Edge Computing-Based Security Framework for Big Data Analytics in VANETs

Sahil Garg, Amritpal Singh, Kuljeet Kaur, Gagangeet Singh Aujla, Shalini Batra, Neeraj Kumar, and M. S. Obaidat

Abstract

With the exponential growth of technologies such as IoT, edge computing, and 5G, a tremendous amount of structured and unstructured data is being generated from different applications in the smart citiy environment in recent years. Thus, there is a need to develop sophisticated techniques that can efficiently process such huge volumes of data. One of the important components of smart cities, ITS, has led to many applications, including surveillance, infotainment, real-time traffic monitoring, and so on. However, its security, performance, and availability are major concerns facing the research community. The existing solutions, such as cellular networks, RSUs, and mobile cloud computing, are far from perfect because these are highly dependent on centralized architecture and bear the cost of additional infrastructure deployment. Also, the conventional methods of data processing are not capable of handling dynamic and scalable data efficiently. To mitigate these issues, this article proposes an advanced vehicular communication technique where RSUs are proposed to be replaced by edge computing platforms. Then secure V2V and V2E communication is designed using the Quotient filter, a probabilistic data structure. In summary, a smart security framework for VANETs equipped with edge computing nodes and 5G technology has been designed to enhance the capabilities of communication and computation in the modern smart city environment. It has been experimentally demonstrated that use of edge nodes as an intermediate interface between vehicle and cloud reduces access latency and avoids congestion in the backbone network, which allows quick decisions to be made based on the traffic scenario in the geographical location of the vehicles. The proposed scheme outperforms the conventional vehicular models by providing an energy-efficient secure system with minimum delay.

Introduction

Smart cities aim to transform the liveability of people by excelling in multiple key areas: safety, sustainability, and economic growth. Transportation is one key factor that significantly affects the socio-economic development of smart cities [1]. Any sort of inefficiency in this critical network can cause enormous loss of time, depreciation in the level of safety, high pollution, and degradation in the quality of life. The increasing need for mobility has brought about significant changes in transportation infrastructures. Intelligent transportation systems (ITS) have evolved as an innovative and promising solution for next generation transport networks [2]. They contribute to the life and economy of the city and ensure the resilience of the smart city. ITS aim to streamline the operation of vehicles by assisting drivers with important information along with other related needed applications for passengers and road safety. As a prospective ITS technology, vehicular ad hoc networks (VANETs) have recently attracted increasing attention from both the research and industry communities [3].

VANETs are recognized as a significant component of the ITS that creates an intelligent space for vehicular communications. They have no fixed infrastructure and thus rely on smart vehicles to provide them with network functionality. VANETs assume moving vehicles as the nodes and turn every moving vehicle into a wireless router to form a mobile network. Due to their infrastructure-less nature, these networks provide a wide range of applications varying from transit security to driver assistance and Internet access. With the rapid evolution of the latest technologies, like cellular networks and cloud computing, vehicular networks and associated applications have multiplied tremendously. However, the key factors including availability, secure data communication, storage and computation complexity, dynamicity, trust and authentication, and so on pose enormous challenges to these networks [4]. Since vehicular networks ensure safety while driving, traffic efficiency, and great convenience by relaying real-time and important data, any outage of these services can result in life-endangering situations.

To empower VANETs with data handling capabilities, efficient data processing techniques are required that can reduce the computational delay and substantially minimize the cost of data transmission and storage. However, highly dynamic topology, variable network density, and unlimited battery power and storage with vehicular nodes are key participants in time-sensitive communication and computation. To meet these ever increasing demands, cloud-based data processing is an appealing strategy. While data storage and central processing are necessary for some use cases, they may be unreliable for others where critical infrastructures are involved as in ITS, where a minor delay in data processing can lead to hazardous impacts [5]. Additionally, the volume of data will proliferate tremendously as the proportion of con-

Sahil Garg and Kuljeet Kaur are École de Technologie Supérieure; Amritpal Singh, Shalini Batra, and Neeraj Kumar are with the Thapar Institute of Engineering and Technology; Gagangeet Singh Aujla is with Chandigarh University; M. S. Obaidat is with Nazarbayev University, the University of Jordan, and the University of Science and Technology Beijing.

Digital Object Identifier:
10.1109/MNET.2019.1800239

nected vehicles increase and use cases evolve. Hence, the need to minimize latency also emerges subsequently.

To meet the demands of emerging communication applications, a geographically distributed computing architecture is needed in which heterogeneous devices are ubiquitously connected. Hence, edge computing, a new computing paradigm, has been introduced to benefit several relevant domains, including mobile computing, big data analytics, and the Internet of Things (IoT) [6]. It is a highly virtualized platform, which supports different services such as-computation, storage, and networking between the end nodes and traditional clouds. Contrary to the centralized cloud infrastructure, edge computing supports services and applications using extensively distributed deployments. Moreover, edge computing manages to distinguish itself from the rest of the technologies by extending proximity to end users and supporting mobility. Current and upcoming applications that demand edge computing include connected vehicles, autopilot vehicles, wireless sensor networks, smart cities, mobile healthcare systems, and so on [7]. In all such applications, edge computing can provide unified interfaces and flexible resources to accomplish heterogeneous computational and storage requests [5].

Communication in VANETs

In VANETs, it is assumed that each vehicle automatically broadcasts its location, speed, and other useful information at fixed intervals of time within the network. Since the transmission of very critical information is involved, it becomes critical that the relayed messages should travel their optimal path without any additional loss or overhead. The IEEE has proposed IEEE 802.11p, Wireless Access in Vehicular Environments (WAVE), for vehicular communication (5.850–5.925 GHz band) for dedicated short-range communication (DSRC). Defined specifically for VANETs, DSRC makes it possible to transfer the messages without the need to join a basic service set (BSS). It relays messages between the sender and receiver via the network layer. There are two types of DSRC devices that are being used for establishing the communication pathway in VANETs: onboard units (OBUs) and roadside units (RSUs). The former devices are mounted on all vehicles and enable vehicles to communicate with other vehicles and RSUs. Here, RSUs are stationary devices that are mounted along the sides of roads. Due to the dynamic and intermittent connected topology of VANETs, two types of communication generally occur in this architecture: vehicle-to-vehicle (V2V) communication and vehicle-to-roadside (V2R) communication. In V2V communication, a direct link between vehicles is established without relying on a fixed infrastructure for providing safety, security, and dissemination applications. As for V2R communication, the vehicle communicates with the RSUs for the purpose of information processing and data storage.

Security Challenges in VANETs

The VANET is a crucial component of ITS that aims to accomplish productivity through intelligent transportation means. However, issues like multihop connectivity, lack of centrality, infrastructure-less nature, and absence of clear line of defense makes this network unstable. Besides these issues, the information passed on this network is very sensitive and crucial such that any damage or attack on this information can lead to huge disasters to human lives [9]. Examples of the existing security proposals with respect to VANETs are illustrated in Table 1.

In V2V communication, a direct link between vehicles is established without relying on a fixed infrastructure for providing safety, security, and dissemination applications. As for V2R communication, the vehicle communicates with the RSUs for the purpose of information processing and data storage.

There are several attacks that can affect the performance and operation of VANETs. Some possible attacks could cause spreading of bogus information, traffic jams, alteration of the positioning information, disclosure of IDs, and so on. Moreover, other attacks, including replay attacks, masquerading attacks, eavesdropping, wormhole attacks, denial of service (DoS) attacks, impersonation attacks, and so on, can cause catastrophic consequences to this network [3]. Moreover, the development of more and more wireless applications on the very exposed wireless medium can cause significant increase in such attacks. Hence, security of VANETs has emerged as a major concern for both industry and academia, as demonstrated in Fig. 1.

Motivation

With the growing popularity of connected vehicles and the emergence of advanced vehicular applications, the need for large amounts of data to be accessed more quickly, substantially, and locally is growing. Data collected by ITS can be characterized by heterogeneous formats, large volume, and real-time processing requirements. Simple data processing, integration, and analytics tools do not meet the needs of the complex data processing tasks of ITS. Crucial constraints must be taken into account for attaining high reliability and long lifetime of the network. Additionally, the volume of data would proliferate tremendously as the proportion of the connected vehicles increases. Hence, the need to minimize the latency also emerges. It is worth emphasizing that edge computing reduces latency, increases throughput, consolidates resources, saves energy, and enhances security and privacy. Edge computing also reduces the back-and-forth communication between RSUs and the cloud, which can negatively affect VANETs' communication performance, where milliseconds matter [5]. Further to this, big data generated from VANETs requires new security models to support its main dimensions: data volume, variety, velocity, and value. In this direction, many network solutions and overlay networks utilized probabilistic data structures (PDSs) to reduce the processing overhead and enhance security [12]. Therefore, confluence of edge computing and PDSs in fifth generation (5G) networks will better support future VANETs.

Contributions

The key contributions of this article are summarized as follows:

- We propose a smart security framework for VANETs that employs edge computing nodes along with fifth generation (5G) technology instead of RSUs to enhance the achievable capabilities of communication and computation.

Contributors	Focus of the research	Computing model used	Tool(s)	Experimental results
Sedjelmaci *et al.* [9]	An efficient and lightweight intrusion detection system (IDS) was proposed to safeguard the network. The proposed system was implemented to detect three attack types: DoS, integrity target, and false alert generation	A lightweight intrusion detection technique along with a set of rules	NS-3.17 Simulator	Simulation results suggest that the proposed mechanism exhibits high detection rate (DR) (> 97 %), low false positive rate (FPR) (1 percent) along with a low overhead
Zaidi *et al.* [3]	Cooperative information exchange-mechanism-based IDS was proposed for rogue node (RN) detection in VANETs	Anomaly detection along with identification of RNs was envisioned using a traffic model based on statistical techniques	OMNET++, SUMO, and VACaMobil Simulators	Performance of the proposed application-layer IDS improved for highly mobile and dynamic networks such as VANETs in terms of DR, FPR, and overhead
Bouali *et al.* [4]	A preventive mechanism for intrusion detection was proposed to predict the vehicles' behavior on the basis of their expected trustworthiness	A Kalman-filter-based approach was used to predict and classify the behavior of vehicles	NS-3 Simulator	Simulation results suggest that the proposed approach presents a high detection rate, low end-to-end delay, and high delivery ratio
Hou *et al.* [5]	Vehicular fog computing (VFC), a collaborative framework for utilizing the capabilities of numerous end-user clients or near-user edge devices, was proposed to use vehicles as the infrastructures	Edge computing was used for improved usage of individual communication and computational resources of the involved vehicles	Quantitative analysis	VFC achieves better connectivity and computational performance in terms of relaying packets, which leads to more reliable communication
Mehdi *et al.* [8]	A game-theory-based trust model to identify and counter the attacker/malicious nodes in VANETs	A game theoretical approach for VANETs that allows the system to intelligently monitor the network's reliability	NS-2 Simulator	The proposed technique performs better in terms of throughput and data drop rate for different attacker and defender scenarios
Sookhak *et al*. [10]	A novel method for addressing the data sharing problem in the context of VANETs was developed. Additionally, the proposed method also involved intelligent delegation of data to a trusted third party.	Bilinear pairing technique and cloud computing were used as the mainstream platform for utility computing	Simulation modeling	Simulation results show that the proposed model was able to perform the re-encryption process promptly and share the important information among vehicles securely

TABLE 1. Outline of some existing security-related proposals for VANETs.

- We analyze two communication scenarios of VANETs, V2V and vehicle-to-edge (V2E), to detect the attacks occurring in the real-time data streams generated by the network. This has been done by using a PDS-based approach, that is, quotient filter (QF) [13], which is well known for fast query processing and merge operations with robust single hash functions.
- We inspect the intrusion handling capabilities of the proposed model with a dedicated case study supported by a simulated vehicular mobility environment.

Organization

The rest of the article is structured as per the following sequence. The following section illustrates the various aspects related to VANETs in terms of 5G-enabled networking and the potential of edge and cloud computing technologies. Additionally, the section also elaborates on the use of probablistic data structures in big data analytics. The proposed model is then presented with detailed technical description. The results and observations that validate the efficacy of the proposed model are then drawn. Finally, the work is concluded.

Various Aspects Related to VANETs

Several aspects related to VANETs are discussed in the following subsections.

5G-Enabled Vehicle Networking

The rapid increase of mobile services poses challenges to our expectations of security and privacy in wireless networks. Network connectivity is one of the crucial challenges for providing secure information transmission in mobile networks. VANETs are anticipated to deal with ever increasing demand of mobile traffic. However, due to certain limitations like incapable onboard devices, limited spectrum resources, and inefficient system management, they possess several integrity and reachability related issues. Further, the limited transmission capacity of 802.11b protocol also poses significant challenges to this network in terms of network capacity and computing ability. Turning to a 5G cellular network seems to be an enticing prospect for this connectivity related issue in vehicular networks.

A 5G network employs small cell technology to offer superior performance in terms of mobility, reliability, availability, delay, and throughput. Further, the device-to-device (D2D) communication and heterogeneous network technology present in 5G systems greatly improve spectrum efficiency, supporting large-scale streaming data in distant communications. This network provides the following configurations to enable high-data-rate applications: shorter transmission latency (< 1 ms) for moving vehicles, 10 Gb/s peak data rate for low mobility, and 1 Gb/s peak data rate for high mobility, which often vary with

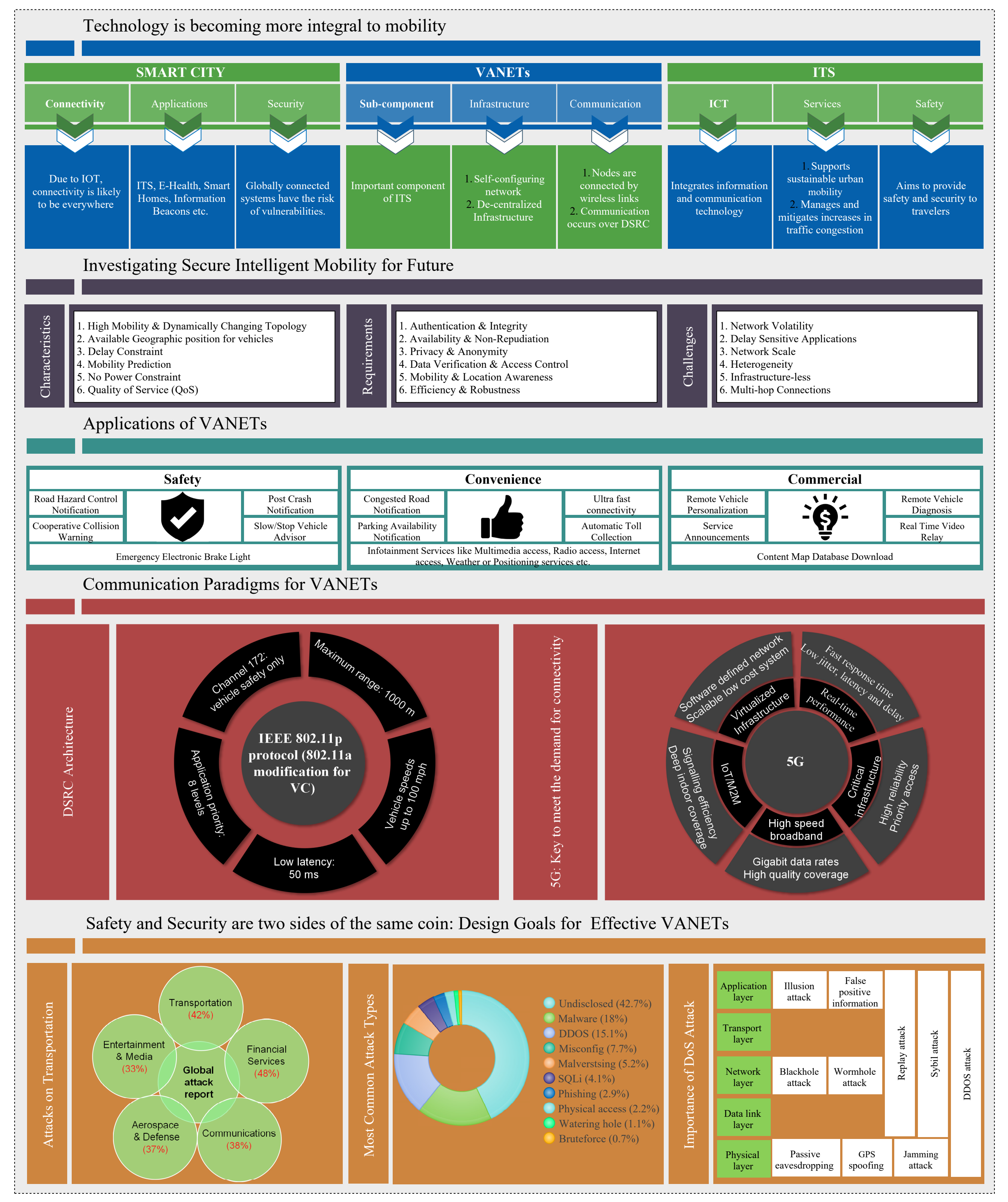

Technology is becoming more integral to mobility
SMART CITY
Connectivity
Applications
Security
VANETs
Sub-component
Infrastructure
Communication
ITS
ICT
Services
Safety
Due to IOT, connectivity is likely to be everywhere
ITS, E-Health, Smart Homes, Information Beacons etc.
Globally connected systems have the risk of vulnerabilities.
Important component of ITS
1. Self-configuring network
2. De-centralized Infrastructure
1. Nodes are connected by wireless links
2. Communication occurs over DSRC
Integrates information and communication technology
1. Supports sustainable urban mobility
2. Manages and mitigates increases in traffic congestion
Aims to provide safety and security to travelers
Investigating Secure Intelligent Mobility for Future
Characteristics
1. High Mobility & Dynamically Changing Topology
2. Available Geographic position for vehicles
3. Delay Constraint
4. Mobility Prediction
5. No Power Constraint
6. Quality of Service (QoS)
Requirements
1. Authentication & Integrity
2. Availability & Non-Repudiation
3. Privacy & Anonymity
4. Data Verification & Access Control
5. Mobility & Location Awareness
6. Efficiency & Robustness
Challenges
1. Network Volatility
2. Delay Sensitive Applications
3. Network Scale
4. Heterogeneity
5. Infrastructure-less
6. Multi-hop Connections
Applications of VANETs
Safety
Road Hazard Control Notification
Post Crash Notification
Cooperative Collision Warning
Slow/Stop Vehicle Advisor
Emergency Electronic Brake Light
Convenience
Congested Road Notification
Ultra fast connectivity
Parking Availability Notification
Automatic Toll Collection
Infotainment Services like Multimedia access, Radio access, Internet access, Weather or Positioning services etc.
Commercial
Remote Vehicle Personalization
Remote Vehicle Diagnosis
Service Announcements
Real Time Video Relay
Content Map Database Download
Communication Paradigms for VANETs
DSRC Architecture
IEEE 802.11p protocol (802.11a modification for VC)
Channel 172: vehicle safety only
Maximum range: 1000 m
Vehicle speeds up to 100 mph
Low latency: 50 ms
Application priority: 8 levels
5G: Key to meet the demand for connectivity
5G
Software defined network
Scalable low cost system
Fast response time
Low jitter, latency and delay
Virtualized Infrastructure
Real-time performance
Critical infrastructure
High reliability
Priority access
High speed broadband
Gigabit data rates
High quality coverage
IoT/M2M
Signalling efficiency
Deep indoor coverage
Safety and Security are two sides of the same coin: Design Goals for Effective VANETs
Attacks on Transportation
Transportation (42%)
Entertainment & Media (33%)
Financial Services (48%)
Global attack report
Aerospace & Defense (37%)
Communications (38%)
Most Common Attack Types
Undisclosed (42.7%)
Malware (18%)
DDOS (15.1%)
Misconfig (7.7%)
Malverstsing (5.2%)
SQLi (4.1%)
Phishing (2.9%)
Physical access (2.2%)
Watering hole (1.1%)
Bruteforce (0.7%)
Importance of DoS Attack
Application layer
Illusion attack
False positive information
Transport layer
Network layer
Blackhole attack
Wormhole attack
Data link layer
Physical layer
Passive eavesdropping
GPS spoofing
Jamming attack
Replay attack
Sybil attack
DDOS attack

FIGURE 1. The need for security in VANETs [11].

vehicle density and vehicle average speed [14]. This recent advancement in mobile communications also allows different deployment architectures of VANETs to support applications with different requirements. Compared to the centralized cloud, the 5G communication environment also enables vehicles to access base stations and communicate with edge nodes.

Cloud Computing and Edge Computing

Cloud computing is defined as an Internet-based computing paradigm that uses a network of high-end servers to deliver remote access computing resources to intended users. With abundant storage and ample computing power, the cloud has

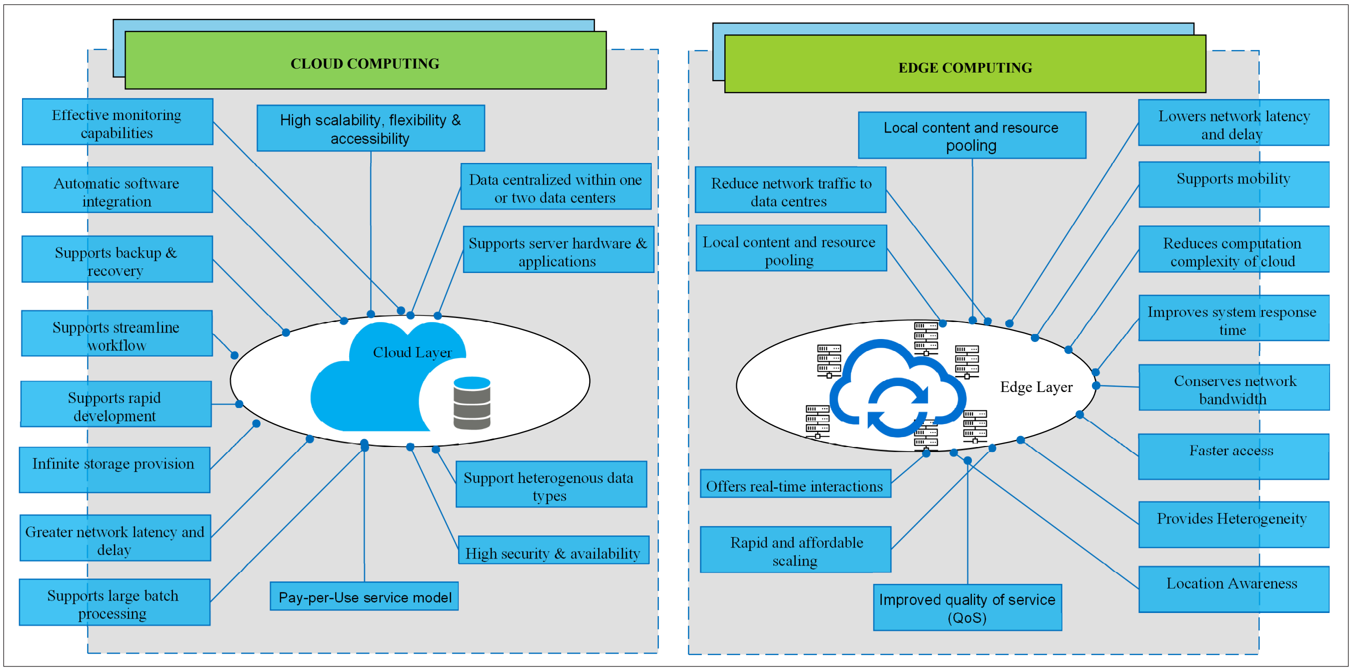

CLOUD COMPUTING
Effective monitoring capabilities
High scalability, flexibility & accessibility
Automatic software integration
Data centralized within one or two data centers
Supports backup & recovery
Supports server hardware & applications
Supports streamline workflow
Cloud Layer
Supports rapid development
Infinite storage provision
Support heterogenous data types
Greater network latency and delay
High security & availability
Supports large batch processing
Pay-per-Use service model
EDGE COMPUTING
Local content and resource pooling
Lowers network latency and delay
Reduce network traffic to data centres
Supports mobility
Local content and resource pooling
Reduces computation complexity of cloud
Improves system response time
Edge Layer
Conserves network bandwidth
Faster access
Offers real-time interactions
Provides Heterogeneity
Rapid and affordable scaling
Location Awareness
Improved quality of service (QoS)

FIGURE 2. Comparative analysis of cloud computing and edge computing.

become an affordable platform for big data analytics. It provides several features including load balancing, data replication, failure recovery, and so on. Cloud computing has been the outstanding service supplier where delay-tolerant applications can be handled with ease. The only problem with cloud computing is bandwidth and latency if data is to be transmitted wirelessly. For example, the immense amount of data generated from vehicles would be costly and time-consuming to be sent to the cloud for processing and analysis. Huge latency is caused by this round-trip movement of data between vehicle and cloud-based platforms, which cannot be afforded in a critical infrastructure like ITS. Since ITS operates based on movements, any delay in information processing can lead to collisions and accidents [11]. Thus, the solution demands instant processing of data streams with quick turnaround.

Edge computing is a distributed infrastructure in which applications are managed at the edge of the network, thus reducing the amount of data transported to the cloud [5]. It is a middle layer between the cloud and the hardware interface where constraints of energy and resources can be relaxed by geographically distributed edge nodes. In the edge environment, intelligence is provided to a local area network to enable efficient data processing, analysis, and storage, which is strongly required when dealing with sensitive data. The decentralization of cloud infrastructure to the edge delivers several benefits to the transportation system: user privacy, speedy computation, prompt handling of latency-sensitive messages, easy deployment of edge services on 5G base stations, and pooling of resources at different layers, among others. This is the reason that big hardware giants like Cisco, Intel, and Dell are also operating to construct gateways and routers that can assist fogging. Some of the relative differences between computational methodology of cloud computing and edge computing are highlighted in Fig. 2.

Probabilistic Data Structures in Big Data Analytics

Statistical analysis and data mining of gigantic datasets involving multiple terabytes of data have become a trivial task in the modern era of technology. These complex operations on data are extensively witnessed in different domains like web analytics and Internet advertising. When the datasets with which an application is dealing becomes very large, deterministic data structures are not feasible because the data is too big to fit in the memory. It becomes even more difficult for streaming applications, which typically require data to be processed in one pass and perform incremental updates. Probabilistic alternatives to deterministic data structures are better in terms of simplicity and constant factors involved in actual runtime as traditional methods might show accurate results, but the time and space trade-off is not acceptable to big data frameworks.

PDSs are extremely handy data structures that reduce the time and space trade-off to a great extent corresponding to storage and retrieval and querying of data [12]. They utilize the different probability-based approaches along with approximation principles and hashing methods. They are suitable for large data processing, approximate predications, fast retrieval, and storing unstructured data, thus playing an important role in big data processing. In particular, these data structures use hash functions to randomize and compactly represent a set of items, and some probability-based approaches to reduce time or space trade-offs. Compared to error-free approaches, these algorithms use much less memory and have constant query time. Moreover, they usually support union and intersection operations, and therefore can be parallelized easily. Some important PDSs include Bloom filters and quotient filters for membership query of massive datasets, count-min sketch for counting the number of times a data item has arrived in the huge datasets, and hyper-log-log for cardinality estimates [15].

Proposed Model

This article proposes a novel technique for secure communication in VANETs. Security attacks can create unpredicted chaos in a V2V network and hamper all vehicular movements, especially in areas that experience heavy rush hours. Attackers specifically target vehicles to breach security and disrupt services. The proposed technique uses a quotient-filter-based authentication mechanism to solve two crucial issues related to this critical infrastructure: whether any unauthorized node has entered the network and whether any attack has been initiated in the existing network.

As depicted in Fig. 3a, an attacker can choose any attack vector from the database to disrupt the services of the network. Let us consider one scenario where an attacker tries to enter the network by using a fake identity or attack the network from outside. To provide protection against such types of mischievous activities, a few novel technical upgrades proposed in the scheme are:

- Edge computing platforms for replacing RSUs, that is, vehicle-to-infrastructure (V2I) refers to the vehicle-to-edge (V2E) computing platform instead of V2R communication
- 5G network for *V2V* and *V2E*-based fast communication
- Use of quotient filter (*QF*) between vehicle to vehicle ($V_i \rightarrow V_j$) and vehicle to edge node ($V_i \rightarrow f_j$) communication to identify an attacked node quickly

Consider a network ($\aleph_V$) having *n* vehicles, that is, ($\aleph_V = \{V_1, V_2, V_3,..., V_n\}$), communicating through the set of *m* edge nodes represented as $f' = \{f_1, f2, ..., f_m\}$. The optimization objective of the proposed framework, with vehicles, edge nodes, and cloud as its components, is to maximize the computational power (C_p) for better decision making, minimize delay (D_e) at all the communication levels, enhance the cryptography-based security ($Ⓢ^C$) mechanism to improve integrity of data, and provide an energy-efficient (*En*) communication framework. Mathematically, the objective function (Ω^O) can be defined as

$$\begin{matrix} \max_{Ⓢ C, C_p} \\ \min_{D_e, En} \end{matrix} \left(\Omega^O\right) \left[V_i \Rightarrow \begin{cases} V_j & \text{(V2V)} \\ f_k & \text{(V2E)} \end{cases} \right. \quad (1)$$

subject to → identification of an infected node. To make the said framework resilient to attacks, a QF-based security mechanism is used at both the vehicle and edge levels. QF is a probabilistic data structure used for membership query from the stored dataset.

Quotient Filter

The QF is a space-efficient and cache-friendly probabilistic data structure that uses the quotienting technique of hashing [12] to store a set $S \subset U$ efficiently. Mapping is done for every element $x \in S$ to $h(x)$, where $h(x)$ is a primary hash function resulting in a set of *p* bits named as the fingerprint of *x*, that is, $h(x) \mapsto \{0, ...,2^p - 1\} \Rightarrow fp(x)$.

$fp(x)$ is an open hash table with $m = 2^q$ buckets where each bucket has $(r + 3)$ bits. In $fp(x)$, the least significant bits are denoted by *r*, and the most significant bits in quotienting are represented by $q = (p - r)$. Insertion operation in the QF is performed by computing the quotient $f_q \leftarrow (\lfloor fp(x)/2^r \rfloor)$ and remainder $f_r \leftarrow (fp(x) \bmod 2^r)$ of every considered element. Here, the index of the bucket used for inserting an element is denoted by f_q, whereas f_r represents a value inserted in the bucket f_q.

Two important terms used for identifying the appropriate positions for insertion and querying in QF are *run* and *cluster*. *Run* refers to a scenario where remainders of different fingerprints having the same f_q are stored contiguously, that is, f_q of two items collide but f_r are distinct. Such collisions are resolved through linear probing. In such scenarios, remainders associated with different f_q are shifted and corresponding meta-data bits are updated for each bucket if required. A *cluster* is a sequence of one or more consecutive runs with no empty bucket between them. A cluster is immediately followed by an empty slot.

A general observation is that a Bloom filter (BF) has more hashing functions as compared to a QF. Since hash functions are generated only once per QF and every signature is implemented using one location, single-memory access is required to check the presence of a signature. Thus, in comparison to BF, QF has higher throughput.

Vehicle-to-Vehicle Communication

Every vehicle $V_i \in \aleph_V$ can send or receive data from any other vehicle V_j in the network. To ensure the security of the communication, each vehicle is equipped with the table of all authenticated nodes registered under the edge (f_k) node. Further, a public key ($Pu^{f_k}_{key}$) is shared with all vehicles belonging to the same edge node to provide a second level of security. Each vehicle, that is, $V_j \in \aleph_V$, is assigned an edge node as it enters its defined range.

Whenever the vehicle enters the edge node's range, it upgrades its table, which includes the list of all the nodes registered with the edge node at that time. The following steps are performed before starting V2V communication as depicted in Fig. 3b.

Step 1: Each vehicle maintains a quotient filter (QF_v) of all registered vehicles to the edge using $vehicle_{id}$ and a public key provided by the edge node, that is, $(QF_v) \leftarrow (V_i \oplus Pu^{f_k}_{key})$.

Step 2: Whenever a vehicle V_i tries to communicate with V_j, a query is performed on the QF, that is, $QF_{V_j} \leftarrow QueryV_i$, before initiating any communication (sending and receiving the packets).

Step 3: If the query in Step 2 returns TRUE, that is, $(V_i \in f_k)$, V_j sends an acknowledgment (*Ack*) to V_i, and data sharing starts.

Step 4: If the query in Step 2 returns FALSE, that is, $(V_i \neq f_i)$, V_j sends a wait signal to V_i and edge level query is performed as follows:

- V_j immediately contacts the edge node f_k to check whether any new vehicle has been registered (V_i). If the edge node returns true, V_j updates its table and starts communication.
- If even the edge is unable to locate V_i, it is marked as an intruder in the network, and an alert is issued indicating that a fake user or node has entered the network, and the id (V_i) along with the packet are passed on to the edge node so that it can send an alert to all the nodes in its range to not receive any message from V_i.

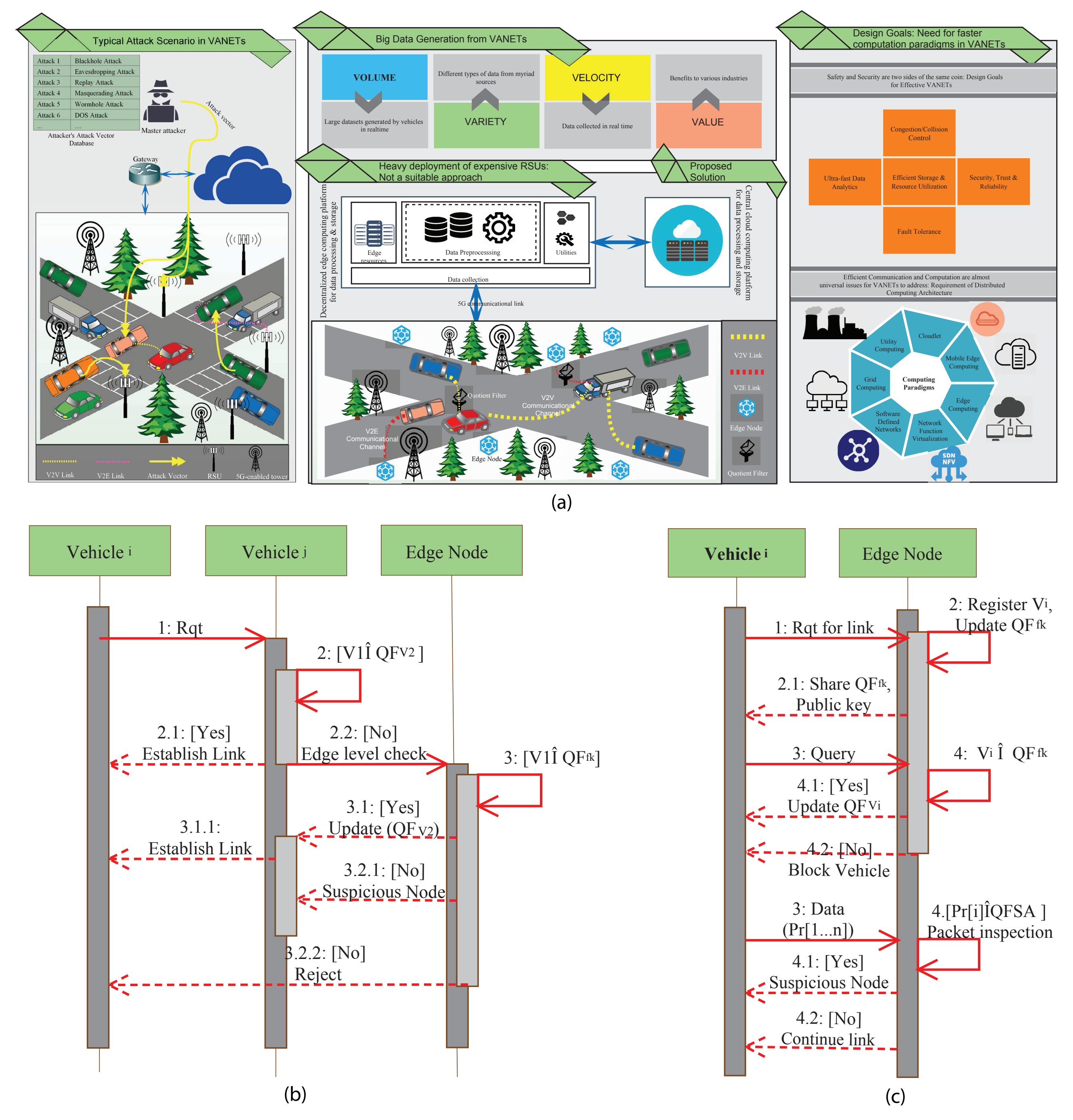

Typical Attack Scenario in VANETs
Attack 1 Blackhole Attack
Attack 2 Eavesdropping Attack
Attack 3 Replay Attack
Attack 4 Masquerading Attack
Attack 5 Wormhole Attack
Attack 6 DOS Attack
Attacker's Attack Vector Database
Master attacker
Attack vector
Gateway
V2V Link
V2E Link
Attack Vector
RSU
5G-enabled tower
Big Data Generation from VANETs
VOLUME
Different types of data from myriad sources
VELOCITY
Benefits to various industries
Large datasets generated by vehicles in realtime
VARIETY
Data collected in real time
VALUE
Heavy deployment of expensive RSUs: Not a suitable approach
Proposed Solution
Decentralized edge computing platform for data processing & storage
Edge resources
Data Preprocesssing
Utilities
Data collection
Central cloud computing platform for data processing and storage
5G communicational link
Quotient Filter
V2V Communicational Channel
V2E Communicational Channel
Edge Node
V2V Link
V2E Link
Edge Node
Quotient Filter
Design Goals: Need for faster computation paradigms in VANETs
Safety and Security are two sides of the same coin: Design Goals for Effective VANETs
Congestion/Collision Control
Ultra-fast Data Analytics
Efficient Storage & Resource Utilization
Security, Trust & Reliability
Fault Tolerance
Efficient Communication and Computation are almost universal issues for VANETs to address: Requirement of Distributed Computing Architecture
Utility Computing
Cloudlet
Mobile Edge Computing
Grid Computing
Computing Paradigms
Edge Computing
Software Defined Networks
Network Function Virtualization
SDN NFV
(a)
Vehicle i
Vehicle j
Edge Node
1: Rqt
2: [V1Î QFV2]
2.1: [Yes] Establish Link
2.2: [No] Edge level check
3: [V1Î QFfk]
3.1: [Yes] Update (QFV2)
3.1.1: Establish Link
3.2.1: [No] Suspicious Node
3.2.2: [No] Reject
(b)
Vehicle i
Edge Node
1: Rqt for link
2: Register Vi, Update QF fk
2.1: Share QFfk, Public key
3: Query
4: Vi Î QF fk
4.1: [Yes] Update QF Vi
4.2: [No] Block Vehicle
3: Data (Pr[1...n])
4.[Pr[i]ÎQFSA] Packet inspection
4.1: [Yes] Suspicious Node
4.2: [No] Continue link
(c)

FIGURE 3. Current and proposed communication scenarios in VANETs: a) the need for a faster and secure communication paradigm in VANETs; b) V2V communication scenario; c) V2E communication scenario.

V2V communication is summarized in Eq. 2 as

$$\begin{matrix} QF_{V_i} \\ Pu_{key}^{f_k} \end{matrix}\,[V_i]\xrightleftharpoons[(1.Req),(2.(Data_{V_i}\oplus Pu_{key}^{f_k}))]{(1.Ack/Wait),(2.(Data_{V_j}\oplus Pu_{key}^{f_k}))}[V_j]^{QFV_j}_{Pu_{key}^{f_k}} \qquad (2)$$

Vehicle-to-Edge (V2E) Communication

In the proposed framework, RSUs are replaced with edge nodes for intermediate communication since edge nodes have much greater processing power than RSUs. Each edge node f_k contains the following information for reliable communication:

- The list of registered nodes ($\aleph_{f_k}$) within range of f_k is stored in QF using quotienting technique (H_q), that is, $QF_{f_k} \leftarrow \forall(V_i \in \aleph_{f_k})H_q(V_i \oplus f_k)$.
- The signature of already traced attacks in data packets is maintained in QF (QF_{SA}) to inspect data coming from a vehicle.
- The public key ($Pu_{key}^{f_k}$) is maintained for secure V2V communication, which is shared with all registered nodes under edge node f_k.
- The private key ($Pr_{key}^{f_k}$) is kept with the edge node for secure V2E communication.

In the proposed framework, the following steps are required to start V2E communication, as shown in Fig. 3c:

- Whenever the vehicle ($V_n \in \aleph_V$) enters the edge node's (f_k) range, the following steps are performed:
  - The edge node (f_k) upgrades its table by adding a new entry in the QF (QF_{fk}), thereby registering the incoming vehicle as an authentic node, that is, $QF_{fk} \leftarrow Hq(V_n \oplus f_k)$.
  - Share the public key (Pu^{fk}_{key}) and update the vehicle database for registered nodes, that is, ($QF_{V_n} \leftarrow QF_{fk}$)
- Whenever any vehicle V_i tries to communicate with f_k the following steps are performed:
  - Before starting communication, the edge node queries its table, and if the id exists, that is, $V_i \in \aleph_{fk}$, it establishes the communication.
  - Otherwise, V_i is marked as an intruder, and all registered authentic nodes are informed about it.
- To detect an infected registered node, that is, ($V_i \in \aleph_{fk}$) used by an attacker for communication, inspection of random packets (PI) coming from vehicles to the edge is done as follows:
  - Select packet (P_r) from data coming from V2E communication, that is, $P_r \leftarrow Random(\exists(V_i \in \aleph_{fk})(Data(V_i))$.
  - Check P_r against the signature of stored attacks, that is, $\kappa \leftarrow (QF_{SA} = Query(P_r))$. If κ returns false, it means that reliable data has been sent, and V_i is not affected by an attacker.
  - If κ returns true, V_i is marked as suspect (infected by an attacker), and packet inspection of V_i is performed more frequently.
  - If infected messages from suspicious vehicle V_i, which is communicating with f_k, are recurrent, it is added onto the suspicious list. The moment the threshold of the messages is crossed, the edge node sends an alert to all the nodes registered to it, that is, $(\forall j(V_j \in \aleph_{fk}) \mid j \neq i)$, to avoid receiving any message from V_i as it realizes that the entire network is under attack.

From the analysis it is clear that the frequency of packet inspection in the considered security model is dependent on the underlying user application services. For instance, for highly crucial and secure applications, the packet inspection is done for every incoming packet/flow. On the other hand, for less delay-sensitive crucial applications, a user can opt for the random walk approach, wherein a packet is randomly selected for packet examination to ensure that it is not tampered by the intruder. V2E communication is demonstrated in Eq. 3, where $E_D = Data_{V_j} \oplus Pu^{fk}_{key}$.

Since edge nodes are quite close to the vehicle compared to cloud nodes, latency is reduced and congestion in the backbone network is avoided, allowing decision making based on the traffic scenario in the geographical location of the vehicles. These nodes immediately send alerts to all the nodes (vehicles) within their range through the 5G network as it enable devices to communicate directly with other devices in proximity through a direct local link. The major reason for the use of QFs in V2V and V2E in this scheme is that they can identify the authentication of communicating nodes by verifying the packets coming from the vehicles.

$$\overset{QF_{V_i}}{\underset{Pu^{f_k}_{key}}{}}[V_i] \xrightleftharpoons[(1.Req),(2.E_D)]{(1.Ack/Rjt),(2.QF_{f_k}),(3.(E_D)),(4.P_I)} [f_k]^{(QF_{f_k}),(Pr^{fk}_{key})}_{Pu^{fk}_{key},(QF_{SA})} \tag{3}$$

Results and Analysis

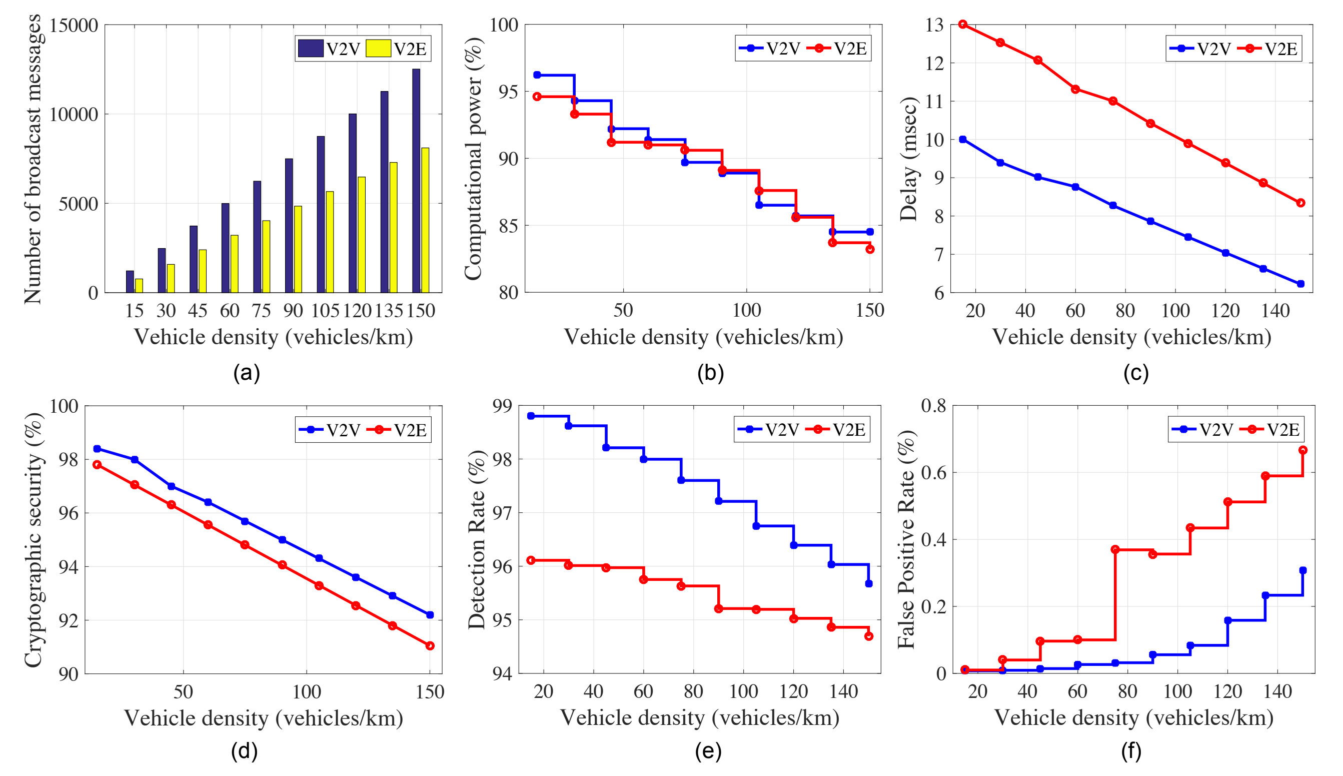
The proposed approach aims to improve the surveillance and security of vehicles in VANETs using edge nodes as they provide more computational power for data processing and decision making. The proposed framework uses a QFr-based security mechanism in both V2V and V2E communication. Intruder detection is done at the V2V level, and for advanced security check, deep packet inspection is done by the edge nodes. In this section, the performance analysis of the proposed framework is provided. To authenticate the security features of the proposed model at various levels, analysis has been performed through network simulations. All the experiments have been performed on an *i7-3612QM* CPU @ 2.10 GHz with 8 GB of RAM. For simulation of the proposed framework, we have used MATLAB. During simulations, a three-layer setup is designed where the first layer indicates the vehicles in the network, the second layer consists of edge nodes, and in the third layer cloud storage is mentioned as the final data center. The obtained results have been averaged over 15 simulation runs, wherein the density of the participating vehicles has been taken in the range of 0 vehicles/km to 150 vehicles/km. Here, the vehicular density in the range of 110–150 vehicles/km denotes the heavy traffic scenario. On the other hand, the vehicular density in the range of 0–60 vehicles/km represent the non-traffic scenario, while the rest of the cases depict the moderate traffic scenario. This variability in the vehicular traffic has been incorporated to extensively evaluate the scalability of the proposed scheme in the vehicular setup.

In our simulations, we consider two kinds of attacks. The first is a vehicle not registered with the edge node, which tries to communicate with other registered vehicles of the network, and the second is when the attacker tries to use a registered vehicle as an intermediary medium to disrupt the network. These attacks are simulated, and results are computed for the network, which show high detection rate in a short time using very little computational resources.

For the purpose of relative comparison, the vehicular density on roads (number of vehicles per kilometer) has been considered for the evaluation of different metrics. Their detailed description is as follows. In Fig. 4a, the number of broadcast messages with the increase in vehicular density is depicted. As is clearly evident from the figure, the broadcast messages increase more rapidly in V2V communications than in V2E communication.

Figure 4b shows the computational power required for querying the QF based security framework. In V2V communication the commutation power decreases slowly in comparison with V2E communication. This is due to the fact that the number of edge nodes is fixed in V2E communications and the computational power decreases a bit faster with the increasing vehicular load.

End-to-end delay in message passing vs. vehicle density between V2V and V2E is depicted in Fig. 4c. As the number of vehicles increases,



(a) (b) (c)

(d) (e) (f)

FIGURE 4. Performance evaluation of the proposed scheme: a) broadcast messages with respect to vehicle density; b) computational power with respect to vehicle density; c) delay in packet communication with respect to vehicle density; d) cryptographic security with respect to vehicle density; e) intrusion detection rate at different levels; f) false positive rate with respect to vehicle density.

communication between vehicles becomes fast; delay in V2E communication is significantly affected because of the fixed number of edge nodes in the network.

The performance evaluation of the proposed scheme in terms of cryptographic security vs. vehicle density is depicted in Fig 4d. As is evident from the figure, the proposed scheme offers good performance across the V2V and V2E communications. This can be attributed to the use of the proposed QF-based security mechanism at both the vehicle and edge levels.

Figure 4e depicts the accuracy in identifying attacks, that is, detection rate vs. vehicle density. To evaluate the accuracy of the proposed approach, a number of security attacks have been simulated at different layers. A vulnerable environment of around 100 attacks is considered to study the effect of the proposed security framework. A network having 15 edge nodes shows a significant detection rate as the number of vehicles in the network is increased.

Figure 4f evaluates the performance of QF vs. vehicle density used in the proposed work (i.e., false positives while querying for an element). Very low false positives are observed as the number of components in the network increases.

Comparison with an Existing Scheme

In order to evaluate the effectiveness of the proposed scheme based on QF, it has been compared to an approach based on BF, hash table (HT), and B+ tree (B+). All schemes possess the same capabilities to secure the vehicular infrastructure except the use of underlying PDS, that is, QF against the BF, HT, and B+. The related results in the considered vehicular network are depicted in Fig. 5, wherein the delay parameter has been considered for comparison. As evident from the figure, the proposed scheme has comparatively less delay relative to the existing schemes. An overall improvement of 19.93, 25.75, and 33.50 percent has been observed, which can be attributed to the following intrinsic properties of QF. It is an improved version of the BF wherein the metadata bits per buckets lead to faster querying results in comparison to the existing counterparts. Also, the QF depicts improved results in terms of merging operations relative to others.

Conclusion

With the emergence of ever growing advanced vehicular applications, the challenges to meet the increased data communication and computational capabilities are increasing. The latest applications developed for better safety, security, and traffic efficiency demand more sophisticated techniques to accommodate these ever changing requirements of complex data processing. Every day, new vulnerabilities are discovered, more breaches are reported, and the network becomes less secure. Although mobile cloud computing facilitates high quality of network connections with remote infrastructures, it increases the network latency and congestion. Further, RSUs widen the network communication capability, but are quite expensive and difficult to deploy along roads, especially on a large scale, such as over a whole city. To meet the desired demands, this article proposes the use of the edge data platform for quick and efficient communication and computation along with quotient filter for providing efficient storage and security. Experimental results demonstrate that the proposed model outperforms the conventional vehicular models by providing an energy-efficient secure system with minimum delay.

References

[1] *Smart Cities and Homes: Key Enabling Technologies*, M. S. Obaidat and P. Nicopolitidis, Eds., Morgan Kaufmann, May 2016.

[2] "Technology and Computing Requirements for Self-Driving Cars," Intel, 2016; http://les-svc.org/wp-content/uploads/2015/06/2016-05-18-Intel-automotive-autonomous-driving-vision-paper.pdf, accessed May 2017

[3] K. Zaidi *et al.*, "Host-Based Intrusion Detection for VANETs: A Statistical Approach to Rogue Node Detection," *IEEE Trans. Vehic. Tech.*, vol. 65, no. 8, 2016, pp. 6703–6714.

[4] T. Bouali, S.-M. Senouci, and H. Sedjelmaci, "A Distributed Detection and Prevention Scheme from Malicious Nodes in Vehicular Networks," *Int'l. J. Commun. Systems*, vol. 29, no. 10, 2016, pp. 1683–1704.

[5] X. Hou *et al.*, "Vehicular Fog Computing: A Viewpoint of Vehicles as the Infrastructures," *IEEE Trans. Vehic.r Tech.*, vol. 65, no. 6, 2016, pp. 3860–73.

[6] K. Kaur *et al.*, "Edge Computing in the Industrial Internet of Things Environment: Software-Defined-Networks-Based Edge-Cloud Interplay," *IEEE Commun. Mag.*, vol. 56, no. 2, Feb. 2018, pp. 44–51.

[7] S. Garg *et al.*, "UAVEmpowered Edge Computing Environment for Cyber-Threat Detection in Smart Vehicles," *IEEE Network*, vol. 32, no. 3, May/June 2018, pp. 42–51.

[8] M. M. Mehdi, I. Raza, and S. A. Hussain, "A Game Theory Based Trust Model for Vehicular Ad hoc Networks (VANETs)," *Computer Networks*, vol. 121, 2017, pp. 152–72.

[9] H. Sedjelmaci, S. M. Senouci, and M. A. Abu-Rgheff, "An Efficient and Lightweight Intrusion Detection Mechanism for Service- Oriented Vehicular Networks," *IEEE Internet of Things J.*, vol. 1, no. 6, 2014, pp. 570–77.

[10] M. Sookhak, F. R. Yu, and H. Tang, "Secure Data Sharing for Vehicular Ad-hoc Networks Using Cloud Computing," *Ad Hoc Networks*, Springer, 2017, pp. 306–15.

[11] IBM X-Force Threat Intelligence, IBM Security, Mar. 2016; http://www. foerderland.de/fileadmin/pdf/IBM XForce Report 2016.pdf, accessed Aug. 2017.

[12] S. Dutta, A. Narang, and S. K. Bera, "Streaming Quotient Filter: A Near Optimal Approximate Duplicate Detection Approach for Data Streams," *Proc. VLDB Endowment*, vol. 6, no. 8, 2013, pp. 589–600.

[13] M. Al-hisnawi and M. Ahmadi, "Deep Packet Inspection Using Quotient Filter," *IEEE Commun. Letters*, vol. 20, no. 11, 2016, pp. 2217–20.

[14] R. Yu *et al.*, "Optimal Resource Sharing in 5G-Enabled Vehicular Networks: A Matrix Game Approach," *IEEE Trans. Vehic. Tech.*, vol. 65, no. 10, 2016, pp. 7844–56.

[15] A. Singh *et al.*, "Probabilistic Data Structure-Based Community Detection and Storage Scheme in Online Social Networks," *Future Generation Computer Systems*, vol. 94, 2019, pp. 173–84.

Biographies

Sahil Garg [S'15, M'18] (sahil.garg@ieee.org) received his Ph.D. degree from Thapar Institute of Engineering and Technology, Patiala, India, in 2018. He is currently working as a postdoctoral research fellow at École de Technologie Supérieure, Université du Québec, Montréal, Canada. He has many research contributions in the area of machine learning, big data analytics, cloud computing, and vehicular ad hoc networks. Some of his research findings are published in top cited journals such as *IEEE TII*, *IEEE TMM*, the *IEEE Internet of Things Journal*, *IEEE Communications Magazine*, *IEEE Network*, *IEEE CE Magazine*, *FGCS*, and *Information Sciences*. He also received the IEEE ICC Best Paper Award in 2018.

Amritpal Singh [S'17] (amritpal.singh203@gmail.com) received his M.E. degree and Ph.D. degree from Thapar Institute of Engineering and Technology, Punjab, India, in 2013 and 2018, respectively. He is currently working as a lecturer in the Computer Science and Engineering Department, Thapar Institute of Engineering and Technology. His research interests include probabilistic data structures, machine learning, and big data.

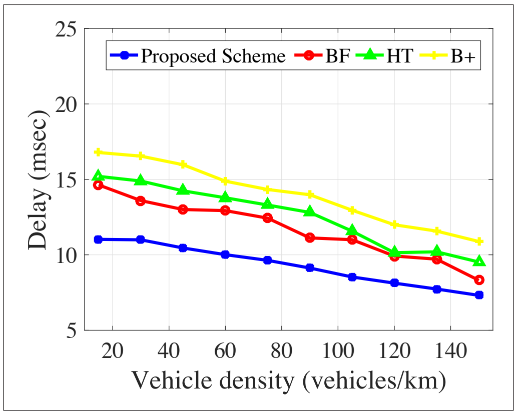


FIGURE 5. Comparative delay evaluation of the proposed scheme with an existing scheme.

Kuljeet Kaur [S'13, M'18] (kuljeet.kaur@ieee.org) received her Ph.D. degree from Thapar Institute of Engineering and Technology, Patiala, in 2018. She is currently working as a postdoctoral research fellow at École de Technologie Supérieure, Université du Québec. She has many research contributions in the areas of cloud computing, energy efficiency, smart grid, frequency support, and vehicle-to-grid. Some of her research findings are published in top cited journals inlcuding *IEEE TII*, *IEEE TMM*, *IEEE TVT*, *IEEE TSG*, *IEEE TPDS*, *IEEE Communications Magazine*, *IEEE Wireless Communications*, *IEEE PS*, *Springer PPNA*, and so on. She received the IEEE ICC Best Paper Award in 2018.

Gagangeet Singh Aujla [S'15, M'18] (gagi_aujla82@yahoo.com) received his B.Tech and M.Tech degrees from Punjab Technical University, Jalandhar, in 2003 and 2013, respectively, and his Ph.D. from Thapar Institute of Engineering and Technology in 2018, all in computer science and engineering. He received the 2018 *IEEE TCSC* Outstanding Ph.D Dissertation Award in 2018 at Guangzhou, China. He has many research contributions in the areas of smart grid, cloud computing, and vehicular ad hoc networks. Some of his research findings are published in top cited journals such as *IEEE TII*, *IEEE TCC*, *IEEE Communications Magazine*, *IEEE CE Magazine*, *FGCS*, and *JPDC*.

Shalini Batra [M'17] (sbatra@thapar.edu) received her Ph.D. degree in computer science and engineering from Thapar Institute of Engineering and Technology, Patiala, in 2012. She is currently working as an associate professor with the Department of Computer Science and Engineering, Thapar University, Patiala. She has guided many research scholars leading to Ph.D. degrees and M.E./M.Tech degrees. She has authored more than 60 research papers published in various conferences and journals. Her research interests include machine learning, web semantics, big data analytics, and vehicular ad-hoc networks.

Neeraj Kumar [M'16, SM'17] (neeraj.kumar@thapar.edu) is working as an associate professor in the Department of Computer Science and Engineering, Thapar Institute of Engineering and Technology. He received his M.Tech. from Kurukshetra University, India, followed by his Ph.D. from SMVD University, Katra, in CSE. He was a postdoctoral research fellow at Coventry University, United Kingdom. He has more than 150 research papers in leading journals and conferences of repute. He is an Associate Editor of *IEEE TKDE*, *IEEE TII*, *IEEE TCC*, the *IEEE Internet of Things Journal*, *IEEE Communications Magazine*, *IEEE Network*, *IEEE CE Magazine*, *FGCS*, *JPDC*, *Information Sciences*, and *Computer Networks*.

Mohammad S. Obaidat [F'05] (m.s.obaidat@ieee.org) received his Ph.D. and M.S. degrees in computer engineering with a minor in computer science from Ohio State University. He is a well-known worldwide academic and scientist. He is currently a full professor at King Abdullah II School of Information Technology, University of Jordan. He has published about 55 books, over 55 book chapters, and over 700 refereed technical journal and conference articles.