gathering and relapse challenges are used in the regulated ML algorithm. characterization issues use this challenge. two classes (hyper-plane) are separated by the SVM classifier.

(**d) Naïve Bayes:**

Bayes' theorem is introduced by the regulated ML algorithm to highlight the regulated ML algorithm. To get effective outcomes, classifier is used [7].

2. Unsupervised learning:

The output data has no indication in the neural system. Several similarities are used to classify the data using primary occupation of the system. Diverse source of information and gatherings are used to inter related the connection in neural system.

(a) Unsupervised Neural Network: The neural framework has no prior suggestion about the yield of the data. The essential control of the framework is to group the data dependent on a few similitudes. The neural framework checks the association between different wellspring of data and get-togethers.

(b) K-Means: One of the least demanding and famous solo ML calculations. The K-implies calculation see k number of centroids, and a brief timeframe later produces every information highlight the nearest assembling, while at the same time keeping up the centroids as little as could be common thinking about the current condition [8].

(c) Singular Value Decomposition (SVD): Quite possibly the most extensively utilized solo learning calculations, at the focal point of various proposition and dimensionality decrease structures that are crucial for overall associations, like Google, Netflix, and others.

3. Semi-Supervised Learning: while training in ML, abundant unlabeled information will combine the small quantity of named information. Unsupervised and supervised learning will fall in Semi-supervised learning. The combination of labelled and unlabeled information is the main objective [9].

4. Reinforcement Learning (RL): The total proze in ML is administrated by Reinforcement Learning. There are three perfect RL models which are followed by the supervised learning and unsupervised learning. [10].

III. ENHANCED INTRUSION DETECTION AND SECURE MOBILE CLOUD COMPUTING SYSTEM

The proposed enhanced intrusion detection and secure mobile cloud computing system using automated ensemble machine learning classification is shown in figure (1). It consisting of normalization, feature selection, ensemble method, automated data classification, encryption and data storage phases as shown in figure (1). Each stage is described as follows:

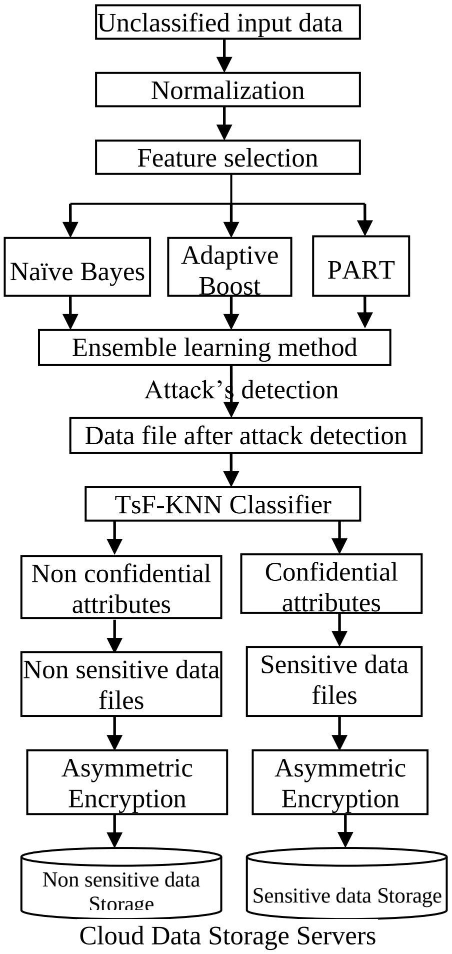

Fig. 1: ENHANCED INTRUSION DETECTION AND SECURE MOBILE CLOUD COMPUTING SYSTEM MODEL

3.1 Normalization

In the primary analysis, dataset with its finite features is utilized and the dimensionality of this dataset was reduced by performing the normalization using the formula as,

$$X_N = \frac{X - min_A}{max_A - min_A}$$

Here, X_N represents the normalized value and X is initial value. max_A *and* min_A represents the maximumand minimum attribute A value before normalization respectively.

3.2 Feature Selection

The automatically or manually selection of features from the output is known as feature selection. The accuracy is decreased by using the irrelevant features from your data. The main intent is to reduce the number of input variables. Statistical measures are used in the Filter-based feature selection methods which are independent and correlated with input variables.

3.3 Ensemble approach

Multiple learning algorithms are used in the statistics and machine learning algorithms. These multiple models are combined together to give effective results. Accurate solutions are used in the Ensemble methods compared to the single model. Machine learning competitions use ensemble methods.

3.4 Automated Data Classification

The TsF KNN calculation is applied to the preparation datasets and after filtration measure this is getting such ordered information. The customary K-NN calculation has enormous registering intricacy at information characterization steps. This enormous computational intricacy will influence low capability in information filtration. There are two kinds of information security the executives strategies out of which AI (Artificial Intelligence) information arrangement technique is utilized in the proposed structure as demonstrated in figure (1).

1) Non-Confidential Data: Public information is general information which is directly accessible to for sharing and replicating with no protection issues, for instance, authentic information, reviews, News Media, etc. This information doesn't need any benchmark and straightforwardly advanced by supported channels. Thusly, this proposed secure hypertext convention and transport layer security convention are sufficient for such information applications in customer worker transmissions for encryption and unscrambling measure.

2) Confidential Data: Significantly Highly Confidential Data is a basic information that deflected by authoritative laws, monetary business arrangements, research information, and HRD representative's information, Medical Records, Donor Information, Bank subtleties, Payment Card Data, etc. Business associations and people should have holds individual inside information may protected due to copyright, sensitive and mystery viewpoints to keep from illicit adjustment, correspondence, stockpiling and use. Since approved authority getting this information is by all accounts utilized in business and individual information at whatever point required.

3.5 Asymmetric Encryption

Utilizing deviated key cryptographic calculations, for example, ECC, RSA, and ELGAMAL with various key sizes boundary used to guaranteeing information mystery and trustworthiness towards highly confidential information. This information classification increments proficient asset use and decreasing information handling time. Morden Public key cryptography framework like Elliptical Curve Cryptography (ECC) which is perhaps the most appropriate and suggested by the U.S.

IV. RESULTS

Data set KDD cup 99 are used to carry out the ensemble learning method classification experiment. Depend on intrusion detection analysis software, headquarters was created defence developed research undertaking company DARPA. Varieties of attacks are simulated from the computer network operators. Connection files are obtained by processing uncooked facts. Few features are extracted more than 41 for every connection. 4 important categories are divided by 24 attackers such as Denial of Service (DOS), User to Route Attack (URA), Remote to Local Attack (R2LA), Probing Attack (PROBA).

Different criteria are introduced by comparing and evaluating the performance of the proposed model which are given below:

- False alarms and miss detection are indicated by False Positive (FP) and False Negative (FN).
- Packets that are correctly identified (as normal) are indicated by the True Positive (TP) and portion of correctly rejected packets are indicated by True Negative (TN).

Accuracy: To evaluate the performance of any algorithm accuracy is one of the primaries assess. Accuracy is calculated by using the below formula:

Accuracy= TN+TP/ TP+TN+FP+FN

Table 1: COMPARISON OF INTRUSION DETECTION ACCURACY

Classifiers	Features	Feature Selection Method	Ensemble learning method
Naïve Bayes	91.89	91.14	91.16
PART	99.35	99.28	99.32
Adaptive Boost	96.98	97.05	97.08
Ensemble approach	NA	NA	99.92

All through this entire technique, the time intricacy of K-NN at the testing stage was decreased and improved the information arrangement precision. The principal benefit of this filtration cycle is diminishing the quantity of reiterations in the calculation and the computational expense. The Mobile preparing test datasets for our investigation was removed from open mobile data by MobiPerf Google vault and not many of the datasets are made by the specialists. This has built up a java test system in java Net beans SDK to gauge execution of the proposed model. For this reason, the examinations were led Windows 7 Professional with incorporates Intel(R) Corei3 processor, processor speed 2.43 GHz and 4 GB RAM. We have utilized the form in cryptography classes in java climate to re-enact ECC, RSA and ELGAMAL with various key sizes for different sizes of versatile preparing datasets documents. This assesses execution of public key cryptographic calculations like ECC, RSA and ELGAMAL with different key size boundaries for various content documents information blocks. ECC offers huge timetable encryption and decoding execution benefits which appeared in fig.2, as correlation with other deviated calculations like RSA and ELGAMAL.

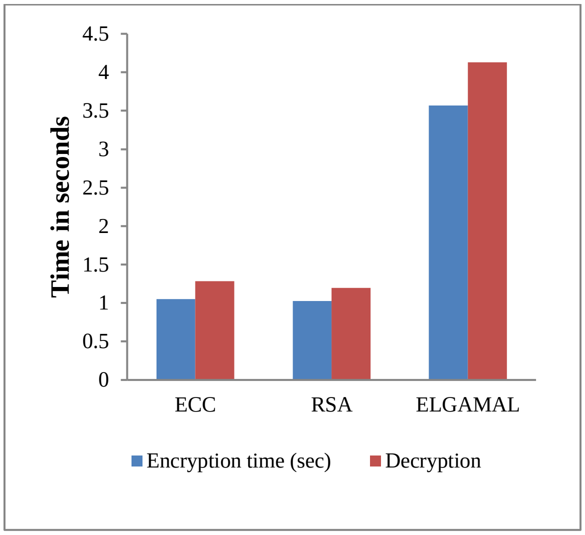

Fig. 2: COMPARISON OF ALGORITHM ENCRYPTION DECRYPTION TIMES WITH 30MB FILE SIZE

V. CONCLUSION

By using machine learning techniques, privacy based secure mobile cloud data repository model is introduced in this paper. ensemble algorithm will detect the intrusion attacks in the cloud. Privacy and integrity of critical data categorization will decrease the computation time based on the classified mobile training datasets through TsF-KNN algorithm. The effects are examined by using ensemble learning method which will bootstrap the opt for the fine components. Confidential and non-confidential data is classified based on data attributes in the TsFKNN algorithm which is an augmentation of high precision and low computational complexity. Ensemble approach will give average performance compared to classifiers. In future, another machine learning approach is enhanced with data security for crucial data handling on cloud storage-based environments.

VI. REFERENCES

[1] Kwangsu Lee, “Comments on “Secure Data Sharing in Cloud Computing Using Revocable-Storage Identity-Based Encryption”, IEEE Transactions on Cloud Computing, Volume: 8, Issue: 4, 2020

[2] Lo-Yao Yeh, Peggy Joy Lu, Szu-Hao Huang, Jiun-Long Huang, OChain: A Privacy-PreservingDDOS Data Exchange Service Over SOC Consortium Blockchain”, IEEE Transactions on Engineering Management, Volume: 67, Issue: 4, 2020

[3] Hamidreza Sadreazami, Arash Mohammadi, Amir Asif, Konstantinos N. Plataniotis, “Distributed-Graph-Based Statistical Approach for Intrusion Detection in Cyber-Physical Systems”, IEEE Transactions on Signal and Information Processing over Networks, Volume: 4, Issue: 1, 2018

[4] Ali Bou Nassif, Manar Abu Talib, Qassim Nasir, Halah Albadani, Fatima Mohamad Dakalbab, “Machine Learning for Cloud Security: A Systematic Review”, IEEE Access, Volume: 9, 2017.

[5] Yi Han, Tansu Alpcan, Jeffrey Chan, Christopher Leckie, Benjamin I. P. Rubinstein, ”A Game Theoretical Approach to Defend Against Co-Resident Attacks in Cloud Computing: Preventing Co-Residence Using Semi-Supervised Learning”, IEEE Transactions on Information Forensics and Security, Volume: 11, Issue: 3, 2016

[6] B. K. Samanthula, Y. Elmehdwi and W. Jiang, "K-nearest neighbor classification over semantically secure encrypted relational data", *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 5, pp. 1261-1273, May 2015.

[7] Tanvir Islam, Miguel A. Rico-Ramirez, Prashant K. Srivastava, Qiang Dai, Dawei Han, Manika Gupta, Lu Zhuo, “CLOUDET: A Cloud Detection and Estimation Algorithm for Passive Microwave Imagers and Sounders Aided by Naïve Bayes Classifier and Multilayer Perceptron”, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, Volume: 8, Issue: 9, Sept. 2015.

[8] Steve S. C. Ou, Brian H. Kahn, Kuo-Nan Liou, Yoshihide Takano, Mathias M. Schreier, Qing Yue, “Retrieval of Cirrus CLoud Properties From the Atmospheric Infrared Sounder: The K- Coefficient Approach Using Cloud Cleared Radiances as Input”, IEEE Transactions on Geoscience and Remote Sensing, 2013, Volume: 51, Issue: 2, 2013

[9] X. Zhu and A. B. Goldberg, "Introduction to semi-supervised learning", *Introduction Semi-Supervised Learn.*, vol. 3, no. 1, pp. 1-130, 2009.

[10] J. Zhang, M. Zulkernine and A. Haque, "Random-forests-based network intrusion detection systems", *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 38, no. 5, pp. 649-659, Sep. 2008.

[11] T. Spyropoulos, K. Psounis and C. S. Raghavendra, "Efficient routing in intermittently connected mobile networks: The multiple-copy case", *IEEE/ACM Trans. Netw.*, vol. 16, no. 1, pp. 77-90, Feb. 2008.

[12] J. H. Cheon, N. Hopper, Y. Kim and I. Osipkov, "Provably secure timed-release public key encryption", *ACM Trans. Inf. Syst. Security*, vol. 11, no. 2, pp. 4, 2008.

[13] X. Lin, X. Sun, P.-H. Ho and X. Shen, "GSIS: A secure and privacy-preserving protocol for vehicular communications", *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3442-3456, Nov. 2007.

[14] Y. Zhang and Y. Fang, "A fine-grained reputation system for reliable service selection in peer-to-peer networks", *IEEE Trans. Parallel Distrb. Syst.*, vol. 18, no. 8, pp. 1134-1145, Aug. 2007

[15] .A. Miyaji, M. Nakabayashi and S. Takano, "New explicit conditions of elliptic curve traces for FR-reduction", *IEICE Trans. Fundam.*, vol. E84-A, no. 5, pp. 1234-1243, May 2001.