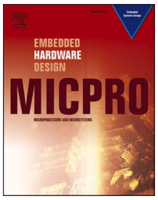
Microprocessors and Microsystems 90 (2022) 104500

Contents lists available at ScienceDirect

Microprocessors and Microsystems

journal homepage: www.elsevier.com/locate/micpro



A security framework for QaaS model in intelligent transportation systems

Majdi Rawashdeh [a], Yazan Alshboul [b], Mohammed GH. AL Zamil [c], Samer Samarah [d], Awny Alnusair [e], M. Shamim Hossain [f]

[1] *Department of Business Information Technology, Princess Sumaya University for Technology, Amman, Jordan*
[2] *Department of Information Technology, Yarmouk University, Irbid, Jordan*
[3] *Department of computer science, Yarmouk University, Irbid, Jordan*
[4] *Department of Information Technology, Yarmouk University, Irbid, Jordan*
[5] *Department of Informatics and Computer Science, Indiana University Kokomo, 2300 South Washington Street, Kokomo, IN, 46904, USA*
[6] *Department of Software Engineering, College of Computer and Information Sciences King Saud University, Riyadh 11543, Saudi Arabia*

ARTICLE INFO

Keywords:
Smart cities
Cyber security
Intelligent sensors
Intelligent transportation system ITS
Microsystems
Internet of vehicles IoV
Sensors security

ABSTRACT

Recent advances in the development of intelligent transportation systems (ITSs) impose complex services that utilize Query-as-a-Service Model in ITS microsystems. Such model is vulnerable to a vast range of security threats such as Man-in-the-Middle attacks. Intelligent sensors and microsystems provide important systems-level functionalities to smart cities applications, which enhance data acquisition and system control services. This paper proposes a communication framework that handles intrusion threats to intelligent sensors during the data acquisition and service provision phases. The contributions of this research are: (1) Proposing a reliable Query-as-a-Service communication model based on Fog computing architecture, (2) Proposing communication protocols that preserve the integrity of exchanged data through intelligent sensors, and (3) Providing a security analysis based on the mobile nature of vehicles in ITS microsystems. Our proposed methodology is a data-driven one, in which entities exchange data models instead of the data itself, thus, minimizing the communication overhead and providing a smart way to tolerate misinformation. We have conducted experiments to analyze the impact of failure rate and the size of exchanged data on our proposed framework. In addition, the computational cost has been tested against the amount of communicated data reports. The results indicated that the proposed framework showed high performance in terms of the impact of data granularity on the failure rate and computational cost. Our proposed methodology achieved 89.6% detection rate, 3.5% false-detection rate, and <0.02 probability of query failure. Accordingly, our proposed framework overcomes the major limitations of traditional cloud-based model.

1. Introduction

In modern smart cities, querying vast amount of data pools is an essential service that promotes smartness and provides on-time connectivity among different intelligent transportation systems (ITSs) [1]. Query-as-a-service (QaaS) integrates data resources, security and privacy services, and communication paradigms to deliver information as requested in the context of ITS [2]. In an edge-computing environment, it involves connecting multiple smart-city microsystems to deliver complex, implicit, and non-trivial answers to clients' inquiries. Therefore, such critical service requires reliable modeling and communication management protocols to ensure its integrity and credibility [3].

Recent advances in intelligent transportation systems (ITSs) enable the development of smart data services and applications, ranging from controlling transportations to deploying self-driving vehicles through the integration between intelligent sensors and microsystems. At early stages, ITSs have been utilized as platforms for vehicular communications to enable a vast range of on-road query-based smart applications [4] such as the architecture in Fig. 1. Such QaaS paradigm resulted in a high data load, unreliable communication in terms of security and privacy, complex multi-hop routing protocols, and extra road-units to synchronize data propagation among vehicles on roads [5]. Consequently, due to the advances in Internet-of-Things (IoT) and edge-computing, the internet-of-vehicles (IoV) provides more reasonable and efficient alternative to develop ITS microsystems in terms of communication load, reliability, scalability, and platform consistency

E-mail addresses: m.rawashdeh@psut.edu.jo (M. Rawashdeh), yazan.shboul@yu.edu.jo (Y. Alshboul), mohammedz@yu.edu.jo (M.GH.A. Zamil), samers@yu.edu.jo (S. Samarah), alnusair@iuk.edu (A. Alnusair), mshossain@ksu.edu.sa (M.S. Hossain).

https://doi.org/10.1016/j.micpro.2022.104500
Received 29 September 2021; Received in revised form 23 November 2021; Accepted 19 February 2022
Available online 12 March 2022

[6–9,33]. Therefore, edge-computing presents an efficient and secured solution for QaaS frameworks.

In smart cities, the edge-computing paradigm integrates heterogeneous data from vast range of intelligent sensors to deliver responds for various types of queries [10]. For instance, a vehicle may request an appointment at a maintenance station in the next city that can serve the current problem at the expected arrival time. Such query may involve several Fogs, or even several clouds, to be executed[11]. Furthermore, replying instant queries requires Fogs to aggregate collected data into a form that facilitates finding the approximate, or even the accurate, answer. Data aggregation, in this context, reflects the fact that hierarchical clustering of data resulted in high performance searching mechanism. Apparently, false data reports that are collected from roads' vehicles or false Fogs' query responds might affect the overall integrity and, therefore, reliability of the delivered services.

Due to its mobility nature, vehicles on the roads are moving and changing their locations while communicating via a wireless medium; making queries highly vulnerable to man-in-the-middle (MITM) attacks. Such attacks affect the overall integrity and, ultimately, the reliability of the delivered services. Furthermore, MITM attacks might breach the privacy protocols by allowing the intruder to learn more about vehicles, their locations, destinations, and much other information

As shown in Fig. 2, there are mainly two basic tasks that are required to be protected to prevent intruders from attacking the IoV query services. The first is the vehicle data reporting task, in which each vehicle sends its data and status once they changed. The second task is the query response propagation from the nearest Fog to the requested vehicle. In fact, intruders are able to change the data reports or the query response in many different ways. For instance, they can randomly flip bits in intercepted packets, which would result in delivering false contents. In addition, an intruder can learn about the location or the destination of the requested vehicle by acting as a proxy between vehicles and Fogs. Such attacks are not requiring any access to the packet's contents. Consequently, authentication protocols [12] might not serve the need of protecting mobile objects from MITM attacks; especially in a mobile environment in which vehicles are changing their corresponding authentication body (Fogs) frequently.

In this research, we propose an efficient framework aiming at protecting IoV querying services in ITS microsystems from MITM attacks. It provides solutions at different levels including the infrastructure, communication, and quality-of-life services. Specifically, the contributions of this paper are as follows:

- Proposing a data acquisition scheme called Query-as-a-Service (QaaS) that can be utilized to protect the propagation of aggregated queries in an edge-computing environment. The proposed algorithmic scheme modularizes the data acquisition process from intelligent sensors into a form that controls how mobile vehicles can preserve their data integrity while querying aggregated data from upper-level entities.
- Proposing a communication management protocol that enables both communicating sides to ensure that reported data from vehicles and query-responses from fogs are not altered. The proposed protocol is designed based on a classification algorithm to restore data into its original and proper state if it has been altered.
- Providing a security analysis method to analyze the threats on QaaS in the proposed edge computing environment. The proposed methodology is based on eliciting the query-failure rate, the response-failure rate, and privacy analysis. The analysis will consider the mobility issues such as responding to a vehicle that changes its location frequently.

This paper is organized as follows: Section 2 highlights the related work and the place of our proposed contributions in the literature Section 3. provides preliminaries about existing communication models, threats exposed to existing models, and the ultimate goals of this research Section 4. explains the protocols that are proposed in this research to handle the problem of meddling communications in Fog-based IoV systems Section 5. provides a security and performance analysis of the proposed framework. Finally, Section 6 concludes the research.

2. Related work

Recently, protecting data, which is considered an essential component in Internet-of-Things (IoT), has paid research attention in Cybersecurity and IoT domains. There are many Cybersecurity threats that may affect the efficiency and data sharing of IoV. Such attacks target IoV like Sybil attacks, impersonation attacks, Denial of service (DoS) attacks, data non-repudiation, replay attacks, forgery attacks, eavesdropping attack, and Man in the Middle (MITM) Attacks [13]. In this section, we discuss previous work that is related to the protection of IoV functions,

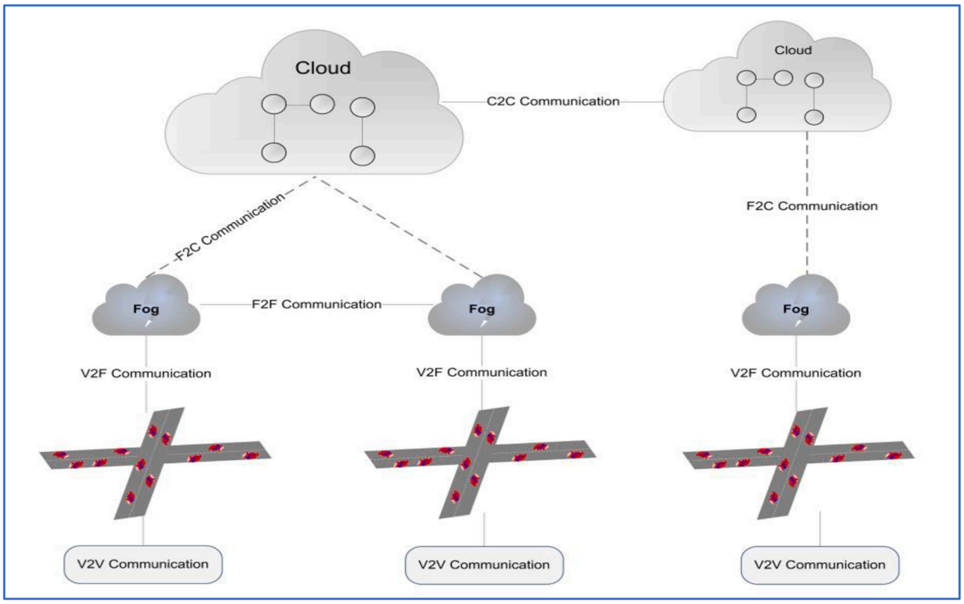


Fig. 1.. ITS internet-of-vehicles general architecture based on cloud edge-computing.

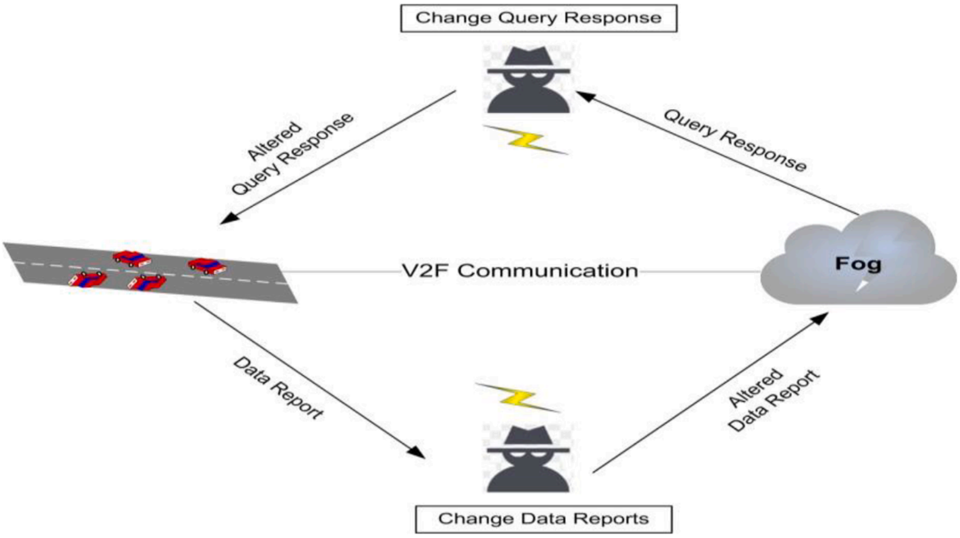


Fig. 2.. Man-in-the-Middle (MITM) attack in ITS QaaS paradigm.

particularly, protecting the integrity of queries from vehicles to data centers (fogs), which are closely related to the proposed model.

Query-as-a-Service (QaaS) in IoV environment is highly vulnerable data service to several types of attacks. In [14], Muller. et al. have discussed the impact of such attacks on the integrity of QaaS. The research showed the need for a decentralized computing architecture to efficiently implement such complex distributed data analysis design. While rule-based solutions provide simplicity [15], it lacks dynamicity to handle timely issues since such smart environment is designed to be self-responsive.

Preserving data privacy, during the aggregation task of query response from several distributed computing nodes, is a major challenge [16]. Accessing and recognizing such data may result in several types of threats and damages; especially at commercial level. Qian et al. [17] have discussed such problem and proposed a novel solution to preserve data that has been aggregated from social systems. While there are several solutions that address the preserving of data aggregation for QaaS systems, they lack efficiency in handling mobility in which the source of data is moving over time.

Lu et al, proposed a privacy-preserving aggregation scheme (EPPA) [18]. EPPA relies on a super-increasing sequence to structure multi-dimensional data communication. Furthermore, EPPA uses the homomorphic Paillier cryptosystem technique to encrypt the structured data. On another research, Lu et al proposed a lightweight privacy-preserving data aggregation scheme (LPDA) employing the homomorphic Paillier encryption, Chinese Remainder Theorem, and one-way hash chain techniques [19].

To prevent abnormal measurements caused by electricity theft or false data injection attacks in smart grid, Ni el at developed a privacy-preserving smart metering scheme to support data aggregation, differential privacy, and fault tolerance. They employed Lifted ElGamal encryption and differential privacy in their model [20]. To defend security threats , Rahman et al, designed a distributed security method using Adversarial Examples [21].

Li et al, developed an efficient algorithm to implement the n × 1-out-of-n OT which help in securing the communication between servers and client's side in IoT environment [22]. The main motivation behind the proposed algorithm is to securely transfer secrets between servers and clients (nodes) to achieve privacy-preserving data aggregation in smart grids.

Providing secure environment of Internet-of-Vehicles has attracted increasing interest to preserving data privacy. Kong et al, proposed a security model to achieve the privacy of location data [23]. Their proposed model uses the modified Paillier Cryptosystem during the vehicular sensory data collection phase and the proxy re-encryption technique during the vehicular sensory data acquisition phase to achieve the location privacy-preserving sensory data aggregation.

Data aggregation in IoV has some security challenges such as data privacy and accuracy of queries and location privacy of vehicles. Hu et al, proposed a secure and lightweight privacy-preserving data aggregation scheme SLPDA. Their proposed combines Chinese Remainder Theorem, masking technology, and Identity-based batch authentication technology [24]. Their model helps in mapping multi-dimensional data into one-dimensional scheme and reduces the authentication overhead.

To prevent chaos on the road caused by malicious users who may mislead the whole communications in IoV, Rawat et al, proposed a data falsification attack detection model that uses hashing technique to improve communication security and contention window size to improve performance [25]. To improve the security of the authentication process in IoV environment, several authors proposed proactive based privacy, authentication and edge caching scheme for IoV in [26–28,34].

Xu et al, proposed privacy-preserving data aggregation for IoV called PAVS. PAVS aim to solve vehicles' location privacy issue in VANET. It employs bilinear pairing and proper- ties of group Zp2 [29]. Kong et al., proposed a range query scheme which helps to accurately retrieve the sensed data from the distributive on-board storage in vehicular ad hoc networks (VANETs) with location privacy preservation [30].

The simplest way to detect the type of wireless sensors embedded in smart devices at modern smart cities is the analysis of the data that are generated by these sensors [31]. For instance, a motion sensor attached in a vehicle can be interpreted as the need to track the vehicle on the road and project its motion. Other sensor is measuring the level of speed, neighboring vehicles and so on. Simply, both facts are correlated.

3. System model, security threats, and design goals

This section describes the system model of vehicle-to-fog communication with essential communication stacks. In addition, the set of security threats, which can benefit from the open-medium data communication, are illustrated. Furthermore, this section concludes the atomic goals that must be achieved in order to ensure the reliability and integrity of V2F communications.

3.1. System model

In this section, we provide a detailed description about the QaaS system model to describe how the QaaS system acquires data, processes it, and delivers services to vehicles on the roads in an edge-computing architecture. As stated before, there are mainly three entities in this architecture: vehicles, fogs, and clouds. Vehicles are the source of data in this system; provide timely data frequently to their corresponding Fogs. On the other hand, Fogs are responsible for accumulating the gathered data into a summarized form; replacing the original huge amount of data with minimal ones. Such process is called data aggregation [32]. Furthermore, Fogs used to exchange their data frequently to integrate their data sources and provide Similar data services.

Fig. 3 shows the modular design of the QaaS system. To preserve the confidentiality of vehicles data, the system differentiates between the data collection task and the query delivery task. In QaaS model, vehicles are responsible only for reporting data to the nearest Fog and requesting a predefined set of queries, so that the system can preserve the privacy of data by limiting and controlling the inquired information. Indeed, a vehicle has to join the Fog's domain and can communicate directly with other vehicles via V2V communication connection. Network Authentication and Authorization has no effect on the QaaS system model.

On the other hand, a Fog is responsible for applying an aggregation method to summarize the collected data, update its data tables, and send aggregated data to other neighboring Fogs so that they keep their data sources synchronized. Such integration among Fogs guarantees consistency in responding to similar queries. Aggregated data, then, will be available for querying. The QaaS model relies on a predefined set of services in which gathering data is directed to serve their responds. For instance, queries are performed on smart city services such as: location-based services, social services, crowd sourcing services, etc.

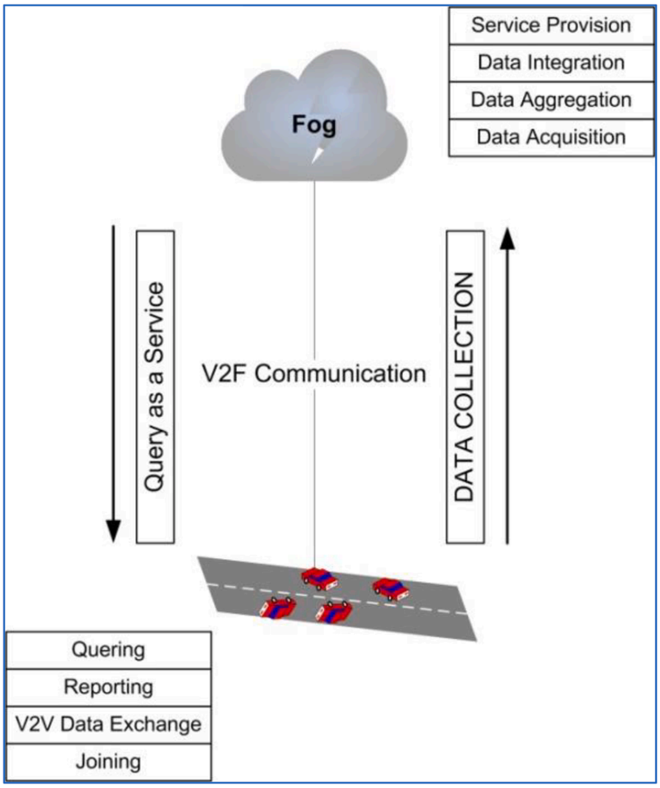


Fig. 3.. QaaS modular system.

3.2. Security threats

There are several security risks that can threaten QaaS model in a cloud-based architecture in which wireless networks are the communication medium. These threats make smart platforms, such as IoV, vulnerable to lose their reliability and integrity. Security threats, in this context, can be classified into two categories: network core and application-level threats.

Network core threats are designed to attack the network through its networking services such as authentication, addressing, and routing. For instance, unauthorized joining of IoV platform is a measure security threat. Such threat will allow intruders to breach someone else's private data or gain a higher privilege to access and alter secured data. On the other hand, spoofing is a major network core security threat. It allows endeavors to mimic legitimate network entities and acting on behalf of them. There are several network-cores spoofing mechanisms such as: ARP (Address Resolution protocol), IP (Internet Protocol Address), DNS (Domain Name Service), HTTPS (Hyper Text Transfer Protocol), and SSL (Secure Socket Layer) hijacking or stripping.

In this research, we are emphasizing our efforts to handle application level threats that are depending on the delivered services. These threats are, commonly, hard to handle due to its diversity and reliance on the type of application. In IoV's QaaS model, there are mainly two types of MITM attacks. The first one has the purpose to change the data reports that are collected from mobile vehicles on the roads. Such changing of data leads to deliver unreliable services and generates false alarms, which might harm and overwhelm the IoV platform. The second threat is focusing on altering the query-responds that are directed from the nearest Fog to the requesting vehicle. The purpose of such attack is two folded. In one-fold the mobile vehicle will receive a false respond that will confuse it and resulted in lack of integrity. In the other fold, if the vehicle recognizes the fault response, it will keep requesting the same service frequently resulted in pressuring the network traffic and resources.

3.3. Design goals

The main goal of the QaaS model, according to the security threats, is to collect data from IoV mobile vehicles and deliver querying services in a way that guarantees MITM attacks will not affect the overall system reliability and integrity. Such goal has to be met using prevention, detection, and fault-tolerance techniques. Therefore, we specify a roadmap that limits the research domain and depicts the boarders of this research by restricting the contributions of this research to meet a set of goals. Specifically, the design goals of the proposed QaaS model are:

- The proposed QaaS should prevent intruders from attacking the transmitted data from mobile vehicles and the query responds from Fog entities. Such protection should cover every threat including network-core and application-level threats.
- The proposed QaaS should be able to detect successful MITM attacks and notify both parties so that no harm occurs. The detection process should be implemented at both communication sides through a reliable scheme.
- The proposed QaaS should apply a fault-tolerance technique that is able to make the system available and deliver its designated services even if the system has been attacked.

To meet such goals at application-level, smart data modeling techniques should be employed to implement the required level of security. The analysis of exchanged data can provide valuable information to discover changes or fault alarms in an efficient way in terms of time complexity and accuracy.

4. Secured QaaS model

This section introduces two protocols: data acquisition and QaaS Communication protocols. The first illustrates how data reports are communicated between vehicles and fog entities in a reliable style, while the second protocol ensures that query requests and responses are received properly by both communication sides. To simplifying the description of protocols, Table 1 shows the mathematical and statistical annotations that have been utilized to describe the threading tasks formally.

4.1. Data acquisition protocol

Gathering data from vehicles on the roads is a complex process in terms of time-complexity and data load. It requires collecting a large amount of data chunks frequently. There are two types of data that are generated from vehicles on the roads: V2V and V2F. Vehicle-to-Vehicle (V2V) data are generated due to the interactions among the vehicles themselves for discovering the area, sharing multimedia data, or warning each other about specific events. On the other hand, Vehicle-To-Fog (V2F) data are reported to promote information sharing at large scale; allowing for further predictions and querying. Such data include sensor data, GPS, Camera and LiDAR of self-driving vehicles, user generated data, road safety data, navigation, and many others. In edge-computing architecture, FOG systems are installed to cover a particular area or service so that edge devices (vehicles) can easily detect their corresponding FOG. Generally, it is common that vehicles in specific area generate similar data. Therefore, aggregating data from multiple vehicles is a significant procedure for delivering on-time query services.

Data aggregation is an essential process for clustering data into a hierarchical form; which facilitates accessing accurate information quickly and reduces overall system latency. Specifically, each fog-node defines a semantic and cohesive group of data cluster, which is contextually related to each other. This form of hierarchical clustering allows for fast retrieval and classification of data since there is no need to search low-level and high volume information. Instead, aggregated data at fog-nodes provides aggregated and representative information about their associated road units (parent-child relationship).

At data acquisition phase, intruders can sabotage the process by changing the data reports that are generated from on-ground vehicles. Such changes will negatively affect the integrity of aggregated data at FOG entity and, therefore, resulted in unreliable query responses. To prevent such destructive action, we propose a data acquisition protocol that is able to avoid the negative consequences of MITM attacks during the data gathering phase.

To illustrate, given a set of queries that are defined as a collection of services in the QaaS model. Further, the static set of data chunks that are regularly collected to serve the QaaS service is. Since each query may require a different combination of data chunks, we define the function in which the assertion holds true. Acknowledging every chunk of data is an overwhelming process and requires a vast amount of communication. Since function points to the required data for a specific service query, we can model the validation process as a prediction algorithm and communicate the data-model instead of the data itself. To implement this protocol, the service component at FOG entity must fit into an appropriate data model; for example: linear or non-linear data models.

Table 1
The notations and their description.

Notation	Description
	The set of predefined queries that represents V2F services.
	The query.
	The total number of predefined queries.
	The data vector that represents all aggregated data on the fog entity.
	The data chunk in the data vector.
	The data vector that represents the local data in vehicle .
	The total number of vehicles that are available instantly on the fog segment.
	A function that holds the required data by a specific service query .
	A vehicle with identification.
	A data model that encompasses the characteristics of a specific data set.
	Modeling relation.
	Margined error.
	The response data vector to a specific query.
	The distance parameter that has been computed at Fog entity.

Algorithm 1 explains the acquisition task and shows how vehicles and fogs validate reports and acknowledgements. Specifically, vehicles have mainly two functions: reporting and validating. At reporting thread, a vehicle used to sense its own environment (status) continuously. Once changed, the vehicle collects its reporting data through the function Sense. Then, the algorithm gathers the data chunks into a data vector and sends it immediately to the nearest Fog. Lines 1–7 in the vehicle thread illustrate this task. The time complexity of the reporting function is linear in terms of the length of the data vector (i.e.).

Moreover, once the FOG entity receives the data vector (report) from a vehicle, it conducts an aggregation task in which data are aggregated according to the services that required utilizing it. A data chunk may contribute in serving more than one service. Such fine-grain redundancy simplifies the service delivery model. The function defines the relationship between data chunks and available services. During the aggregation process, a data model that is suitable to the existing data is built; defining an approximate prediction based on acquisition time. Then, the predicted error (or misclassification rate) is computed so that if the prediction error is less than or equal to, the system will skip. In other words, this action tolerates the impact of several attacks that create a marginal effect on the data services. Finally, the Fog entity disseminates the model and the error value to associated vehicles. Lines 1 to 6 in the FOG thread illustrate this procedure. The time complexity of this part is defined as the number of data chunks that are required by each service multiplied by the number of existing services. In the worst case, the time complexity is, where ***K*** is the number of existing services (Queries).

Finally, every vehicle must ensure that their reports have received successfully and accurately. For this reason, once the vehicle received the disseminated model, it compares its own model with the received one. If the error rate is not acceptable (i.e.), then the vehicle notifies the FOG with the last data report indicating that the model is inaccurate. Lines 1 to 4 in the Validate thread illustrate the validation process. The time complexity of the validation process is.

Algorithm 1. Data acquisition protocol

```
FOG:
1|Thread :=Aggregation (D) {
2|   for each for each $h(q_{h(q)} \mid \forall(i) \leq k\ _i) \mid \forall(i) \leq k$
3|       $M_i \dashv$ fit(h($q_i$,D,time)
4|   $\alpha$=Actual Error(D, Mi,qi)
5|   Disseminate (V,$M_i$,$\alpha$) }

| Vehicle:
1|Thread:= Report($V_r$,D){
2|    if (status changes) Then
3|       Sense ($V_r$)
4|       for each $d_i \in D$
5|          fill-in($d_i$)
6|       Send(D)
7|    End if}

1| Thread:= Validate($V_r$,D,$M_i$,$\alpha$)){
2|       $M_x \dashv$ fit(D)
3|          if $|M_x$*- $M_i| \geq \alpha$
4|                  Report($V_r$,D)
```

As shown before, the time complexity of the acquisition protocol depends on two variables; ***D*** and ***K***. since the number of queries in the proposed model is fixed, ***K*** is considered always less than ***D*** (i.e. $|D| \ll K$). Therefore, the overall time complexity of the data acquisition protocol is $O(|D|)$.

4.2. QaaS communication protocol

As stated before, both data reporting and services responding tasks are subject to intrusion attacks. However, while the data acquisition (in Algorithm 1) has handled the former task, it is necessary to define the communication model that secures the service responses from being manipulated. Exchanging data models instead of acknowledgements guarantees the integrity between local data and associated data on the Fog entity. Furthermore, such style maintains reliable and timely synchronization among both fogs and vehicles. Communication and data integrity among fogs and clouds is out of this research scope, since it depends heavily on the core networking layers that might be different with respect to the services providers.

Algorithm 2. QaaS communication protocol

```
Thread := Query_Response($V_r$,$q_i$,$D_v$){
Compute $h(q_i) \dashv D_v$
Distance= $|h(q_i)$-ReData$|$
Reply($V_r$,Distance,RData) }
Vehicle: Thread := Request($V_r$,$q_i$,$D_v$) {
Send($V_r$,q,$D_v$)
Wait until receiving the Response
Validate($V_r$,FogDist$_i$,RData)}
Thread := Validate($V_r$,FogDist$_i$,RData){
Compute $h(q_i) \dashv D_v$
Distance= $|h(q_i)$-RData$|$
if(Distance $\neq$FogDist)
Request($V_r$,$q_i$,$D_v$)}
```

Algorithm 2 explains the service provision model in which a vehicle requests a service query by providing its identification, query identification, and its local data vector. At the receiver side, the Fog computes the function using the vehicle's local data vector. The Fog entity responses with mainly two data objects: the distance between query response vector and the function in addition to the query response itself.

Validating such communication paradigm is simply performed by computing the function locally, which must be equivalent because it has been computed using the same local data vector. The distance between both and the response data vector are computed. If both Fog and Vehicle distance parameters are equal, this implies that the response does not change. Otherwise, the vehicle repeats the request until receiving a reliable reply.

5. Security analysis and performance evaluation

This section provides an analysis to evaluate the proposed methodology. The analysis emphasizes on two performance evaluation properties: failure rates and computation cost. The failure rate analysis focuses on testing the querying failure rates with respect to time and average number of reported data. The computation cost analysis has been conducted to measure the performance of the proposed algorithms in terms of computation power versus the amount of exchanged data during requesting and responding phases.

To test our proposed methodology, we designed a simulated environment for a group of taxis that span five districts in San Francisco. These districts are Chinatown, Golden Gate, Financial, Nob Hill, and Western Addition. According to a real map, the simulator mimics the communication between these edge-computing sources and their associated fogs; assuming that each district is controlled by a fog and all fogs are managed by a cloud system. The QaaS functionality is defined to serve different classes of requesters: taxi drivers, customers, police, and telecommunication authorities. The simulation model has been designed with 50 vehicles and a maximum of 200 requests per hour. Moreover, the simulator plants a random number of bugged communications that present intrusions and MIMT attacks. The number of attacks is defined as a random number in the range of 10–30 per hour.

Three experiments have been performed to measure the performance of the proposed framework. The first measures the ability of the proposed model to detect intrusion threats. The second experiment is focusing on the reliability of communication during the data gathering phase. It is crucial since failing to deliver the data reports negatively affects the overall accuracy of QaaS-based services. Finally, the security layer will, normally, add extra computational cost. For this reason, we conducted a cost analysis to show how this layer will affect the overall all complexity of such systems.

5.1. Detecting security threats

In this experiment, we measure the performance of the proposed method in terms of its ability to detect intrusions and MITM attacks. As stated before, the simulator ran for 100 times with maximum number of vehicles is 50 and a maximum of 200 requests per round. The number of involved fogs is 5, which depends on the location of reporting vehicles. The simulator generates random attacks at each round. Two indicators have been computed: detection rate and false detection. The first measure the ability of the tool to detect actual threats, while the other measures the false detection rate or the percentage of proper requests that the tool mistakenly considers them as threats. Both measurements have been calculated according to the following formula:

As shown in Table 2, there are a strong relationship between the number of requested services and the false detection rate. On the other hand, we found no statistically significant relationship between detection rate and other attributes. This is leading us to conclude that the proposed framework does not affect by the amount of data and requests during the system lifecycle. Fig. 4 shows how the false detection rate affected by different attributes that have been generated through 100 running rounds.

5.2. Data querying failure analysis

The querying failing analysis provides evidence on the communication reliability; especially when communication between vehicles and fogs fail to deliver the data report. To apply such analysis, we assumed that every fog covers a specific segment. Given a set of segments, in which each Fog covers a segment such as represents this relation. Then, we assume that the probability of failing to receive the data report is less than the threshold parameter. Finally, we define the reporting time as the parameter, which defines the interval of time that vehicles conduct the data acquisition task.

Having these parameters allow for modeling the problem as a Poisson distribution as a function of the number of data reports generated by vehicles in a specific segment in a specific time period. Given the number of data reports as the function, denotes the number of data reports in segment and denotes the number of data reports generated in segment at the time period interval.

To apply Poisson probability distribution, we define the probability function, denotes the probability of failure at specific time period, which is represented as follows: where .

Fig. 5 depicts the querying failure probability rate against time. As shown in Fig. 5, the framework achieved low failure that approach to zero at high time interval. Further, at average amount of time, the curve has shown liner-stabilization. On the other hand, the performance in terms of failure rate was at worst at low time interval. The reason behind the low performance at the low time interval is highly related to the low quality of the generated data models due to lack of enough data to generate accurate models, which results in high margined error. While after a small amount of time, these models performed better to stabilize the failure rate. Notice that, as a consequence of having a failure, the system resends the request again until it succeeds to pass the validation

Table 2
Security threat detection.

N	Number of Communicating Vehicles	Number of Requested Services	Number of Involved Fogs	Actual Number of Threats	Detection Rate	False Detection
1	44	186	4	12	%83.3	%1.6
2	32	111	5	25	%92.0	%2.7
3	48	200	4	20	%90.0	%2.0
4	41	163	5	14	%93.0	%3.1
5	29	101	5	16	%87.5	%3.0
6	34	173	3	11	%81.8	%2.9
7	31	96	2	17	%82.4	%2.1
8	18	58	1	10	%90.0	%3.4
9	49	179	3	30	%93.3	%2.2
10	36	162	5	27	%88.9	%3.1
…	…	…	…	…	…	…
100	50	198	5	29	%89.6	%3.5

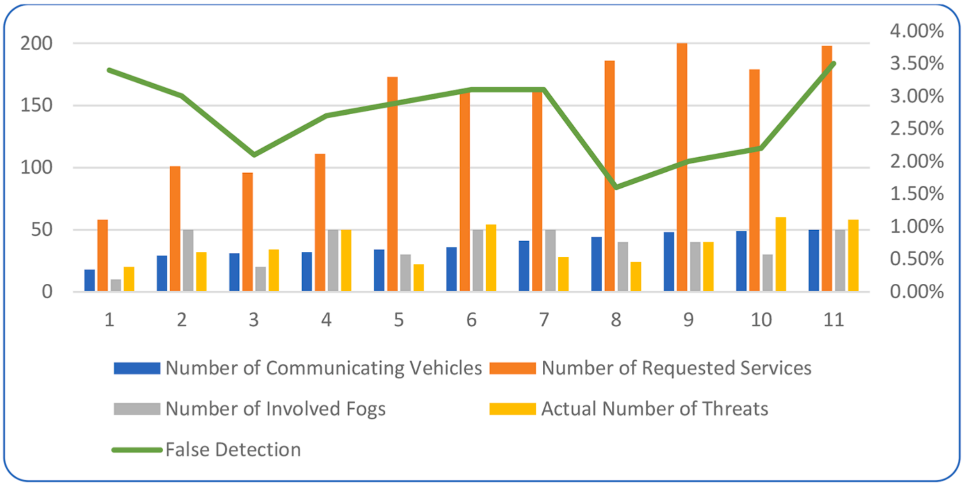


Fig. 4.. False detection rate analysis.

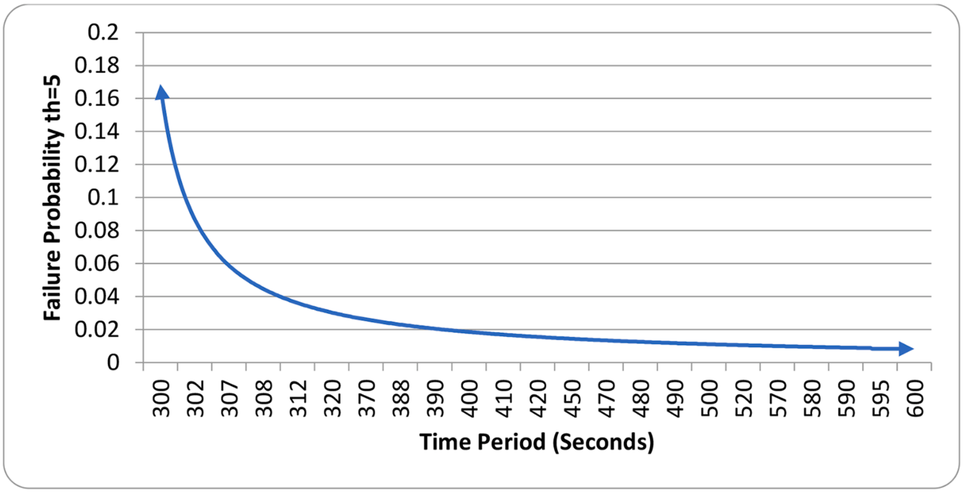


Fig. 5.. Failure rate with respect to time dimension.

process.

To measure the failure probability with respect to the average number of data reports, we consider the parameter as the average number of data reports generated in a specific segment. Furthermore, we define the probability function to denote the probability of data reporting with the assertion: . Accordingly, .

Fig. 6 depicts the relationship between the average number of data reports and the querying failure rate. As shown in the figure, the impact of the report size is significantly low at high number of reports, while it is at worst at low number of data reports. Since data reports from vehicles act as the only source for building the data models, the impact of increasing the number of data reports is slightly positive as the amount of data increase. Such reports provide the input for the data models, which became more accurate as the number of reports increase.

5.3. *Computational cost analysis*

This analysis shows a comparison between two system models: Fog-

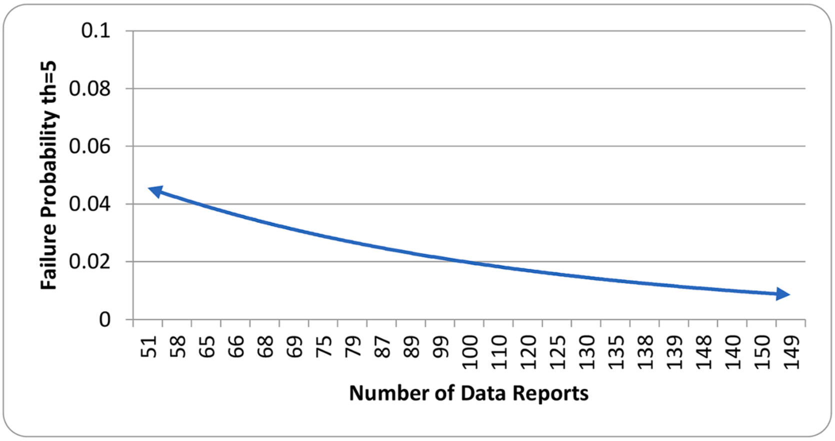


Fig. 6.. Failure rate with respect to the average number of data reports.

based and Cloud-based (Traditional centralization paradigm). The purpose of conducting such analysis is to prove that decentralizing the communications and the security tasks has a positive impact by protecting the communications and, also, accelerate data exchanging and processing tasks. For this reason, we conduct two experiments to measure the impact of the number of data reports and the size of the service responds on the processing speed.

The experiments took place on a machine with the following specifications: Intel (R) Processor Core (TM) i5-7200U @ 2.5 GHz, 2701 Mhz, 2 cores and 4 logical processors, Microsoft Windows 10 (Home Edition), and 8 GB RAM.

As shown in Fig. 7, the computational rate in traditional communication model increased exponentially as the size of the data reports increased. On the other hand, in Fog-based communication, the computation rate increased linearly. The reason behind such difference is that Fog-based communications rely on distributing processing and communication tasks over multiple entities (Fogs) resulted in a distributed processing paradigm. This type of communication adds synchronization overhead but remains better than centralizing the processing toward single entity.

As shown in Fig. 8, the communication overheads that are required to synchronize fogs entities resulted in slightly low performance at low sized responds. On the other hand, as the size of responds increased, the fog-based processing remains stable at fixed rates while the traditional centralized processing increased exponentially. Distributing data over multiple entities decrease the processing time significantly.

6. Conclusion

This paper introduces a framework that models the communication of vehicles-to-fogs for the purpose of minimizing the effect of man-in-the-middle attacks. The paper proposes a data-driven protocols that exchange models instead or data chunks, which minimize the communication load and provides an intelligent way to tolerate errors. The paper provides detailed and formal descriptions of the proposed communication protocols. The validation phases in these protocols were able to guarantee the reliability and integrity of such vulnerable communication environment. The security analysis that has been conducted showed that the proposed framework behaved better than traditional ones in terms of its adaptability to failure rates and the size of exchanged data. Furthermore, it shows an acceptable performance in terms of time complexity and computational time with respect to service provision task.

Acknowledgement: This work was supported by the

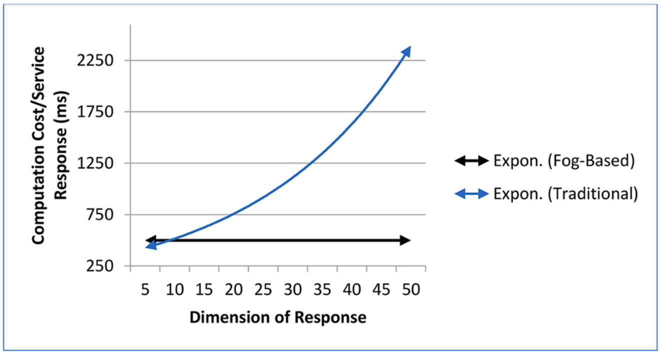


Fig. 8.. Computation rate versus the size of service responds in fog-based and traditional communication models.

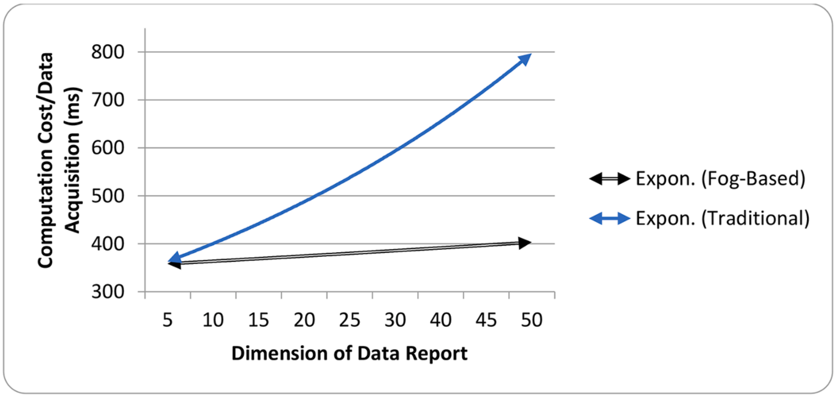


Fig. 7.. Computation cost versus size of data reports in fog-based and traditional communication models.

Researchers Supporting Project number (RSP-2021/32), King Saud University, Riyadh, Saudi Arabia.

Declaration of Competing Interest

None.

Acknowledgments

This work was supported by the Researchers Supporting Project number (RSP-2021/32), King Saud University, Riyadh, Saudi Arabia.

References

[1] Y. Zhang, et al., Multi-aspect aware session-based recommendation for intelligent transportation services, IEEE Trans. Intell. Transp. Syst. 22 (7) (2021) 4696–4705.

[2] H. Lin, et al., Blockchain and deep reinforcement learning empowered spatial crowdsourcing in software-defined internet of vehicles, IEEE Trans. Intell. Transp. Syst. 22 (6) (June 2021) 3755–3764.

[3] M.G.A. Zamil, Multimodal daily activity recognition in smart homes, in: Proceedings of the 6th International Conference on Control, Decision and Information Technologies (CoDIT), IEEE, 2019, pp. 922–927.

[4] C. Lochert, H. Hartenstein, J. Tian, H. Fussler, D. Hermann, M. Mauve, A routing strategy for vehicular ad hoc networks in city environments, in: Proceedings of the IEEE IV2003 Intelligent Vehicles Symposium. Proceedings (Cat. No. 03TH8683) (pp, IEEE, 2003, pp. 156–161.

[5] F. Sakiz, S. Sen, A survey of attacks and detection mechanisms on intelligent transportation systems: VANETs and IoV, Ad Hoc Netw. 61 (2017) 33–50.

[6] I. Kabin, Z. Dyka, D. Klann, J. Schaeffner, P. Langendoerfer, On the complexity of attacking commercial authentication products, Microprocess. Microsyst. 80 (2021), 103480.

[7] X. Xu, Y. Hao, et al., Energy efficient task caching and offloading for mobile edge computing, IEEE Access 6 (2018) 11365–11373.

[8] X. Jian, et al., Blockchain-empowered trusted networking for unmanned aerial vehicles in the B5G era, IEEE Netw. 35 (1) (2021) 72–77. January/February.

[9] Lin, K., Luo, J., Hu, L., Hossain, MS, Ghoneim, A. (2027), Localization based on social big data analysis in the vehicular networks. IEEE Trans. Ind. Inf..13 (4) 1932-1940.

[10] M.G. Al Zamil, S. Samarah, Application of design for verification to smart sensory systems, in: Proceedings of the Qatar Foundation Annual Research Conference Proceedings Volume 2014 Issue 2014, Hamad bin Khalifa University Press (HBKU Press, 2014, 1ITPP0366.

[11] K. Lin, et al., Green video transmission in the mobile cloud networks. *IEEE transactions on circuits and systems for video technology* 1st, 27, IEEE, 2016, pp. 159–169.

[12] M.A. Rahman, et al., Blockchain and IoT-based cognitive edge framework for sharing economy services in a smart city, IEEE Access 7 (2019) 18611–18621.

[13] T. Garg, N. Kagalwalla, P. Churi, A. Pawar, S. Deshmukh, A survey on security and privacy issues in IoV, Int. J. Electr. Comput. Eng. 10 (5) (2020) 5409–5419, https://doi.org/10.11591/IJECE.V10I5.PP5409-5419.

[14] I. Müller, R. Marroquín, G. Alonso, Lambada: interactive data analytics on cold data using serverless cloud infrastructure, in: Proceedings of the 2020 ACM SIGMOD International Conference on Management of Data, 2020, pp. 115–130.

[15] X. Wang, C. Wang, J. Zhang, M. Zhou, C. Jiang, Improved rule installation for real-time query service in software-defined internet of vehicles, IEEE Trans. Intell. Transp. Syst. 18 (2) (2016) 225–235.

[16] L. Xiong, S. Chitti, L. Liu, Preserving data privacy in outsourcing data aggregation services, ACM Trans. Internet Technology (TOIT) 7 (3) (2007) 17–es.

[17] S. Qian, et al., Social event classification via boosted multimodal supervised latent dirichlet allocation, ACM Trans. Multimedia Comput. Commun. Appl. 11 (2) (2015) 1–22.

[18] R. Lu, X. Liang, X. Li, X. Lin, X. Shen, EPPA: an efficient and privacy-preserving aggregation scheme for secure smart grid communications, IEEE Trans. Parallel Distrib. Syst. 23 (9) (2012) 1–11, https://doi.org/10.1109/TPDS.2012.86.

[19] R. Lu, K. Heung, A.H. Lashkari, A.A. Ghorbani, A lightweight privacy-preserving data aggregation scheme for fog computing-enhanced IoT, IEEE Access 5 (2017) 3302–3312, https://doi.org/10.1109/ACCESS.2017.2677520.

[20] J. Ni, K. Zhang, K. Alharbi, X. Lin, N. Zhang, X.S. Shen, Differentially private smart metering with fault tolerance and range-based filtering, IEEE Trans. Smart Grid 8 (5) (2017) 2483–2493, https://doi.org/10.1109/TSG.2017.2673843.

[21] M.A. Rahman, et al., Adversarial examples—Security threats to COVID-19 deep learning systems in medical IoT devices, IEEE Internet Things J. 8 (12) (2020) 9603–9610.

[22] R. Li, C. Sturtivant, J. Yu, X. Cheng, A novel secure and efficient data aggregation scheme for IoT, IEEE Internet Things J. 6 (2) (2019) 1551–1560, https://doi.org/10.1109/JIOT.2018.2848962.

[23] Q. Kong, R. Lu, M. Ma, H. Bao, A privacy-preserving sensory data sharing scheme in internet of vehicles, Future Gener. Comput. Syst. 92 (2019) 644–655, https://doi.org/10.1016/j.future.2017.12.003.

[24] P. Hu, et al., A secure and lightweight privacy-preserving data aggregation scheme for internet of vehicles, Peer-to-Peer Netw. Appl. 13 (3) (2020) 1002–1013, https://doi.org/10.1007/s12083-019-00849-6.

[25] D.B. Rawat, M. Garuba, L. Chen, Q. Yang, On the security of information dissemination in the internet of vehicles, Tsinghua Sci. Technol. 22 (4) (2017) 437–445.

[26] Y Zhang, R Wang, M.S. Hossain, MF Alhamid, M Guizani, Heterogeneous information network-based content caching in the internet of vehicles, IEEE Trans. Veh. Technol. 68 (10) (2019) 10216–10226. Oct. 2019.

[27] L. Hu, et al., Proactive cache-based location privacy preserving for vehicle networks, IEEE Wirel. Commun. 25 (6) (2018) 77–83.

[28] X. Yang, et al., Automatic visual concept learning for social event understanding, IEEE Trans. Multimed. 17 (3) (2015) 346–358.

[29] C. Xu, R. Lu, H. Wang, L. Zhu, C. Huang, PAVS: a new privacy-preserving data aggregation scheme for vehicle sensing systems, Sensors 17 (3) (2017) 1–18, https://doi.org/10.3390/s17030500.

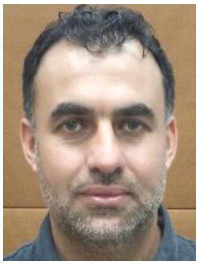
[30] Q. Kong, R. Lu, M. Ma, H. Bao, Achieve location privacy-preserving range query in vehicular sensing, Sensors 17 (8) (2017) 1–16, https://doi.org/10.3390/s17081829.

[31] A. Yassine, et al., IoT big data analytics for smart homes with fog and cloud computing, Future Gener. Comput. Syst. 91 (2019) 563–573.

[32] P. Castillejo, G. Johansen, B. Cürüklü, S. Bilbao-Arechabala, R. Fresco, B. Martínez-Rodríguez, J. Häggman, Aggregate farming in the cloud: the AFarCloud ECSEL project, Microprocess. Microsyst. 78 (2020), 103218.

[33] Z. Yu, et al., Mobility-aware proactive edge caching for connected vehicles using federated learning, IEEE Trans. Intell. Transp. Syst. 22 (8) (2021) 5341–5351.

[34] Y. Hao, et al., Smart-edge-CoCaCo: AI-enabled smart edge with joint computation, caching, and communication in heterogeneous IoT, IEEE Netw. 33 (2) (2019) 58–64.

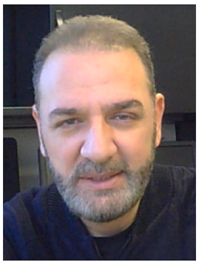


Majdi Rawashdeh received his Ph.D. degree in Computer Science from the University of Ottawa, Canada. He is currently an associate professor at Princess Sumaya University for Technology (PSUT), Jordan. His research interests include social media, user modeling, recommender systems, smart cities, and big data.

Yazan Alshboul is an assistant professor in the department of Information Technology and the coordinator of cybersecurity program at Yarmouk University. He earned his Ph.D in information systems/ information assurance and computer security from Dakota State University in USA. His research interest is in cybersecurity and data analysis.

Mohammed GH. AL Zamil is full professor in the department of computer science at Yarmouk University (YU) in Jordan. He obtained his Ph.D degree in Information Systems from Middle East Technical University, Ankara, Turkey (2010). His master degree is in Computer Science from YU. He had B.Sc. degree in computer science from YU. His research interests include: Data Mining, Wireless Sensor Networks, Model Checking, Software verification, and Software Engineering.

Samer Samarah is a full professor in the department of Information Technology at Yarmouk University, Jordan. He obtained his PhD in Computer Science from University of Ottawa, Canada in 2008. Dr. Samarah has many published journals and conferences in the area of data mining and wireless networks. His research focuses on discovering behavioral patterns from data collected by Wireless Sensor Networks, Vehicular ah-hoc Networks, Data Analytic and IoT. Dr. Samarah is a referee for many international journals and conferences.



Awny Alnusair received his Ph.D. degree in Computer Science from the University of Wisconsin-Milwaukee. He is currently an Associate Professor of Informatics and Computer Science at Indiana University Kokomo. His research interests fall in the broad domain of Software Engineering. The specifics, however, are continually changing but generally include Program Comprehension, Software Maintenance, Programming Languages, Information Retrieval, Cloud Computing, and the Semantic Web. He has recently focused on research projects that deal with Cybersecurity, Big Data Analytics, and Data Mining in Vehicular Ad Hoc Networks.

M. Shamim Hossain is currently a Professor with the Department of Software Engineering, College of Computer and Information Sciences, King Saud University, Riyadh, Saudi Arabia. He is also an adjunct professor with the School of Electrical Engineering and Computer Science, University of Ottawa, ON, Canada. He received the Ph.D. in Electrical and Computer Engineering from the University of Ottawa, ON, Canada in 2009. His research interests are on cloud networking, smart environment (smart city, smart health), AI, deep learning, edge computing, Internet of Things (IoT), multimedia for health care, and multimedia big data. He has authored and coauthored more than 325 publications including refereed journals (270+ SCI/ISI-Indexed papers, ***150+ IEEE/ACM Transactions/Journal papers***, 12+ ESI highly cited papers, 1 hot paper), conference papers, books, and book chapters. Recently, he co-edited a book on "*Connected Health in Smart Cities*", published by Springer. He has served as the cochair, general chair, workshop chair, publication chair, and TPC in several IEEE and ACM conferences. He is the chair of IEEE Special Interest Group on Artificial Intelligence (AI) for Health with IEEE ComSoc eHealth Technical Committee. Currently, he is the Organizing Co-Chair of the Special Sessions with IEEE I2MTC 2022. He is also the Co-Chair of the 1st IEEE GLOBECOM 2021 Workshop on Edge-AI and IoT for Connected Health. He was the Co-Chair of the special session "AI-Enabled technologies for smart health monitoring", held with IEEE I2MTC 2021. He was the co-chair of the 3rd IEEE ICME Workshop on Multimedia Services and Tools for smart-health (MUST-SH 2020). He is the Technical Program Co-Chair of ACM Multimedia 2023. He is a recipient of a number of awards, including the Best Conference Paper Award and the **2016 *ACM Transactions on Multimedia Computing, Communications and Applications (TOMM) Nicolas D. Georganas Best Paper Award,*** *the* ***2019 King Saud University Scientific Excellence Award (Research Quality)****, and the Research in Excellence Award from the College of Computer and Information Sciences (CCIS), King Saud University* (**3 times in a row**). He is on the editorial board of *the IEEE Transactions on Instrumentation and Measurement (TIM), IEEE Transactions on Multimedia (TMM),ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM), IEEE Multimedia, IEEE Network, IEEE Wireless Communications, IEEE Access,* Journal of Network and Computer Applications (Elsevier), International Journal of Multimedia Tools and Applications (Springer), Games for Health Journal, and International Journal of Information Technology, Communications and Convergence (Inderscience). Previously, he served as a guest editor of *ACM Transactions on Internet Technology, IEEE Communications Magazine, IEEE Network, IEEE Transactions on Information Technology in Biomedicine* (currently *JBHI*), *IEEE Transactions on Cloud Computing*, International Journal of Multimedia Tools and Applications (*Springer*), Cluster Computing (*Springer*), Future Generation Computer Systems (*Elsevier*), *Elsevier*), Sensors (*MDPI*), and International Journal of Distributed Sensor Networks. He is a senior member of the IEEE, and Distinguished member of ACM. He is an IEEE ComSoc Distinguished Lecturer (DL).