ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

A Light Weight KNN Based Continuous User Authentication Scheme for Vehicular Clouds

[1] Shailaja S.Mudengudi, [2]Mahabaleshwar S.Kakkasageri
[1] Assistant Professor, Department of Electronics and Communication, Tontadarya College of Engineering,Gadag-582101 (Karnataka), INDIA
[2] Associate Professor, Department of Electronics and Communication Engineering, Basaveshwar Engineering College (Autonomous), Bagalkot-597-102, (Karnataka), INDIA

Abstract: Authentication of the genuine nodes is one of the important aspect in addressing the security issues in Vehicular Cloud Computing. The authentication methods followed involves the use of secret key generated by the credentials of the vehicle owner which is used for encryption or decryption of messages or to sign them. If this key gets compromised and malicious node gets access to this secret key then the complete service is disrupted. In order minimize the effect of compromised key we present a light weight continuous authentication scheme based on KNN classifier algorithm to validate the user, which can be installed along with the existing authentication schemes with little modifications. The simulation results show the proposed KNN classifier is the simplest and fastest classifier in comparison with other classifiers which makes the presented framework light weight. Using feature selection there is slight decrease in the accuracy but it does not degrade the performance of the framework to a greater extent. Security analysis shows that the proposed scheme is able to withstand most of the attacks efficiently in terms of Network Configuration Time, Kappa statistics and F- measure.

Keywords: VANET • Security • Authentication • KNN classifier

I. INTRODUCTION

In the recent years there is an exponential increase in the number of vehicles especially in metropolitan cities. Managing such huge traffic is tedious and needs to be addressed. Intelligent Transport System (ITS) intends to support the management of the large vehicular network by providing services like traffic and road conditions, route planning, navigation, road safety, vehicle control & monitoring applications, and many more. It also has many applications for users to ease the driving experience such as e-commerce, interactions between different users, information services and multimedia services like messaging, voice over IP etc. Applications like Apple CarPlay, Android Auto facilitates are connected to car services. VANET (Vehicular ad-hoc network) is an integral part of the ITS. It uses resources like On Board Unit (OBU) along with Road Side Unit (RSU) infrastructure to provide the above mentioned services [1]. The communication link is very short termed. The vehicles in the vehicular network range from high end to low end employed with lot of

advanced sensors which facilitates sensing, computing further they can exchange the data with others. As the nodes in the VANET are continuously moving at a extremely high speed it becomes very difficult to handle issues such as mobility management, data aggregation, data validation, routing, privacy and security [2]. In order to avail all the above mentioned services there is need for resources and infrastructure which provide the platform in order to utilize the services offered by VANET. So it is hardly feasible to employ all the resources on each and every vehicle, which is uneconomical also. It would be rather more effective, efficient and economical if the resources were rented. There would be great reduction in the cost if the users were allowed to request services and pay only for them. This can be achieved by Cloud Computing (CC). Applications like Google Apps, Microsoft exchange, many social media all are based on cloud, but its actual analysis and study is recently taken up by the professionals [3]. From the sources presented in [4] the worldwide market of public cloud services is

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

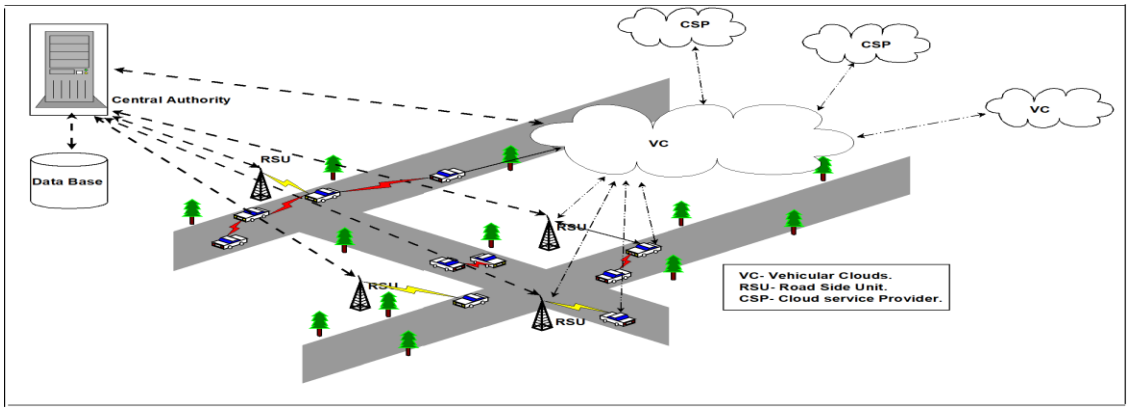

Fig 1: Vehicular Cloud Network

predicted to grow from 182.4 billion dollars to total 214.3 billion dollars in the year 2019 which is 17.5 % increase. It is clear from the figures that cloud computing has become foundation to meet the business needs of the Information Technology (IT) companies. In fact it is difficult to find a business model which is directly or indirectly not influenced by the cloud computing technology, which makes it inevitable. Gartner in [5] predicts that the market is going to increase three fold times by 2022 in comparison with growth of overall IT industry. With such high potential the one concern for its adoption by the companies is security and privacy. Mobile cloud computing technology combines cloud computing, mobile computing, and wireless network to provide various services to the user. But it faces several challenges such as authenticity, reliability and privacy which preclude its usage. With such huge advancement in the wireless technologies in order to make driving experience more smart and intelligent in automobile industry a new hybrid network Vehicular Cloud Computing (VCC) has been developed where VANET is combined with Mobile cloud computing and wireless sensor networks. A strong authentication method increases the security of VCC and eliminates most of the security threats. When any new user registers with a trusted Central Authority (CA) it is given with unique Identification number (ID) which may be used to generate a unique key. Before start of any service from the cloud it is to be assured about the user. So a set of messages will be transferred initially to authenticate the user where this ID or unique key will be used. Message authentication on the other hand authenticates the message received. Every message sent is signed by a unique signing key allotted to the sender by the CA. But in both the cases if for any reason the unique key is compromised then the result is hazardous. To minimize the effect of compromised key in this paper we present a two tier authentication scheme based on K-Nearest Neighbour, which can be installed along with the existing authentication schemes with little modifications. Our contributions are as follows:

1. We present a light weight authentication scheme based on KNN classifier which will be working along with the existing secret key based authentication framework. This scheme will be minimizing the loss due to secret key compromise. The overhead is negligible as the data need for processing is already present within the VCC.
2. Security analysis shows that the proposed scheme is able to withstand most of the attacks efficiently.
3. The authentication delay and error rate is very less compared to existing schemes.

The remaining paper is organized as follows: In section 2 we present the background of the security and privacy issues in VCC and data classifiers. Section 3 the detail proposed authentication framework is presented. Section 4 presents the simulation process followed by results and discussion of the results. Lastly we conclude in section 5.

2. RELATED WORK

2.1 Vehicular Cloud Computing.

Due to the huge increase in the number of vehicles abundant resources present with the vehicles are

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

expected to be under utilized. Such resources can be shared with the users who are in need of the resources in an economical way which can be achieved by cloud computing [6]. Olariu et al., in [7] has presented such model where VANET and Cloud Computing are combined to form Vehicular Cloud Computing (VCC). But no architecture or structural model is presented. In [8] authors have presented an architecture for the VANET based on cloud computing. The architecture is divided into Vehicular Clouds formed by the resources on the vehicles, vehicles nodes use traditional Clouds and Hybrid Vehicular Clouds. Resources can be pooled by the vehicles which may be static or in motion. Static vehicular clouds provide long term services as they have long life time compared to dynamic vehicular clouds which can provide disposable clouds. In hybrid vehicular cloud on the other hand vehicular cloud access the traditional clouds. Each vehicle is equipped with OBU (On Board Unit), advanced sensors such as temperature, pressure, smart phone sensors, sensors inside the vehicle etc. The OBU can communicate using broadband wireless communication links such as Wi-Fi, WiMax, Wireless Access in Vehicular environment (WAVE) and Dedicated Short Range Communication (DSRC). VANET and CC have security and privacy related issues which are specific to their domain. As VCC is a combination of both technologies, the security and privacy issues are unique. Particularly the unstable communication links established due to high mobility of the nodes make it difficult to attain the security. Methods or frameworks used to deal with security and privacy issues used for conventional VANET and CC cannot be applied for VCC. In [9], author presents the security and privacy issues relevant to VCC. VCC is aimed to provide critical services related to safety of the drivers and provide comfortable driving experience. In due course critical data will be pushed to or pulled from the clouds by the vehicular network. There is always risk of this crucial data being exposed to attacks where data may be compromised which may endanger the purpose of VCC making management of transport system even more tedious which may put passengers life in to danger. It gets even worst if the data is private which may lead to many other kind of attacks like denial of service (DOS), Sybil attack, illusion attack, and man-in the-middle attacks.

2.1.1 Security and privacy issues in Vehicular Cloud Computing.

VCC leverages on group of autonomous vehicles, which are readily interested in renting the large amount of excess resources available with them. This makes both parties economically benefited; one who is renting the resources as the resources were under utilized and the one who is using them as they need not invest in buying the resources. This ‘pay as use’ model helps to rent the resources as per need and only pay as per usage. As discussed earlier one of the major challenges faced by the VCC is security and privacy as in the case of other wired and wireless networks. Author in [10] classifies them in to three layers (i) Network layer consisting of VANET, Wireless Sensor Network (WSN) (ii) Transmission channel layer (iii) Cloud Computing layer. Security issues in network layer include confidentiality, authentication, non-repudiation, localization, data verification. Major threats for VCC are DOS attack, identity spoofing, data tampering,repudiation, Sybil attack, information disclosure. The author further claims that it is impossible to provide security to VCC if confidentiality, integrity and authentication and privacy issues are not addressed. The issues seems to be similar to those experienced by WSN, cloud computing and VANET, but when carefully studied and analyzed they are different and can be said as VCC specific. The work presented in [11] also lists down the threats with respect to security of VCC. An attacker model is presented to illustrate all possible kinds of attacks on VCC topology. The main security and privacy in VCC include Denial of service (DOS), Identity Spoofing, Tampering, Repudiation, Information Disclosure, Elevation of privileges, Authentication of Nodes, Establishing trust between the nodes, Scalability, Heterogeneous Network Nodes. Among the above, several issues can be addressed by having a strong authentication framework for the nodes [12][13] such as identity spoofing, Trust establishment, Repudiation. Several authentication methods and frame works have been proposed. Authentication can be broadly classified in to two categories: user authentication and message authentication [14].

As already discussed VCC has evolved from CC, VANETS and WSN. In this section we discuss the above technologies, existing security and privacy issues

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

and solutions for the same. 2.1 Wireless Sensor Networks It is large network composed of sensors having limited storage, computation and communication capabilities. Application of WSN is huge and still growing ranging from life critical medical applications to military, industry, science, pollution monitoring. Data generated by the sensors will be collected and aggregated to derive useful information. But the data is liable to be attacked and protecting it is crucial and critical. The attacks are broadly classified into Communication Attacks, Attacks against Privacy, Sensor-Node-Targeted Attacks, Power Consumption Attacks, Policy Attacks, and Cryptographic Attacks on Key Management [15]. In [16], a security framework for WSN is presented. Symmetric key encryption is used to secure communication between sensor mote and sink. A counter common between the mote and sensor is used for key generation. The framework attains semantic security, mote authentication, resistant to replay attack. In order to elevate the security issues like broadcast source impersonation and data modification attacks. In order to overcome these issues many Broadcast Source Authentication Protocols are present. One such protocol is presented in [17], which concentrate on authentication of the broadcast source at every hop rather than verifying at the last i.e at destination. Due to which the loss or the damage is limited to next one hop only. This solves the problem of overloading the buffer and battery life depletion due to attacks intended to resource-draining. A list of security requirements, attacks and possible countermeasures for the attacks are listed in [18]. In WSN, data collected and aggregated is of utmost importance. Data authentication, freshness, authentication of the sender, controlled access to the data is among the top of the list. The methods that can be achieved by cryptography using shared or symmetric keys, data aggregation etc. WSN is also prone to Sybil attack, Wormhole attack, DOS, HELLO Flooding attack, Sink-hole attack.

Sink-hole attack is accomplished by a single sensor node possessing more than one legal identity attempting to degrade the security of data in WSN. Sybil attack creates multiple pseudo identities there by undermining the reputation of the system and can be elevated using authentication and encryption methods. Attack in which two hostile nodes present an attractive path for the legitimate nodes by providing a path with less latency or with less number of hops is the wormhole attack. These hostile nodes relay the incoming data packets, which they drop arbitrarily. Both nodes communicate using a dedicated channel. The hostile nodes are difficult to trace as the attack is executed only by the hostile nodes without attacking any of the genuine nodes and they are invisible to other legitimate nodes. Several counter measures are discussed in [19] such as restricting the movements of packets either by geographical or temporal leashes. Geographical leashes restrict the packets in a confined area or distance which gives rise to the requirement of localization system. Whereas temporal leashes requires local clocks which are accurate enough to restrict the packets lifetime. Instead of equipping each sensor node with location aware systems like Radio Frequency or Global Positioning System, a worm hole defense method is presented in [20] called WODEM which has some nodes called as detectors which are installed with location aware system. Initially in scanning phase detectors scan their counterpart side detectors. The outcome is loss exponent and traversed number of hops of that path with which wormhole attack is further detected. Wormhole attack exists if the longest possible distance for that hop count is less than the actual distance. Location of the attacker node is also detected by transmission range and time-to-live range of the control packets. Denial of service leads to a great problem in applications like health, military, government etc. Here a malicious node consumes all the services and resources making flooded and not available for authentic users. DOS attack is being addressed and a protocol for access control is presented in [21]. It uses mutual two way authentication and a query authentication process between the sensor nodes to achieve it. HELLO Flooding attack can be called as a subtype of DOS attack where a large number of advertisement packets are sent to the cluster head so as to degrade its service and performance as it is unable to receive request from other nodes [22]. A dynamic routing method is proposed in [23] to defend the flooding attack, which detects the malicious nodes by measuring average transmission route request packets. Detected the malicious node will be added in the detection list so as to reduce its privileges further. 2.2 Security and Privacy Issues in VANET's VANET play a crucial role in increasing the safety of drivers and

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

increase the efficiency of traffic management. So security of the traffic related data and privacy of the user is a hot area of research. In [24], author discusses the importance of strong authentication system in the VANET. Several types of existing authentication schemes based on pseudonymous, group signature, ID-PKC, hybrid anonymous and their shortcoming have been discussed. To address issues like guessing and replay attack, a low cost elliptical curve based mutual authentication framework has been presented in [25]. It uses elliptical curve and XOR operation to satisfy the security requirement such as user anonymity, conditional trace-ability and produce a good trade-off between security and performance parameters. Identity and location information of vehicle owner is protected in [26] by using homomorphic encryption. The randomized authentication is followed by the protocol which ensures that the vehicle cannot be tracked by entities such as service providers, servers or even other peer vehicles. But in case of dispute identity tracing in real time is easily achieved. Authentication schemes are broadly divided in to four categories in [27]. HAB (Huge Anonymous Based) authentication protocol requires anonymous certificates in large numbers and their respective secret keys. The certificates do not carry any information related to the user which assures their privacy. The management of certificate credentials is done by trusted authority. If a message is to be signed, randomly any certificate from certificate pool is chosen and the corresponding secret key is taken. To verify the sender sign the corresponding public key is used by the receiver. The problem with this system is handling the huge numbers of certificates and revoking them. In group signature method of authentication, a member from the corresponding group signs the message anonymously. The identity of the signing member is revealed neither in signing process nor in verification process. But during any dispute, trusted party can reveal the identity of the signing group manager. This system also faces the problem of grater delay in verification process, when there is large number of revoked vehicles. The complexity further increases as it becomes rather difficult to change or do modification for the secret key, because it should be done by the trusted party using a secure channel. On the contrary if the group size is made small, identifying the members of the group becomes easy making it liable to attack. Vehicles authenticate each other using the expensive on line RSU's. These schemes are more reliable as they do not involve vehicle to vehicle communication links. The computations are comparatively less and the RSU's are capable of long range communication with high capacity and transmission rate with sufficient processing resources. Event Data Recorder (EDR) and the Tamper-Proof Device (TPD) are used in TPDB (Tamper Proof Device Based) authentication scheme. The EDR provides tamper-proof storage providing data which will be helpful in investigation of accidents. TPD is capable of performing cryptographic processing, maintain the secret, public, master keys and generate pseudo identities. The real identity of the signer is not revealed, preserving the privacy and anonymity as well. But the scheme also faces the challenge in revocation process because critical data such as ID, master keys are involved. The discussed methods emphasis the role of authentication in preserving privacy and enabling security in the networks.

2.3 Security and privacy issues in Mobile Cloud

Computing Mobile Cloud Computing (MCC) is extended version of Cloud Computing (CC) where the cloud is formed by mobile nodes to provide demanded services which are elastic and can be measured. In [28], author presents the difference between traditional CC and MCC. An architectural framework has also been presented for MCC which three different layers: mobile user, mobile network and CSP layer. Among the challenges faced by MCC for its adoption security, trust and privacy tops the list. User authentication and authorization play a vital role in the attaining the above challenge. A detailed study on different authentication methods for MCC is presented in [29]. The MCC authentication is broadly divided in to user side and cloud side with further subdivisions based on credentials used for authentication. Expressing the importance of authentication the author states that several major security, privacy and trust issues such as impersonation, intractability, replay, unlink-ability and man-inthe-middle attack. Port-knocking authentication [30] is a method in which the port receiving the data packets need not be open all the time. It can either accept or reject the packet. Due to its light weight property these are more preferable in MCC then traditional authentication methods.

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

2.4 Data Classifiers

There is substantial growth in the data volume due to advancement in technology. It may be the banking sector, sensors in health applications, networks, storage there is huge amount of information that needs to be handled effectively. Data analysis has to be done so as to obtain accurate conclusions which reduce time from such huge data. Data mining is one such statistical method which uses classifiers to extract the information precisely. There are many classification algorithms which give different results for different data sets. A single classification algorithm may not address all the aspect of classification. So before choosing the classifier it is best to examine and compare different classification algorithms and choose the one with best accuracy compared to others [31]. Classifiers applications range from speech recognition to document classification. They are broadly classified in to binary (only two distinct classes) and multi class classifier (more than two distinct classes). Data mining is also being used to study its customer's behavior using the data collected. Data pattern are to be extracted from the database collected which may use methods like classification, clustering, rule based, regression etc. In classification method Bayesian network, Support Vector Machine (SVM), K-NN classifier, Decision tree, random tree [32] are used.

3. PROPOSED KNN BASED AUTHENTICATION SCHEME FOR VEHICULAR CLOUDS

In this paper, we present a light weight authentication scheme for VCC, which continuously authenticates the user based on K Nearest Neighbor classifier algorithm. The VCC architecture for proposed work is shown in figure 1. It consists of Vehicle Nodes (VN), Road Side Units (RSU), Cloud Service Provider (CSP) and Central Authority (CA). We assume that an encryption based authentication scheme which may use any of the methods like the light weight and efficient privacy preserving authentication method presented in [33] is already implemented as a first tier of our system. Key management Centre (KMC), a trusted authority takes care of all the registration of all vehicles, key management, RSU, maintain details of all vehicles. For storage and cryptography execution, Tamper Proof Device (TPD) is used which is secured against any kind of attack. Road side Unit is capable of communicating with vehicles OBU and KMC as well. RSU takes part in communication of messages and updating the keys. As the secret key based authentication is already on the system we will not be addressing the procedure. We assume that the user is authenticated using the secret key and the communication is established. The VCC is serving the user with required services opted by the user. Our proposed framework assumes:

1. Each entity in the VCC network should register with the CA.
2. The Data Base (DB) storage associated with CA is TPD having all the credentials required for the proposed framework. It stores details about all the vehicles, CSP's, RSU's. Only CA can access the credentials from this storage. CA also takes care of the billing, based on the usage of the resources.
3. Vehicle nodes can communicate with each other in limited vicinity and with the RSU using short range communication such as WAVE, DSRC.
4. Only RSU's are able to communicate with the CA. So all the service request from the vehicle nodes will be done via RSU's.
5. The VCC is formed by resources pooled by the vehicles, RSU's, CSP's.

As new vehicle register to the CA, an Identification number (ID) is given to the vehicle. The Data Base (DB) contains data regarding past history of each vehicle's cloud associated activities. It is stored safely in DB which is tamper proof and secure. Algorithm for the proposed work is as shown in algorithm 1. The K-NN algorithm [34] works as follows in the proposed work. K-NN is a supervised classifier, which classifies the test sample based on the majority nearest neighbors which have the same posterior distribution.

1. Start
2. The first tier of authentication is performed using secret key generated using ID.
3. Collect the information regarding the ongoing service and store it in database DB.
4. For every 't' seconds run the proposed KNN based authentication scheme.
5. If a malicious node is detected in step 3, terminate the services from the cloud. If the

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

node is authentic user continue to step 3. Otherwise terminate the services.

6. End

Algorithm1: Proposed Authentication Scheme

The number of neighbors considered depends on 'k' value, which specifies the number of nearest neighbors to be considered. Let there be 'g' classes C1, C2,....Cg. Suppose the test sample has 'n' attributes which is to be classified to any of these 'g' classes based on the maximum class to which the nearest neighbors belong. The unknown sample S has 'n' attributes or features as shown in equation 1.

$$S_{ID} = [s01, s02, s03 \ldots s0n] \quad (1)$$

S_{ID} is the testing data set for the KNN classifier algorithm which is trained with data set D_{ID} which is the past record of the user present in DB. ID represents the identification of the user.

- Collect the current credentials regarding the ongoing service. This will be our test sample S.
- Get the data set 'D' from the DB which will be training data set.

$$D_{ID} = \begin{bmatrix} m01 & m02 & c1 \\ m11 & m12 & c2 \\ mp1 & mp2 & cg \end{bmatrix}$$

- Select the value of 'k' which represents the number of neighbors to be considered.
- Calculate the similarity between the test sample and training sample for each criterion in the row using equation 2.

$$\mathrm{D(Ai, At)} = \sum_{i=0}^{k} \alpha j\ (Si)\mathrm{sim}(Si\ , \mathrm{D}(i, j)) \quad (2)$$

Where aj (Si) ϵ 0,1. If Si ϵ Ωg is true then aj (Si =1)else aj (Si =0) Sim (X,Xi) represents the similarity functions denoting or representing the similarities between the test sample 'SID' and training sample 'DID'. The similarity function is calculated using any of the below distance functions. There are many distance functions such as Euclidean Distance, Hamming Distance, Manhattan Distance. The Minkowski Distance is the generalized distance function given in equation 3 which finds the distance between the samples 'X' and 'Y'.

$$\mathrm{Distance(X, Y)} = \left[\sum_{i=1}^{n} |\mathrm{xi} - \mathrm{yj}|^{\mathrm{p}}\right]^{\frac{1}{p}} \quad (3)$$

It works on the normalized vector space as properties like zero length for zero vectors, scalar factor, and triangle inequality are satisfied. Variations in equation 3 generate different distance functions as follows. In equation 3, if p=1 then we get Manhattan Distance as equation 4.

$$\mathrm{DistanceM(X, Y)} = \sum_{i=1}^{n} |\mathrm{xi} - \mathrm{yj}| \quad (4)$$

If p=2, we get Euclidean distance as in equation 5.

$$\mathrm{DistanceE(X, Y)} = \sqrt{\sum_{i=1}^{n} (\mathrm{xi} - \mathrm{yj})^{\mathrm{p}}} \quad (5)$$

Dataset Machine learning algorithms are used and designed to detect and unmask taint intentions by the malicious nodes which make the authentic nodes malignant and deprived of privileges. Intrusion Detection System can detect offline misuse attacks and online anomaly based attacks [35]. There is need for data set to evaluate the IDS. A comparison between such data sets is presented in [36] where MIT network packet data set and KDD CUP 99' network data sets are used.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

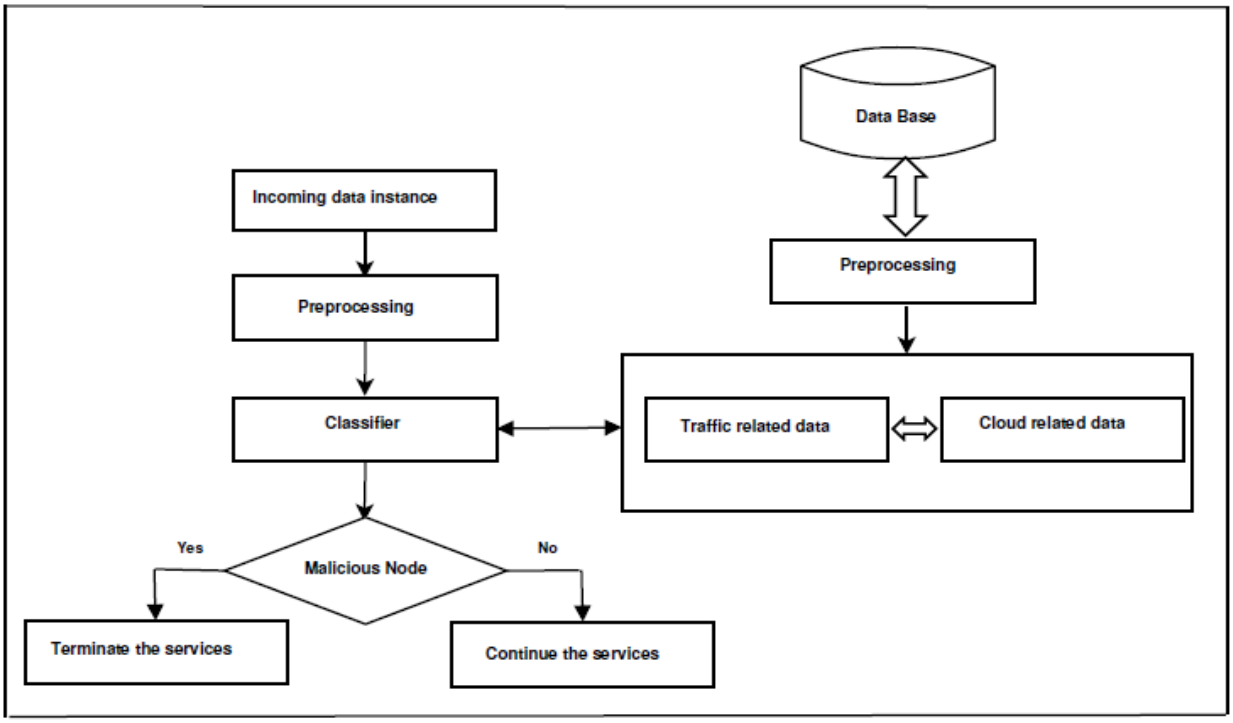

Fig. 2: Proposed KNN based Authentication Scheme for Vehicular Clouds

The KDD data set has a prominent role in the research of security analysis in the network. It is basically used by machine learning algorithms as training and testing data sets. A detailed analysis of the KDD CUP 99' data preformed and the inherent problems associated with it are presented in [37]. Author presents the statistics and analysis of the KDD CUP 99' which are summarized in table 1. Each data instance contains 41 features and characterized as either normal or attack instance. The data set addresses 4 types of attack: 1. Denial of Service: The perpetrator prowls the resources so as to disrupt the services given to a legitimate user [38]. 2. User to Root (U2R): The normal user attempts to use super privileged services or resources not legally subscribed. 3. Remote to Local (R2L): Attacker who is remotely located tries to gain access on targeted machine by generating some vulnerability by continuously sending packets. 4. Probing: Attacker tries to collect information or vulnerabilities from the network which is used to hack the system [39]. The attributes of the data set can be divide in to basic, traffic and content related. Basic attributes are a derivative of the connection related information. Traffic attributes are derived from window interval which is further divide in to same host or service based. But the interval of 2 seconds is considered which will be unable to detect attacks with frequencies greater than 2 seconds. To avoid this issue, connection window is considered instead of time window. Some attacks are data oriented rather than connections. Some attacks like R2L and U2R are executed only in one connection. These attacks can be detected based on information related to suspicious activity like number of failed password entry, reading secured data memory etc. which are the content attributes. The amount of redundant data in KDD CUP 99' training and testing data set is huge. This leads to biased outcome from the learning algorithms.

4 SIMULATION PROCEDURE

We consider the VCC network as shown in figure 1. The procedure adopted for the proposed work is as shown in figure 2. – Vehicle Node: Similar to a traditional vehicle node as in VANET. It is capable of communicating with neighboring nodes and the RSU. It is having OBU with plenty of resources to compute and to store. – Vehicular Cloud (VC): The resources of the vehicles and the RSU's willing to lend on rent are pooled together to form the VC. All the communication happens via RSU's. – Road Side Units (RSU): The RSU's act as Gateway for all the communications of messages and the transferring of

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

services. The RSU's are also equipped with good amount of resources which will be shared with the cloud.
Cloud Service Provider (CSP): Same as Traditional cloud, where resources are pooled and follows pay as use model. Various CSP's are linked with VC and Vehicle Node (VN) can render services from CSP's also via the VC. – Central Authority (CA): The most trusted entity responsible for coordinating all the activities in the VCC. Service request are the tasks handled by CA. – Data Base (DB): Each VN, CSP, RSU, VC details are stored in DB. Each entity registered with CA will be having a list of credentials like ID, services interested, services provided, cloud service related details. The proposed work acts as second tier of authentication process. The first tier is authentication of the two entities, service provider and the service requesting node. This step is out of our focus area. We use different standard data sets on the Weka tool [40], which is a collection of machine learning algorithms. The outcome of the processing is used to decide the authenticity of the nodes.

4.1 Simulation setup

Simulation of the proposed framework is done for various scenarios of authentication in "C++" programming language. The VN are spread randomly on two lane road moving in both directions. Each VN is capable of communicating with neighboring vehicles and RSU's using DSRC links up to a distance of 1Km. Vehicle with 4G/5G facility can directly communicate with VC. Data is collected by the CA from the RSU's about the vehicles regarding the past history. This data base is the training data for the proposed framework. The algorithms will be activated for every 't' seconds. Data collected in this time window will be our testing data. The following steps are executed:

1. Collect the training data from the RSU's and previously stored data in CA.
2. Pre- process the data.
3. Collect the data collected in the time window 't'.
4. Apply the data for the proposed authentication scheme based on KNN. Calculate differences between the data and the Data sets present in the DB. Sort the differences and choose the first k values. Search for the most frequent class. This will be the output class 'Cg' of the data being processed.
5. The output confirms about authentic or suspicious node.

5 **Result and Discussion**

Results of the proposed KNN based authentication scheme are presented in the section. The scheme has been compared with several authentication schemes.
– Network configuration delay: It is the time taken by the network to collect the data related to authentication from the DB. It is calculated in seconds.
– F-measure: It is the accuracy measure given by harmonic mean of Recall (R) and Precision (P). 'R' and 'P' are defined using confusion matrix also known as error matrix. Confusion matrix is the summarized vision of classifier performance in a tabular form. It presents the level of confusion a classifier encounters. It is generated by the number of correctly and incorrectly classified predictions by the classifier with respective classes.

Class	Authentic Node prediction	Malicious Node prediction
Authentic Nodes	TP	FN
Malicious Nodes	FP	TN

Table 1: Confusion Matrix

TP- True Positive, FN- False Negative, FP- False Positive, TN- True Negative.

Several other performance measures are derived from CM. Some are mentioned below. Accuracy of a classifier is the measure of how accurately the classifier predicts the class and is given by equation 6.

$$\text{Accuracy(A)} = \frac{(\text{T P} + \text{T N})}{(\text{T P} + \text{T N} + \text{F P} + \text{F N})} \quad (6)$$

Recall is the measure of how good the classifier is in recalling the positives from the knowledge of training data. It is given by equation 7.

$$\text{Recall(R)} = \frac{\text{T P}}{(\text{T P} + \text{F N})} \quad (7)$$

Precision on the other hand tells us about how pr´ecised the classifier is in predicting the positive class. It is calculated using equation 8.

$$\text{Precision(P)} = \frac{\text{T P}}{(\text{T P} + \text{F P})} \quad (8)$$

F-Measure is calculated using equation 9.

$$\text{F} - \text{Measure(FM)} = \frac{(2 * \text{R} * \text{P})}{(\text{R} + \text{P})} \quad (9)$$

Kappa statistics: As we have seen the F-Measure is not considering TN. Some times F-Measure is not considered as ultimate tool to compare the classifiers because it has flaws such as it is biased by majority class, focuses on only one class, consider even distribution of the classes which is not true in real time scenario. So another

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

measure of performance considered is the Kappa statistic used to evaluate the classifiers among themselves. It is comparison between observed accuracy (A) which is the total number of correctly classified instances to expected accuracy which is the accuracy expected of any random classifier given by equation 10.

$$\text{Kappastatistics} = \frac{(\text{ObservedAccuracy} - \text{ExpectedAccuracy})}{2a(1 - \text{ExpectedAccuracy})} \quad (10)$$

Classifiers considered: Proposed KNN authentication is compared with Support Vector Matrix (SVM), Naive bayes , J48, Random Forest as these are the most commonly used classifiers.

– Training Instances and Features: Each instance is the list of all the features considered in a predefined sequence. The number of instances and features considered are varied in simulation.

A good classifier has a high Kappa statistics, Accuracy, F-Measure which range from 0 to 1, where 0 is the least and 1 represents the highest value. We have already seen the equations for distance function in section 3. In figure 3 a plot of network configuration delay versus number of features is shown. In figure 4 a plot of network configuration delay versus number of instances is shown. We can observe that as the number of features and instances increases there will be increase in the delay. This is the result of increase in the data to be processed. Now we will consider the performance of proposed authentication scheme in comparison with different classifiers for different attack scenarios. The performance parameter considered is the kappa statistics and F-Measure. A plot of Kappa values against Different classifiers is presented in figure 6, 7, 8 and 9. It shows that the proposed scheme is having kappa value greater than Naive bayes, J48 and SVM. The kappa value is almost equal to Random forest classifier. But the Network configuration delay of Random Forest classifier is very large when compared to proposed authentication scheme which makes it not suitable for VCC as time is crucial. The same is true for the F-Measure metric performance which is shown in figure 11, 12, 13 and 14. The F-Measure is relatively same as that of Random Forest Classifier but more than Naive bayes, J48 and SVM. The proposed KNN based Authentication scheme defends the attacks better than other classifiers. We can also observe that there is a slight decrease in the F-Measure and the Kappa value when the numbers of attributes or features are reduced. But such reduction is not degrading the performance of the system to a greater extent. The attacks are handled effectively by the proposed KNN based algorithm. The variation in F-Measure and Kappa statics with changing ‘k’ values is shown in figure 16 and 17 respectively for different distance functions. It can be observed that as ‘k’ increases the efficiency of the framework decreases. It is important to choose appropriate ‘k’ value to get optimal result. Euclidean distance function and Manhattan distance function show more stability when compared to Chebyshev distance. function making it unsuitable.A plot of Accuracy versus ‘k’ values in the KNN algorithm is presented in figure 15 for different distance functions With data set having 22683 training data set, 6492 testing data set with 42 attributes. We can observe Euclidean distance, Manhattan distance and Mahalanobis distance have almost same accuracy rate. Whereas

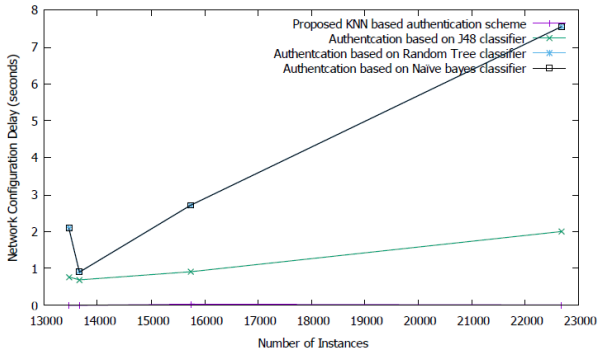

Fig. 3: Network Configuration Delay Vs. Number of Instances

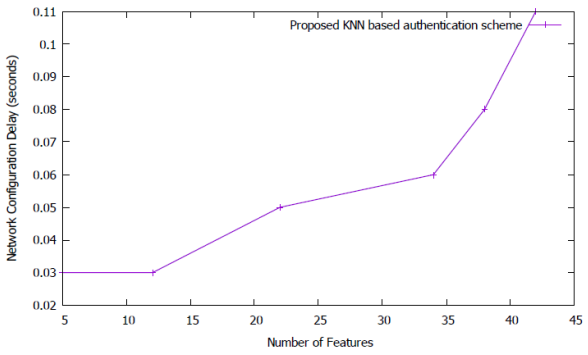

Fig. 4: Network Configuration Time Vs Number of Features

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

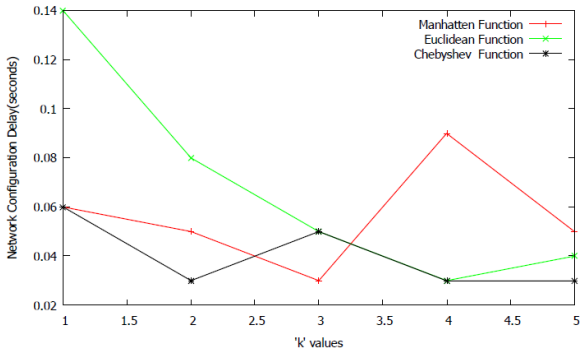

Fig. 5: Network Configuration Time Vs 'k' for Different Similarity Functions

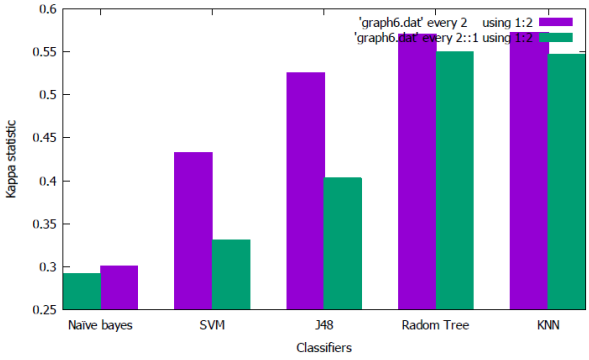

Fig. 6: Kappa statistics for DOS attack Vs. Classifiers

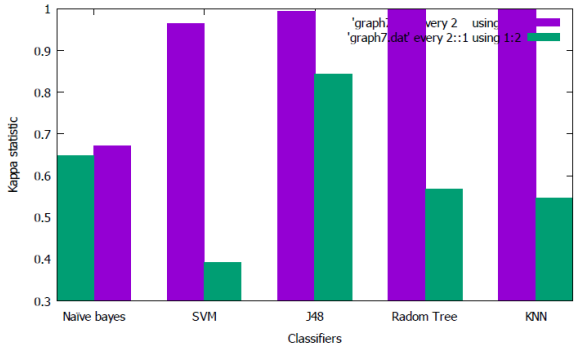

Fig. 7: Kappa statistics for Probe attack Vs. Classifiers

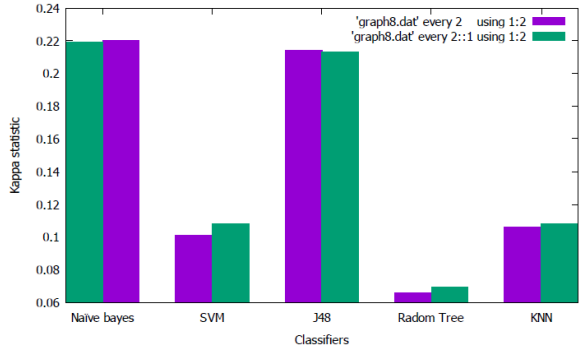

Fig. 8: Kappa statistics for R2L attack Vs. Classifiers

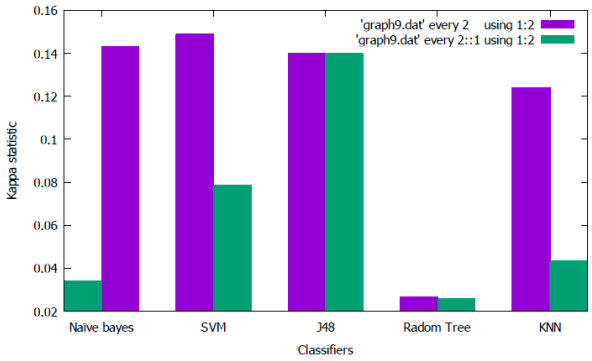

Fig. 9: Kappa statistics for U2R attack Vs. Classifiers

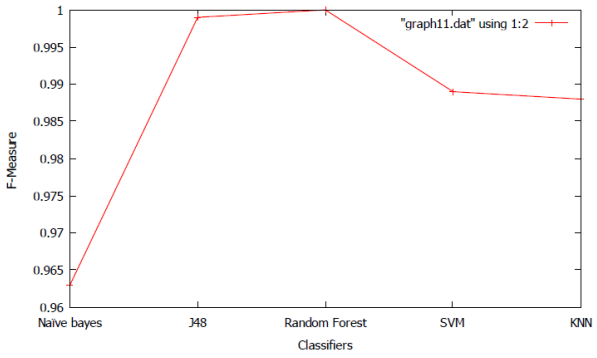

Fig. 10: F- Measure Vs. Classifiers

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

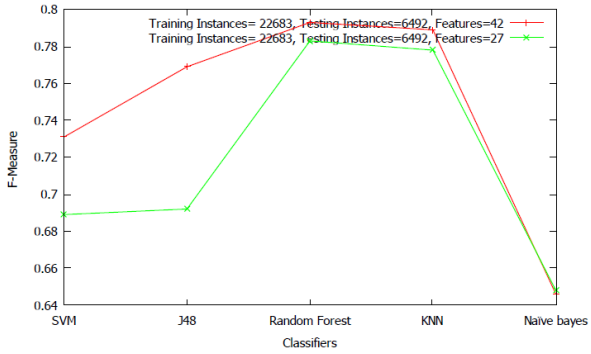

Fig. 11: F-Measure for DOS attack Vs. Classifiers

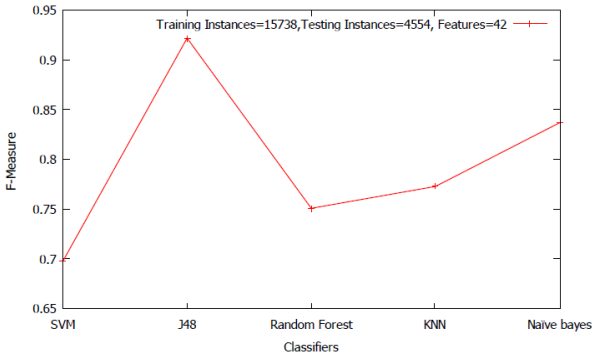

Fig. 12: F-Measure for Probe attack Vs. Classifiers

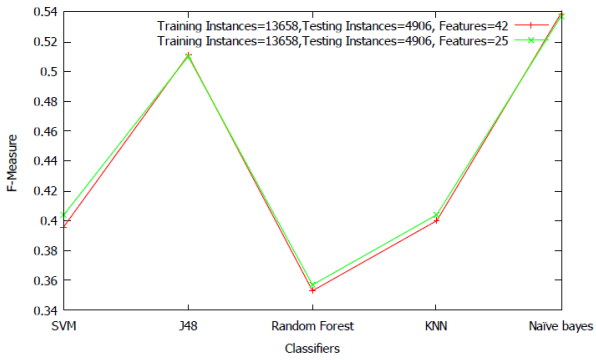

Fig. 13: F-Measure for R2L attack Vs. Classifiers

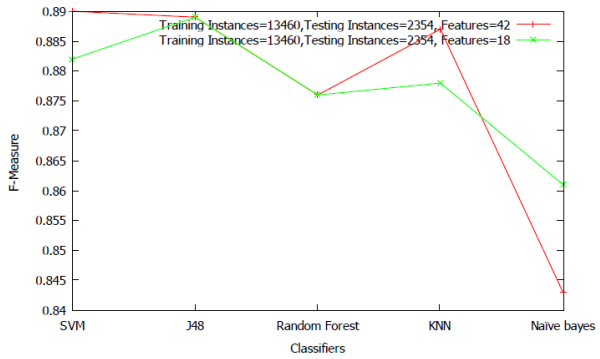

Fig. 14: F-Measure for U2R attack Vs. Classifiers

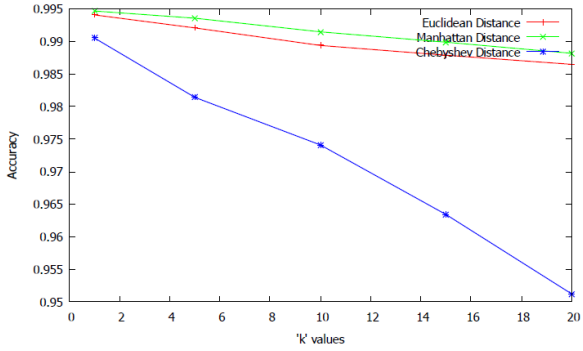

Fig. 15: Accuracy Vs. Total Number of Instances

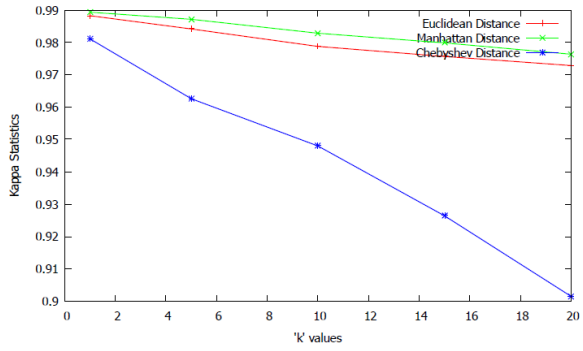

Fig. 16: Kappa statistic for Different Distance Functions

ISSN (Online) 2394-6849

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 7, Issue 9, September 2020

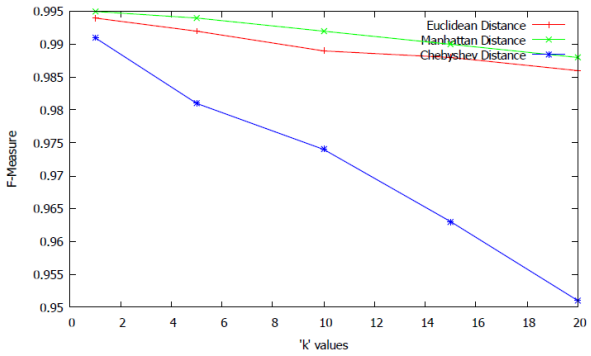

Fig. 17: F-Measure for Different Distance Functions

Chebyshev distance is having the least accuracy as observed. There is rapid decrease in Accuracy, F-Measure and Kappa statics as 'k' value increases.

6 CONCLUSION

In this paper, we have put forth the importance of authentication in achieving the security in VCC. Methods followed for authentication is presented. But most of the had the issue of compromised secret key. The presented framework address the issue by continuously authenticating the user by monitoring the behavior of the user. The KNN classifier detects the presence of malicious node using the identity of authentic user and terminates the services there by reducing the loss due to compromised key. The information required for the data base of the classifier is already present in the KB. So there is no need for separate collection of information. Thus the overhead of storage is reduced. The simulation results show the proposed KNN classifier is the simplest and fastest classifier in comparison with other classifiers which makes the presented framework light weight. Using feature selection there is slight decrease in the accuracy but it does not degrade the performance of the framework to a greater extent. But the number of computations are reduced which increase the speed of the authentication process.

REFERENCES

1. Li, H., Pei, L., Liao, D., Sun, G., & Xu, D. (2019). Blockchain meets VANET: An architecture for identity and location privacy protection in VANET. *Peer-to-Peer Networking and Applications*, *12*(5), 1178-1193.
2. Manvi, S. S., Kakkasageri, M. S., & Pitt, J. (2009). Multiagent based information dissemination in vehicular ad hoc networks. *Mobile information systems*, *5*(4), 363-389.
3. Tchifilionova, V. (2010, March). Security and privacy implications of cloud computing–Lost in the cloud. In *International Workshop on Open Problems in Network Security* (pp. 149-158). Springer, Berlin, Heidelberg.
4. Schwartz, P. M. (2013). EU privacy and the cloud: Consent and jurisdiction under the proposed regulation.
5. Pring, B., Brown, R. H., Frank, A., Hayward, S., & Leong, L. (2009). Forecast: Sizing the cloud; understanding the opportunities in cloud services. *Gartner, Inc., Research Report G*, *166525*.
6. Tahmasebi, M., & Khayyambashi, M. R. (2019). An efficient model for vehicular cloud computing with prioritizing computing resources. *Peer-to-Peer Networking and Applications*, *12*(5), 1466-1475.
7. Olariu, S., Khalil, I., & Abuelela, M. (2011). Taking VANET to the clouds. *International Journal of Pervasive Computing and Communications*.
8. Eltoweissy, M., Olariu, S., & Younis, M. (2010, August). Towards autonomous vehicular clouds. In *International Conference on Ad hoc networks* (pp. 1-16). Springer, Berlin, Heidelberg.
9. Ahmad, F., Kazim, M., & Adnane, A. (2015). Vehicular cloud networks: Architecture and security. In *Guide to Security Assurance for Cloud Computing* (pp. 211-226). Springer, Cham.
10. Goumidi, H., Aliouat, Z., & Harous, S. (2019). Vehicular cloud computing security: A survey. *Arabian Journal for Science and Engineering*, 1-27.
11. You, I., & Li, J. (2016). Special issue on security and privacy techniques in mobile cloud computing.
12. Park, J. J. (2018). Fusion algorithms and high-performance applications for vehicular cloud computing. *The Journal of Supercomputing*, *74*(3), 995-1000.
13. Whaiduzzaman, M., Sookhak, M., Gani, A., & Buyya, R. (2014). A survey on vehicular cloud computing. *Journal of Network and Computer applications*, *40*, 325-344.
14. Xie, Y., Xu, F., Li, D., & Nie, Y. (2018). Efficient message authentication scheme with conditional privacy-preserving and signature aggregation for vehicular cloud network. *Wireless Communications and Mobile Computing*, *2018*.
15. Han, S., Chang, E., Gao, L., & Dillon, T. (2006). Taxonomy of attacks on wireless sensor networks. In *EC2ND 2005* (pp. 97-105). Springer, London.
16. Tripathy, S. (2007, December). LISA: lightweight security algorithm for wireless sensor networks.

In *International Conference on Distributed Computing and Internet Technology* (pp. 129-134). Springer, Berlin, Heidelberg.
17. Bekara, C., Laurent-Maknavicius, M., & Bekara, K. (2008, November). H 2 BSAP: A hop-by-hop Broadcast Source Authentication Protocol for WSN to mitigate DoS attacks. In *2008 11th IEEE Singapore International Conference on Communication Systems* (pp. 1197-1203). IEEE.
18. Alajmi, N. (2014). Wireless sensor networks attacks and solutions. *arXiv preprint arXiv:1407.6290*.
19. Hu, Y. C., Perrig, A., & Johnson, D. B. (2003, March). Packet leashes: a defense against wormhole attacks in wireless networks. In *IEEE INFOCOM 2003. Twenty-second Annual Joint Conference of the IEEE Computer and Communications Societies (IEEE Cat. No. 03CH37428)* (Vol. 3, pp. 1976-1986). IEEE.
20. Yun, J. H., Kim, I. H., Lim, J. H., & Seo, S. W. (2006, December). Wodem: Wormhole attack defense mechanism in wireless sensor networks. In *International Conference on Ubiquitous Convergence Technology* (pp. 200-209). Springer, Berlin, Heidelberg.
21. Suryaprabha, E., & Kumar, N. S. (2019). Enhancement of security using optimized DoS (denial-of-service) detection algorithm for wireless sensor network. *Soft Computing*, 1-11.
22. Armknecht, F., Girao, J., Stoecklin, M., & Westhoff, D. (2006, September). Re-visited: Denial of service resilient access control for wireless sensor networks. In *European Workshop on Security in Ad-hoc and Sensor Networks* (pp. 18-31). Springer, Berlin, Heidelberg.
23. Gill, R. K., & Sachdeva, M. (2018). Detection of hello flood attack on LEACH in wireless sensor networks. In *Next-Generation Networks* (pp. 377-387). Springer, Singapore.
24. Ali, I., Hassan, A., & Li, F. (2019). Authentication and privacy schemes for vehicular ad hoc networks (VANETs): A survey. *Vehicular Communications*, *16*, 45-61.
25. Zhou, Y., Long, X., Chen, L., & Yang, Z. (2019). Conditional privacy-preserving authentication and key agreement scheme for roaming services in VANETs. *Journal of Information Security and Applications*, *47*, 295-301.
26. Prema, N. K. (2019). Efficient secure aggregation in VANETs using fully homomorphic encryption (FHE). *Mobile Networks and Applications*, *24*(2), 434-442.
27. Pournaghi, S. M., Zahednejad, B., Bayat, M., & Farjami, Y. (2018). NECPPA: A novel and efficient conditional privacy-preserving authentication scheme for VANET. *Computer Networks*, *134*, 78-92.
28. Noor, T. H., Zeadally, S., Alfazi, A., & Sheng, Q. Z. (2018). Mobile cloud computing: Challenges and future research directions. *Journal of Network and Computer Applications*, *115*, 70-85.
29. Alizadeh, M., Abolfazli, S., Zamani, M., Baharun, S., & Sakurai, K. (2016). Authentication in mobile cloud computing: A survey. *Journal of Network and Computer Applications*, *61*, 59-80.

REFERENCES

[1] A. Nagpal and G. Gabrani, “Python for Data Analytics , Scientific and Technical Applications,” 1999.
[2] J. Unpingco, “Some Comparative Benchmarks for Linear Algebra Computations in MATLAB and Scientific Python,” pp. 503–505, 2008, doi: 10.1109/DoD.HPCMP.UGC.2008.49.
[3] R. Chudoba, R. Rypl, and M. Vorechovsky, “Using Python for scientific computing : Efficient and flexible evaluation of the statistical characteristics of functions with multivariate random inputs Using Python for scientific computing : efficient and flexible evaluation of the statistical characteristics of functions with multivariate random inputs,” no. February, 2013, doi: 10.1016/j.cpc.2012.08.021.
[4] T. Oliphant and C. Analytics, “Python for Scientific Computing,” no. December, 2014, doi: 10.1109/MCSE.2007.58.
[5] N. Ari and N. M. Mscs, “Symbolic python,” pp. 1–8, 2014.
[6] A. Sheela, “Combination of NumPy , SciPy and Matplotlib / Pylab - a good alternative methodology to MATLAB - A Comparative analysis.”
[7] O. Important *et al.*, “MATLAB vs Python : Why and How to Make the Switch MATLAB vs Python : Comparing Features and Philosophy,” pp. 1–54, 2020.
[8] https://www.analyticsvidhya.com/blog/2020/03/google-colab-machine-learning-deep-learning/