Chapter 13
Privacy Protection in Vehicular Ad-Hoc Networks

Gongjun Yan
University of Southern Indiana, USA

Bhed Bahadur Bista
Iwate Prefectural University, Japan

Danda B. Rawat
Georgia Southern University, USA

Wu He
Old Dominion University, USA

Awny Alnusair
Indiana University – Kokomo, USA

ABSTRACT

The first main contribution of this chapter is to take a non-trivial step towards providing a robust and scalable solution to privacy protection in vehicular networks. To promote scalability and robustness the authors employ two strategies. First, they view vehicular networks as consisting of non-overlapping subnetworks, each local to a geographic area referred to as a cell. Each cell has a server that maintains a list of pseudonyms that are valid for use in the cell. Each pseudonym has two components: the cell's ID and a random number as host ID. Instead of issuing pseudonyms to vehicles proactively (as virtually all existing schemes do) the authors issue pseudonyms only to those vehicles that request them. This strategy is suggested by the fact that, in a typical scenario, only a fraction of the vehicles in an area will engage in communication with other vehicles and/or with the infrastructure and, therefore, do not need pseudonyms. The second main contribution is to model analytically the time-varying request for pseudonyms in a given cell. This is important for capacity planning purposes since it allows system managers to predict, by taking into account the time-varying attributes of the traffic, the probability that a given number of pseudonyms will be required at a certain time as well as the expected number of pseudonyms in use in a cell at a certain time. Empirical results obtained by detailed simulation confirm the accuracy of the authors' analytical predictions.

DOI: 10.4018/978-1-4666-4691-9.ch013

1. INTRODUCTION AND MOTIVATION

Recent statistics show that in 2008 there were over 238 million passenger cars and trucks in the US, a vehicular fleet that increases yearly by almost seven million new cars (US Department of Transporation, Research and Innovative Technology Association, 2011). In an effort to help their vehicles compete in the marketplace, car and truck manufacturers are offering more and more potent on-board devices, including powerful computers, a large array of sensors, radar devices, cameras, and wireless transceivers. These devices cater to a set of customers that expect their vehicles to provide a seamless extension of their home environment populated by sophisticated entertainment centers, access to Internet and other similar wants and needs (Arif et al., 2012; Wang, 2010). The powerful on-board devices support new applications, including location-specific services, on-line gaming, delivering multimedia content and various forms of mobile infotainment made possible by the emergence of vehicular networks (Li et al., 2005). In the near future, a vehicle will be capable of intelligent data-mining based on its owner's preferences (Wen et al., 2011), identifying favorite hotels, shopping malls, restaurants (e.g. Chinese restaurants featuring Szechuan-style cuisine) and, perhaps, a convenient parking lot (Yan et al., 2011). Knowing the driver's preferences, around lunchtime the vehicle will automatically send queries to the roadside infrastructure and other vehicles to find a list of Chinese restaurants nearby (Li et al., 2005; Wen et al., 2011).

The increased Internet presence that enables the above applications invites various forms of privacy attacks mounted by unscrupulous characters in order to identify the location of various parties that might be exploited for financial gains. One well-known such attack has for goal to establish that a family is away from their home so that a burglary can be perpetrated; yet another one has for goal to obtain compromising information that can later be used to blackmail the driver. Invariably, these privacy attacks are mounted by exploiting the various forms of correlation that exist between the identity of a vehicle and that of its driver.

While a great deal of research has been devoted to information security in vehicular networks (Choi et al., 2006; Hubaux et al., 2004; Raya et al., 2006; Sun et al., 2010a, Yan et al., 2008; Yan et al., 2009a), far less attention has been given to privacy issues (Xie et al., 2010; Yan and Olariu, 2011). One of the reasons for this state of affairs is the mistaken idea that the privacy issues encountered in vehicular networks are similar to the ones experienced in cellular telephony and WiFi communications and, therefore, the same solutions can be applied. For example, it has been suggested that instead of radio communications, drivers use their cell phones to access the Internet. However, using a cell-phone while driving may not only be illegal, as it is currently in some states, it has also been identified as one of the principal causes of traffic accidents.

A more careful analysis reveals that many of the privacy challenges experienced in vehicular networks are either brought about or exacerbated by the increased on-line presence of drivers, the high mobility of the vehicular fleet as well as the short transmission requirements of the Dedicated Short Range Communications (DSRC) limiting transmission to between 300m and 1,000m (Yan and Olariu, 2011).

In summary, there are unique challenges to privacy protection in vehicular networks including (Yan et al., 2013; Arif et al., 2012):

- **High Vehicular Mobility:** This challenge renders the network connection inherently unstable and make pseudonyms difficult to manage and update (Yan and Olariu, 2011; Rawat et al., 2011). Therefore, the communication is not reliable;
- **Large and Fluctuating Population of Vehicles:** This challenge will make the scalability requirement of privacy solutions difficult to meet (Yan et al., 2012);

- **Traffic Flow Characteristics:** The vehicles in the network are bound by certain traffic flow parameters that are highly related to the geographical location differences;
- **Unmistakable Correlation Between a Vehicle and its Driver:** Tracking the driver often means tracking its vehicle and vice-versa.

1.1 Our Contributions

As mentioned, assigning pseudonyms to vehicles is a time-honored solution to providing privacy protection both in real-life and in vehicular networks (Huang et al., 2011). The major shortcoming of existing approaches is *scalability*. This is because the high rates of vehicular mobility combined with the short communication range of DSRC make the task of issuing, maintaining and revoking pseudonyms extremely difficult. The short-range vehicular communications rely on multihop routing which will involve multiple vehicles in the network. However, the more vehicles involved, the higher the potential for a privacy attack. The growing size of our vehicular fleet in conjunction with a steadily increasing Internet presence require privacy provisioning schemes to be both scalable and robust.

Considering the challenges outlined above, we take a non-trivial step towards protecting network layer and above privacy of vehicles. Our first main contribution is to provide a robust and scalable solution to privacy protection in vehicular networks. To promote scalability and robustness we employ a combination of two strategies:

- We view vehicular networks as consisting of non-overlapping subnetworks each local to a geographic area referred to as a *cell*. Depending on the topology and the nature of the area, these cells may be as large as few city blocks or, indeed, may comprise the entire downtown area of a smaller town. Each cell has a server that maintains a list of pseudonyms that are valid for use in the cell. Each pseudonym consists of the cell's ID and of a random host ID;
- Instead of issuing pseudonyms to vehicles proactively, as virtually all existing schemes do, we issue pseudonyms only to those vehicles that need them, and therefore request them. This strategy is suggested by accumulated empirical evidence suggesting that only a fraction of the vehicles in an area will engage in communication with other vehicles and/or with the infrastructure and, therefore, need pseudonyms. The others do not.

Our second main contribution is to model analytically the time-varying request for pseudonyms in a given cell. This is important for capacity planning purposes since it allows managers to predict stochastically the probability that a given number of pseudonyms will be required at a certain time as well as the expected number of pseudonyms in use in a cell at a certain time. Empirical results obtained by detailed simulation confirmed the accuracy of our analytical predictions.

Let us elaborate a bit to give the reader a better feel for what we do. Guided by the divide-and-conquer strategy, we partition the geographic area of interest into smaller entities that we call *cells*, where the size of a cell is dictated by the characteristic of the environment and will be discussed later. The municipality-wide vehicular network is partitioned into several subnetworks, each local to a cell. Each cell has its own pseudonym server that assigns, on demand, pseudonyms to the vehicles resident in the cell. To easily locate a network node, the address of a vehicle includes two parts: the cell ID and the pseudonym as a host ID. Pseudonym servers in various cells are connected by wired connection. Packets can be transmitted following the cell ID. Inside a cell, vehicles are located using their host IDs. This simple scheme helps scalability, robustness and the reliability of communication. If necessary, scalability can be enhanced by further dividing cells into microcells.

Vehicles that wish to communicate using the wireless channel need to request pseudonyms from the cell server prior to communicating with either the infrastructure or with other vehicles in the cell. However, it is worth noting that a vehicle does not need to request pseudonyms if it is merely receiving information from the network, such as listening to music, receiving traffic condition updates, etc.

The real identity of vehicle, either its host name, or its IP address, or both, will be hidden by the pseudonym. In this chapter, the identities of the vehicles are temporary and random numbers, the identity collector/attacker will not be able to track the vehicle by sniffing the network.

Operating similarly to the Dynamic Host Configuration Protocol (DHCP), the cell server needs to have an accurate estimate of number of pseudonyms, the probability and the expected number of pseudonym requests. Knowing dynamic information at any time, the server can maximize the utilization of resources (for example IP addresses as pseudonyms, bandwidth, etc.), as the resources are costly. But the difficulty lies in the time-varying nature of the pseudonym requests, mirroring the time-varying population of vehicles in the cell. One of our main contributions is to derive analytically the time-varying expected number of pseudonym requests in a given cell as a function of time. Empirical simulation results have confirmed the accuracy of our analytical derivations.

The remainder of this work is organized as follows: Section 2 reviews relevant results from the recent literature. Sections 3 and 4introduce the system model assumed in the chapter as well as the privacy treat model whose effects we mitigate. They also discuss requirements of a privacy-preserving system. Section 5 presents the details of our privacy provisioning scheme. Further, Section 6 and 7 offers our analytical derivations both of the privacy scheme and of the stochastic model that we propose for pseudonym usage in a cell as a function of traffic intensity. Section 8 offers an empirical evaluation of our model using extensive numerical simulations. Finally, Section 9 offers concluding remarks and directions for future work.

2. STATE OF THE ART

The goal of this section is to review a number of approaches to preserving privacy in vehicular networks. The *mix zone* concept has been proposed recently for privacy protection (Le et al., 2011; Dahl et al., 2010; Palanisamy and Liu, 2011; Sun et al., 2010b). The mix zone is intended as a strategy to break the link between an old pseudonym and a new pseudonym assigned by the roadside infrastructure. Since the attackers can store past pseudonyms and link the new pseudonym to a vehicle that has been tracked, a mix zone can thwart attempts at tracking vehicles by mixing with other vehicles (Ribagorda-Garnacho, 2010; Sampigethaya et al., 2007). In spite of its novelty, the mix zone concept has not met with great success. One of the problems seems to be the synchronization of pseudonyms. If the traffic flow is high, there may have communication delay of synchronization of pseudonyms.

A *silent period* or *silent zone* is another solution to breaking the link between old and new pseudonyms (Dok et al., 2010; Song et al., 2009; Dahl et al., 2010). Group navigation can protect the privacy of several vehicles as a group and can decrease the overhead of pseudonym changes by individual vehicles (Studer et al., 2009; Sampigethaya et al., 2005; Guo et al., 2007). A group leader will normally be elected to represent the whole group. On the other hand, there are potential problems. The privacy of the group is highly dependent o the integrity of the group leader. If the group leader is compromised, the privacy of the whole group will temporally compromised.

Dok *et al.* (Dok et al., 2010) tried to merge the three strategies discussed above to provide better privacy protection. The mix zone is often selected as a place to assign pseudonyms. The group signature can allow vehicles in a group to use the same key to communicate (Studer et al., 2009; Sampigethaya et al., 2005; Guo et al., 2007). This is especially useful when the vehicles are in a mix zone or silent zone. Since the privacy information such as combination of identity, location and time can be stored and used to link the pseudonyms, adopting a random silent period can break the link between a pseudonym and the real identity of a vehicle.

The privacy analysis at an intersection has been discussed in (Dahl et al., 2010). By using a combination of public and private key encryption, the roadside infrastructure in charge of the intersection will assign keys to vehicles entering the intersection. A formal analysis of the transmission has been conducted by using ProVerif (Blanchet et al., 2008). RFID as a electronic device/tag has also been applied in privacy protection in vehicular networks (Arapinis et al., 2010; Brusò et al., 2010; Delaune et al., 2010).

Lu *et al.* (Lu et al., 2008) proposed a privacy preservation protocol in vehicular networks. The basic idea of the proposed method is to dynamically generate anonymous keys between the On-Board Units and the Roadside Units, which can provide fast anonymous authentication and privacy protection while minimizing the required storage for short-time anonymous keys. The authors proposed a filtering algorithm to prevent communication information from encrypting junk information (Lu et al., 2012b, Lu et al., 2012a). However, strictly speaking, the proposed methods are more of a security-preserving strategy than a privacy preservation method as the network identity (such as IP address, or network name) can still be used in tracking the identity of a driver. By contrast, our proposed method focus on network layer pseudonym protection that guarantees that the identity of the vehicles and drivers are both protected.

Lin *et al.* (Lin et al., 2011; Lu et al., 2010) proposed an interesting routing protocol which preserves privacy. To help location-based routing protocol, they proposed a social-tier-assisted packet forwarding protocol by using some social spots, such as well-traversed shopping malls and busy intersections in a city. Without knowing the receiver's exact location information, a packet can be first forwarded and disseminated in the social tier. When the receiver visits one of social spots, it can successfully receive the packet. Vehicle can somehow preserve conditional privacy. However the method is at least two-hops communication and the delivery ratio, packet jitters and average delay are questionable. The method is also based on driver behavior. In many cases, drivers never visit social spots. Our proposed privacy-preserving scheme can overcome the potential problems raised by the social-related method. First, our scheme is reliable and independent of driver behavior and, second, the delivery ratio, packet jitters and average delay shown in simulation are greatly improved as the communication are one-hop.

3. SYSTEM MODEL

The main actors that we deal with in this chapter are the vehicles described in Subsection 3.1, the roadside infrastructure deployed by the municipality and/or third-party players to provide various services to the traveling public as discussed in Subsection 3.2 and the cell model which is the workhorse of our solution and will be discussed in Subsection 3.3.

3.1 The Vehicle Model

An important new concept in the automotive industry is *neighborhood awareness*. This allows a vehicle to know about the presence, location and even speed of neighboring vehicles. It is common-knowledge that present-day vehicles come equipped with powerful on-board resources. Specifically, we assume vehicles to be endowed with the following features (Yan et al., 2013; Arif et al., 2012):

- A GPS navigation system including a GPS receiver and GPS maps;
- A computer center, which will provide data processing, computing and storage;
- A wireless transceiver, using DSRC for fast communications;
- Public Key Infrastructure. We assume that the public key and private key of each vehicle and infrastructure are assigned and maintained by Certificate Authority (CA);
- A virus checker. However, virus protection is outside the scope of this chapter and will not be pursued further.

It is worth noting that the vehicle model adopted in this work is highly realistic and that, in fact, it is not uncommon to have all the assumed features in the high-end vehicular fleet. For example, Toyota Motor Corporation developed Pre-Crash Safety system (Toyota, 2007) which uses millimeter-wave radar to sense vehicles and obstacles on the road ahead back in 2002. Furthermore, GPS and computer center are popular vehicle accessories nowadays. To put things in perspective, and to illustrate the phenomenal technological advances that make their way into present-day vehicles, suffice it to mention that only a few short years back, a comparably-equipped vehicle used to be referred to as a "smart vehicles" (Hubaux et al., 2004; Yan et al., 2008).

3.2 Roadside Infrastructure Model

An important role in our scheme is played by the road side infrastructure deployed by the municipality to provide various services to the traveling public. While at present these services are minimal, we expect that in time they will develop to a full-blown roster of municipality and third-party services ranging from traffic updates to information about local events, parking availability, medical facilities, restaurants, and the like. The roadside infrastructure may be queried about road construction, congestion and can provide, on demand, travel estimates to various points on interest and up-to-the-minute parking lot availability.

The roadside unit has a powerful transceiver and the electronics needed to communicate with the vehicles in the cell. The down-link channel is of the broadcast type, the up-link channel (i.e. from the vehicles to the roadside unit) is contention-based. The roadside infrastructure of neighboring cells is connected by conventional high-bandwidth fiberglass cable.

Last, but certainly not least, the roadside infrastructure houses a pseudonym server that issues, on demand, pseudonyms to the vehicles currently in the cell.

3.3 Cell Model

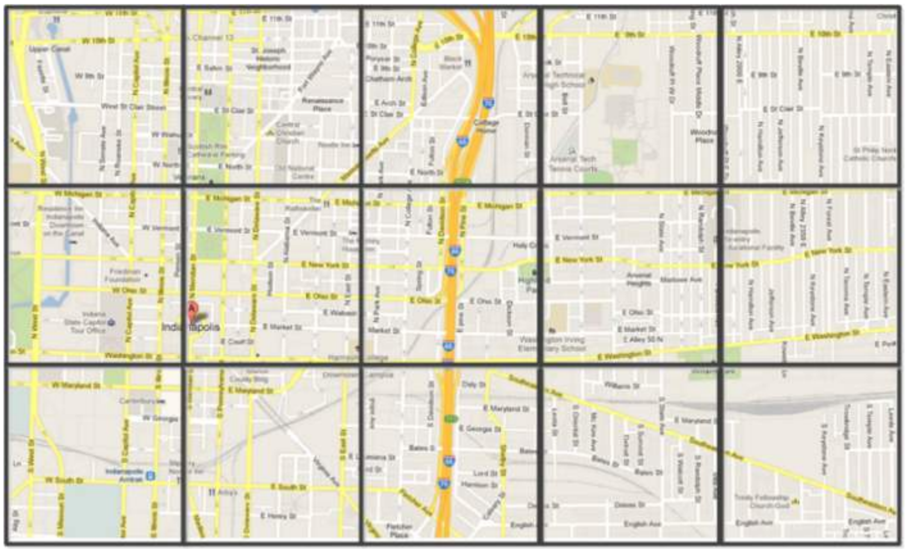
The system discussed in this chapter assumes an urban environment. While we assume the downtown area of a large city, the same reasoning can be applied to a small town. The only difference is that in a small town, there might be only one cell or, indeed, a few cells, while in a larger urban setting the downtown area may be partitioned into several cells.

For reasons of efficiency, similar in nature to the ones motivating cellular telephony service providers, the city-wide vehicular network is partitioned into many smaller subnetworks, each local to a cell, as shown in Figure 1. The details of this partitioning will be discussed in Section 5.1.

Some of the advantages of the cell-based communication include:

- **Localized Communication:** This follows from the observation that drivers tend to be more interested in local information, such as traffic congestion, accidents, parking lot, or gas station, etc.;
- **Enabling Scalability***:* No matter how far a destination is located, a vehicle can access it with help from the cell;
- **Customizing Security and Privacy Strategies:** Different cells can be deployed

Figure 1. Illustrating the partition of downtown Chicago into cells.(Yan et al., 2013; Arif et al., 2012)



with different security and privacy protocols. For example, users along a busy highway will want fast communication with least delay caused by encrypting and decrypting message (Yan et al., 2009b). A simpler cryptographic algorithm can be employed in this area.

4. THREAT MODEL AND REQUIREMENTS

During its residency in a cell, a vehicle may contact many other vehicles and the roadside infrastructure. Therefore, it is important to protect the privacy of vehicles by ensuring that their communication cannot be tracked. The privacy protector will not be able to protect privacy in the situation that vehicles may move out of scope of the protector or the protector will serve large amount of vehicles. Obviously, the privacy protection must be scalable because of the high mobility and large population of vehicles. The threat categories are:

- **Linking Pseudonyms:** The attackers can store the past identities (e.g. pseudonyms) of a vehicle and record the received new identities to link the past pseudonyms with the new pseudonyms (Yan et al., 2008);
- **Global Exploring***:* The attackers can obtain full control of the network including roadside infrastructure, service servers, etc. The attackers therefore can explore any network infrastructure to track the identity of a vehicle;
- **Passive Eavesdropping***:* The attackers can install a powerful radio receiver to passively eavesdrop the identity and location information of other vehicles nearby;
- **Spoofing User Identity***:* The attackers pretend to be another user to obtain data and illegitimate advantages. One classic example is "man-in-the-middle attack" in which the attackers pretend to be Bob when communicating with Alice and pretend to be Alice when communicating with Bob. Both Alice an Bob will send decryptable messages to the attackers. Another similar example is "email address spoofing" in which the attackers fill with forged

return user's identity and create unreachability errors.

With the development of vehicle registration plate recognition, the attackers can track a vehicle by physically following the vehicle. The physical track is not able to prevent by simply use software method (Xi et al., 2007).

4.1 Requirements of a Privacy-Preserving Solution

Vehicular networks are complex systems with time-varying dynamics. While privacy protection in vehicular networks is our main concern in this chapter, there are other related, but equally important issues that any solution must consider. These issues translate into the following three requirements (Yan et al., 2013; Arif et al., 2012):

- **Hiding the real identity of vehicles in an effective way:** A vehicle's real identity must be replaced by a pseudonym which is a random number;
- **Routing packets in an efficient way:** It is helpful to think of the identity of a vehicle like something like an IP address of a network device in the TCP/IP protocol. The identity will be used to route packets and to locate the host;
- **Protecting the privacy of vehicles and their drivers in a scalable way:** Due to high vehicular mobility and a large number of vehicles on our roadways and streets, privacy-preserving strategies must be scalable and robust.

4.2 Level of Private Message

Based on the sensitivity, there are several levels of messages: Public, Personal, Private, Confidential and Private.

1. **For Public:** No sensitivity level is assigned to the message. This message can be transmitted to the whole network.
2. **For Personal:** The recipient will treat the message as personal information. The identity of the sender may or may not shown in the message. It is up to the sender to decide the appearance of the identity.
3. **For Private:** The recipient will treat the message as private information. No identities will be recovered.
4. **For Confidential and Private:** The recipient will treat the message as confidential and private information. No identities will be released and only the authenticated recipient can read this message.

5. OUR PRIVACY-PRESERVING SCHEME

Referring to Figure 1, our scheme partitions the geographic area of interest into non-overlapping cells. In turn, the municipality-wide vehicular network is partitioned into subnetworks, each local to a cell. Each subnetwork maintains a list of pseudonyms and assigns, on demand, a pseudonym to each vehicle in the cell. In principle, the pseudonym is valid for the duration of the residency in the cell. Each pseudonym is composed of two parts (Yan et al., 2013; Arif et al., 2012):

- **The ID of the Cell:** used as a geographic network prefix ID and mask ID that specifies the maximum number of vehicles presented in the cell;
- **The Pseudonym Assigned to a Vehicle:** used as a host ID which uniquely identifies the vehicle while in the cell. One example of the pseudonym is an IP address. With such a pseudonym, vehicles can route packets based on the network ID and receivers can be easily located inside a cell by host ID.

It is worth mentioning that the use of pseudonyms is not mandatory and the decision to use them or not is left with individual drivers who can opt to communicate without protection privacy. Cities, especially some busy districts, often broadcast information for example traffic congestion information, parking lot information, through FM radio or through the broadcast IP address. If a vehicle merely receives information passively, the vehicle does not have to request a pseudonym. It is only when the vehicle is about to send out requests or messages (for example, a direction to a specific address) that the vehicle needs to request a pseudonym. This is obvious if the pseudonym is an IP address. Bearing in mind the fact that pseudonym usage is not mandatory, we can assume that some vehicles will not request a pseudonym for the most part of their trip.

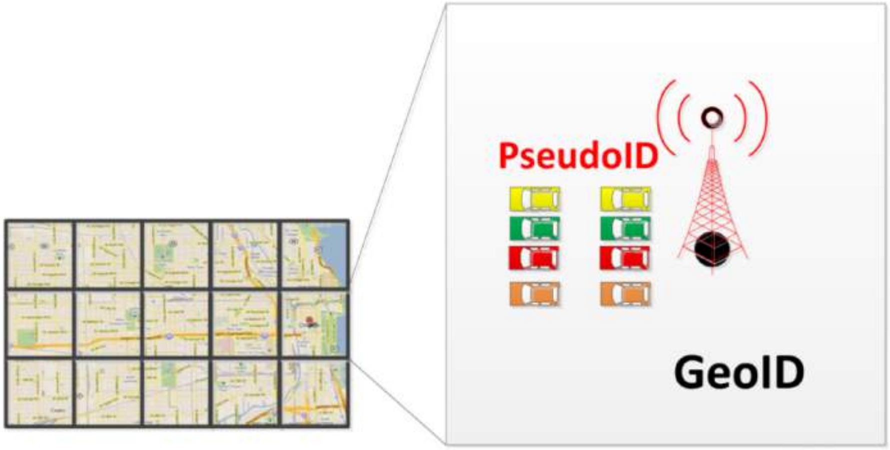
5.1 Cell Communication Enhances Privacy

In this subsection, we introduce cell communication which enhances privacy by reducing the linkage between the identity and location of vehicles. As we have already mentioned, in an urban scenario the whole vehicular network is partitioned into cells on the digital map, as shown in Figure 1. The digital maps with cell partitions are installed on vehicles, just as GPS navigator systems install digital maps with many other useful information such as gas stations, hotels, etc. The size of the cell is a system parameter and depends on a number of factors. For example, in the US the vehicles are assumed to use the DSRC protocol limiting the effective communication range to 1000m and so the cells must have edges of roughly 700 meters so that any two vehicles inside the cell can communicate with each other and with the pseudonym server regardless of their position within the cell. Thus, in a cell vehicles can directly communicate with the pseudonym server in one hop.

As illustrated in Figure 2, each cell has a transmission tower and a pseudonym server to provide privacy protection and wireless communication routing. Each vehicle can automatically find which cell the vehicle belongs to by checking the GPS location and the cell's map.

In this chapter, we assume that a pseudonym uniquely identifies a vehicle, such as IP address, host-name in network. We briefly list the steps involved in requesting a pseudonym (Yan et al., 2013; Arif et al., 2012):

Figure 2. Each cell contains a trusted pseudonym server. GeoID is the unique identity of a cell while PseudoID is the unique identity of a vehicle inside the cell.(Yan et al., 2013; Arif et al., 2012).



- The vehicle first encrypts a request by using the pseudonym server's public key. The request will also include a secret key (random number) that the vehicle will use to uncover new pseudonym. When the pseudonym server receives this request, the server will apply its private key to decrypt the message. In this step, the requesting vehicle will send its real identity. For security reasons, the pseudonym server will store an encrypted version of the real vehicular identity;
- The pseudonym server will authenticate the vehicle. If the vehicle is eligible to receive a new pseudonym, the pseudonym server then allocates an available random pseudonym to the vehicle. Having encrypted the new pseudonym with the secret key the pseudonym server then sends the encrypted message back to the vehicle. The pseudonym include a non-negative expiration time, i.e. a time-to-live (TTL);
- The vehicle applies the new pseudonym after decrypting and uncovering the new pseudonym. When the pseudonym expires, a new request has to be sent by the vehicle. In this chapter, we take the view that the TTL is set to infinity, that is, pseudonyms are valid to use during the entire residency in the cell. However, in general this need not be the case.

Based on the above discussion, it is reasonable to assume that the pseudonym servers will not be compromised and that the public key infrastructure (PKI) including public key and private key pairs cannot be cracked.

5.2 Pseudonyms Hide the Identity of Vehicles

The major goal of this subsection is to discuss how the real identity of a vehicle is hidden and how the requirements specified in Subsection 4.1 are satisfied. As already mentioned, and as illustrated in Figure 3, the identity of a vehicle consists of two main parts: a cell(i.e. subnetwork) ID (GeoID prefix) and a host ID (Host ID). The subnetwork ID is shown as GeoID and the host ID is combined with subnetwork ID by adding the subnetwork ID to the host ID. The HostId is a random number generated and maintained by the cell pseudonym server. Therefore, the combination of the host ID and the subnetwork ID can also be thought of as a pseudonym.

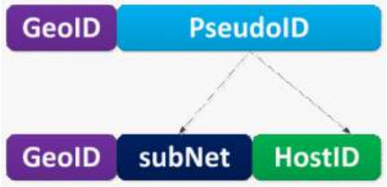
To illustrate the previous discussion let us follow an example. Referring to Table 1, a vehicle ID is 64 bits, e.g. *0932.0968.0115.1300* (dot-hexdecimal notation). The subnetwork ID is specified by subnet mask, i.e. Geonet mask *FFFF.FFFF.FFFF.0* in table 1. Therefore, the GeoID prefix is obtained by a logical "AND" operation between the vehicle ID and the Geonet mask. The host ID is shown as *0000.0000.0000.1300*. The broadcasting address inside the cell will be *0932.0968.0115.FFFF* which can be obtained by replacing the host ID by all "1"s. It is clear that the network localization can use the GeoID prefix. Once the cell as a subnetwork has been identified, the Broadcast Address can be used to find the host vehicle.

Once a vehicle has obtained a pseudonym, it can use it to communicate with other vehicles and/or with the roadside infrastructure until it exits the cell or the pseudonym expires. As mentioned already, cell pseudonym servers are con-

Table 1. Identity pseudonyms: parameters and values.(Yan et al., 2013; Arif et al., 2012)

	Dot-HexDecimal notation
Vehicle ID	0932.0968.0115.1300
Geonet mask	FFFF.FFFF.FFFF.0000
GeoID prefix	0932.0968.0115.0000
Host ID	0000.0000.0000.1300
Broadcast Address	0932.0968.0115.FFFF

Figure 3. Illustrating the pseudonym structure. A 64 bits number is partitioned into two parts: GeoID which is the cell's ID and the PseudoMask which includes a subnetwork mask and a pseudonym ID which is a random number.(Yan et al., 2013; Arif et al., 2012).



nected with each other by a wired network. Messages can be transmitted by following the network ID on wired networks. Inside a cell, the destination vehicle can be located by host ID. However, the network localization is very efficient since the size of a cell is small when compared to the whole network. Therefore, our method is both scalable and robust.

5.3 Dividing a Cell into Microcells to Improve Scalability

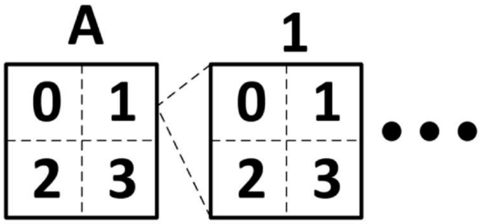
Seen from the perspective of a cell, the population of vehicles is fluctuating with time as vehicles constantly move in and out the cell. According to recent statistics published by the National Highway Traffic Safety Administration, we assume that the percentage of vehicles that will request pseudonym is about 24.5% of the population of vehicles in the cell (National Highway Traffic Safety Administration, 2012) although in our analytical derivations we take a more general view and let the probability that a vehicle requests a pseudonym be time-dependent.

When the request for pseudonyms reaches its peak, it may place a great deal of pressure on the pseudonym server. We present a method to improve scalability by dividing, on a per-need basis, a congested cell into microcells in a manner similar to that of cellular networks. A cell can be divided into microcells, as shown in Figure 4. Each microcell has its own pseudonym server which has the same capability to process pseudonym requests, validation, and updates. The division to microcells can recursively proceed until the request rejection rate drops to zero.

5.4 Location Division Multiple Access (LDMA) to Reduce Collisions

Given that the number of vehicles in a cell can be large, the *broadcast storm* (Lu and Poellabauer, 2010) can be a potential problem that can greatly reduce the efficiency of communications. To mitigate the problem, we propose to use *Location Division Multiple Access* (LDMA), inspired by the classic Time Division Multiple Access (TDMA) protocol, to schedule the communications in order to reduce the number of wireless communication collisions. Like TDMA, data stream is divided into frames. Each frame is further divided into time slots. Each slot is assigned to a small region of a cell. Vehicles in the small region will share the same time slot. Therefore, the idea is to divide a cell into smaller sub-cells, i.e. the cell is divided into 8 slots, to reduce the wireless communication collision. Comparing with TDMA, LDMA only assigns time slots to a sub-cell, a smaller region to the original cell, instead of an individual user. For scalability, a sub-cell can be divided into finer granularity super-sub-cells and a time slots can be partitioned into smaller time fractions to

Figure 4. Illustrating the partition of a congested cell into microcells.(Yan et al., 2013; Arif et al., 2012).



host super-sub-cells respectively. The structure of LDMA is illustrated in Figure 5.

5.5 Quiet Times Prevent the Tracking of ID and Location

To avoid tracking of vehicular ID and location, a silent zone will be adopted. The major function of the silent zone is to break the link between the old ID/location and the new ID/location. Therefore, it is natural to have the silent zones located at intersections because the semi-random mobility of vehicles can enhance the breakage of the correlation between the old ID and new ID. Inside the silent zone vehicles stop using their IDs. But this stoppage will not prevent vehicles from exchanging emergency messages. Only ID related applications will be temporally blocked. This is acceptable for most applications because the duration of the blockage will be only a few seconds. When a vehicle exits the silent zone, it will start using the new ID. The vehicles that have no new ID will keep using their current (i.e. old) ID. (Figure 6)

5.6 Pseudonym Synchronization

A number of agencies can participate in generating pseudonyms. These agencies can be governmental transportation authorities, such as DMV or BMV. These governmental authorities are ideal agents to serve as pseudonym servers. Pseudonyms can also be issued by roadside infrastructure built by DMV/BMV and forced to be updated every expiration period even if the car is not on the road. The vehicular maintenance habits of drivers will not be changed and drivers will not be forced to perform other actions to get their vehicle certified and provided with a suitable set of pseudonyms.

Figure 5. Illustrating LDMA. Data stream is divided into frames. Each frame is divided into time slots. Each time slot is assigned to a sub-cell which is formed by partitioning a cell into 8 sub-cells. Vehicles in the cell slot can communicate only within its time slot.(Yan et al., 2013; Arif et al., 2012).

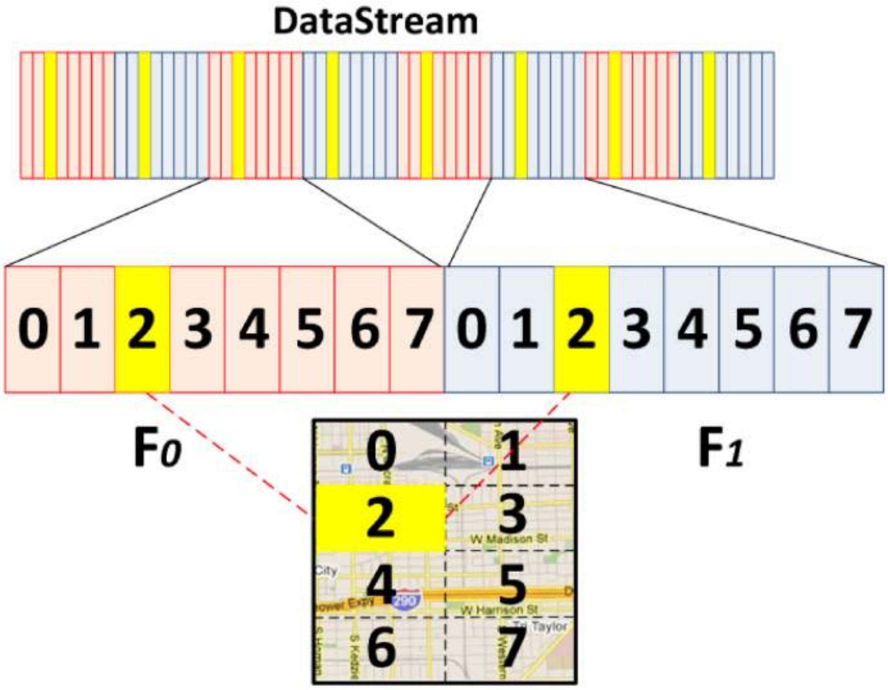
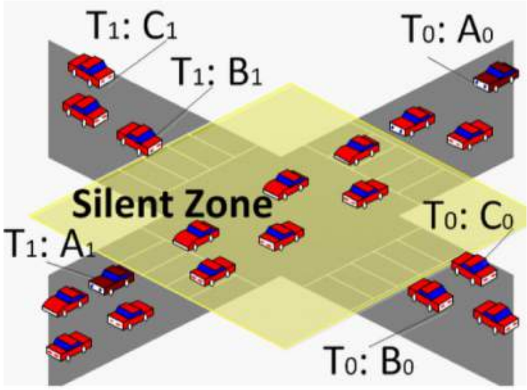


Figure 6. Illustrating the silent time $T_1 - T_0$ and silent zone. (Yan et al., 2013; Arif et al., 2012).



However, there are several problems associated with a central pseudonym authority, the single-point-of-failure concern being one of them. Yet another problem may be that the server IDs are not synchronized among cell servers. This problem will disable packet routing. To synchronize pseudonym servers, a possible solution is make a Pseudonym Server (PS) chain which includes several levels of PSs. Each local pseudonym server receives certificates and server IDs from the higher level servers. Each of them has a series of PSs. A tree structure can be organized as suggested by Figure 7.

Vehicle manufacturers can be the authorized to issue pseudonym as well. Manufacturers will receive permission and certificates from governmental transportation authorities and become a subdivision of PS. In addition, non-profit organizations can also act as authorized organizations. Similar to the vehicle manufacturer, non-profit organizations can obtain permission and certificates from governmental transportation authorities and become a subdivision of a PS.

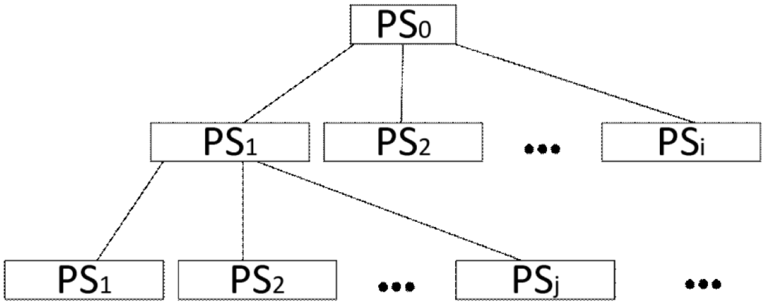
At fabrication time, each vehicle will receive a pseudonym from the manufacturer or some governmental agency by using PKI encryption. Pseudonym assignment is on the basis of the unique ID and a certain expiration time. The pseudonym has to be periodically renewed at local pseudonym servers such as cell pseudonym servers, DMV/BMV pseudonym servers as sub-PSs.

5.7 Pseudonym Update and Expiration

There are three ways a pseudonym can expire. First, pseudonyms are time-sensitive. As mentioned before, each pseudonym is assigned a TTL value. When the TTL decreases to zero, the pseudonym expires and the vehicle will be automatically deregistered. A pseudonym is also associated with the issuing cell ID, i.e. GeoID shown in Figure 2.

Second, when a vehicle exits a cell and enters a new cell, the vehicle can requests a new pseudonym which will submit the old pseudonym as well. The pseudonym server in the new cell will identify the

Figure 7. Illustrating the Pseudonym Server (PS) tree structure which ensured that server IDs are synchronized.(Yan et al., 2013; Arif et al., 2012).



previous cell ID and TTL. If the TTL is still valid, a synchronizing message will be sent to the old cell to notify the old pseudonym server in the old cell. The old pseudonym server will deregister the vehicle that just entered the new cell.

Since each time of applying new pseudonym is not based on the full capacity of the pseudonym pool, we will need to estimate the size of the pseudonym pool. Assume that the expected number of pseudonyms is $E[X(t)]$ where $X(t)$ means the pseudonym server receives k requests at time t. Therefore, in order to accommodate the difference between theoretical predictions and the actual number of requests experienced by the cell, the size of the pseudonym pool may be taken to be $(1+\delta)E[X(t)]+1$ where $0<\delta<1$ is an application-dependent parameter that will have to be fine-tuned as empirical evidence accumulates.

5.8 Preventing Denial 0f Service Attacks

As the pseudonym server is a single service provider in a cell, it can be subjected to Denial of Service (DoS) attacks which are relatively easy to mount and hard to prevent. The damage DoS can inflict includes resource depletion such as wasting computing, memory, storage, and network resources, service exception and starving, among many others. Our main strategies to mitigate the effects of DoS attacks are the following:

- **Location Authentication:** The senders of the pseudonym request must be inside the cell. The location authentication can detect the sender's location by using both GPS coordinates and signal strength. The location authentication can block most, if not all, the requests from outside the cell;
- **Filtering Pseudonym Requests:** The pseudonym server only provides pseudonym services. Any other type of requests will be rejected;
- **Request Throttling:** The pseudonym server only assigns a reasonable request quota to every street in the cell. If the quota of a street is exceeded, it would not use other street's quota.

6. PRIVACY ANALYSIS: MODEL 1

In this section, we present a model to analyze the cell. Suppose roads be partitioned into slots. Each slot can only hold exact one vehicle. Therefore, when a vehicle present at a place, the vehicle actually take one slot. Consider a single cell where the number of possible slots for vehicles is finite and the total number of slots is N. Denote by $X(t)$ the number of slots in use at time t. Then, our physical intuition is not violated by assuming that $\{X(t);t>0\}$ is a birth-death process. It also seems reasonable to assume that, the cars arrive at a rate of $\lambda(t)$, independent of the number of cars already in the traffic area, and if the traffic area contains k cars, then the departure rate is $k\mu(t)$, where $\mu(t)$ is a function of t.

The state space of this process is $S=\{0,1,\ldots,N\}$. For every positive integer k, $1\leq k\leq N$, the event $X(t)=k$ occurs if the traffic area contains k cars at time t. We let $P_k(t)$ denote the probability that the event $X(t)=k$ occurs, that is

$$P_k(t) = P\{X(t) = K\}$$

To make the mathematical derivations more manageable, at this point we assume that $\lambda(t)=\lambda$ and $\mu(t)=\mu$. Thus, the transition rate matrix of this birth-death process is

$$Q = \begin{pmatrix} -\lambda & \lambda & \cdots & 0 & 0 \\ \mu & -\lambda+\mu & \cdots & 0 & 0 \\ 0 & 2\mu & 0 & 0 & \vdots \\ \vdots & \ddots & \vdots & \vdots & 0 \\ 0 & \cdots & -\left[\lambda+(n-1)\mu\right] & (N-1) & \mu 0 \\ 0 & \ldots & N\mu & -N\mu & \end{pmatrix}$$

And the Fokker-Planck equation is as follows:

$$\begin{cases} \dfrac{dP_0(t)}{dt} = -\lambda P_0(t) + \mu P_1(t) \\ \dfrac{dP_j(t)}{dt} = \lambda p_{j-1}(t) - (\lambda + j\mu)P_j(t) + (j+1)\mu P_{j+1}(t) \\ \dfrac{dP_N(t)}{dt} = \lambda P_{N-1}(t) + N\mu P_N(t) \end{cases}$$

If the limit distribution $P = (P_0, P_1, P_N)$ exist, that is $P_j = \lim_{t\to\infty} P_{ij}(t)$ and it is independent of *i*, then we have $dP_j(t) / dt = 0$ as $t \to \infty$. Hence, in the above equation let $t \to \infty$ we have

$$\begin{cases} -\lambda P_0 + \mu P_1 = 0 \\ \lambda P_{j-1} - (\lambda + j\mu)P_j + (j+1)\mu P_{j+1} = 0 \\ \lambda P_{N-1} + N\mu P_N = 0 \end{cases} \tag{1}$$

Let g, $g_j = \lambda P_{j-1} + j\mu P_j$ we have

$$\lambda P_{j-1} - (\lambda + j\mu)P_j + (j+1)\mu P_j + 1$$

$$= \left[\lambda P_{j-1} - j\mu P_j\right] - \left[\lambda P_j - (j+1)\mu P_{j+1}\right]$$

$$= g_j - g_{j+1} = 0$$

Hence $g_0 = 0$, and $g_j = g_{j+1}$, for $1 \le j \le N-1$. Now the equation system (1) can be written as

$$\begin{cases} P_1 = \dfrac{\lambda}{\mu} P_0 \\ P_j = \dfrac{\lambda}{j\mu} P_{j-1} = \dfrac{1}{j!}(\dfrac{\lambda}{\mu})^j P_0 \\ P_N = \dfrac{1}{N}(\dfrac{\lambda}{\mu})^N P_0 \end{cases}$$

It is known that $\sum_{j=0}^{N} P_j = 1$, that is

$$\left[\frac{\lambda}{\mu} + \frac{1}{2!}\left(\frac{\lambda}{\mu}\right)^2 + \frac{1}{N!}\left(\frac{\lambda}{\mu}\right)^N\right] P_0 = 1$$

Hence, solve the above equations, we have

$$P_j = \frac{\dfrac{1}{j!}(\dfrac{\lambda}{\mu})^j}{\sum_{i=0}^{N} \dfrac{1}{i!}(\dfrac{\lambda}{\mu})}$$

7. PRIVACY ANALYSIS: MODEL 2

The main goal of this section is to offer a stochastic analysis of the privacy scheme discussed in the previous sections. The expected number of pseudonyms and the probability of a certain number of pseudonyms in a cell at any time are our main interests. The main reasons we need to model the expected number and probability of vehicles in a cell at time *t* include:

- **Reducing Security Risks:** If the pseudonym server includes a list of unused pseudonyms, attackers may take the unused pseudonyms and pretend that the pseudonym is legally assigned. If we can reduce the number of unused pseudonyms, the risk of security will be reduced as well;
- **Reducing Control Costs***:* Compared with the strategy that a cell request N pseudonyms every time, the strategy that a cell requests the expected number pseudonyms will include far fewer pseudonyms and, consequently, the overhead of generating, transmitting, and processing will be correspondingly smaller;
- **Reducing Maintenance Costs***:* We assume that pseudonyms cost money and that, while the pseudonym service is free to the public the municipality and/or other third party entities will have to pay for them. It is obvious, therefore, that request-

ing and maintaining a smaller number of pseudonyms, tailored to the expected number of vehicles in the cell, will be cheaper than maintaining the maximum number of pseudonyms.

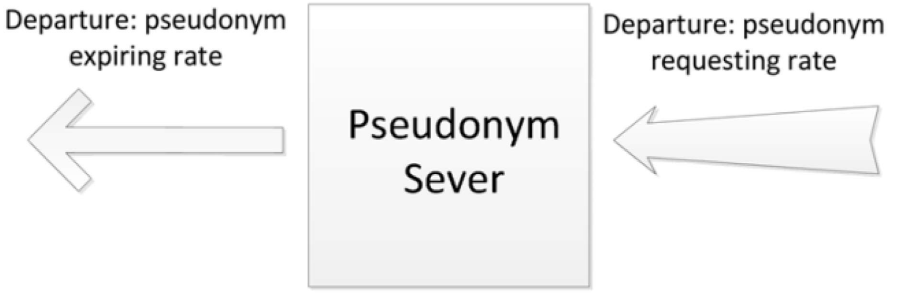
7.1 Defining the Model

In this section, we define the stochastic model which will be used to derive the analytical close form. We take stochastic process to the process of pseudonym application and expiration. The stochastic model is defined as follows, shown in Figure 8:

1. We are interested on the pseudonym resources on the pseudonym server side. Therefore, we focus on the pseudonym server. Cars in cells mostly and passively receive information such as traffic congestion information from infrastructure. Cars will not need pseudonyms at this stage. When a driver wants to send requests or communicate with infrastructure/other cars, a vehicle will reactively request a pseudonym and start to communicate. In other words, cars request pseudonym in an on-demand way. The service server is the whole process of registering, validating, and unregistering pseudonyms for vehicles.
2. The pseudonym server will assign a pseudonym to the vehicle *only after* a vehicle request a pseudonym. Therefore, the *arrival rate of the stochastic process is defined as arrival rate of demands for pseudonyms from vehicles in a cell*.
3. A pseudonym will become expired either the TTL decreases to zero or the pseudonym server is notified by other pseudonym servers. Therefore, the *departure rate of the stochastic process is defined as expiring rate of pseudonyms*.
4. The *service time in the stochastic process is the duration of registering and unregistering the pseudonyms*.
5. Vehicles can be in the incoming traffic and become brand-new vehicle in the cell or the vehicle that has been parked the cell for a long time and get restarted.

It is important to point out that there exists some special cases that the model does not apply. One example is that phenomenal events in a cell, such as super bowl of football, air show, etc., will create exceptional arrival rate and departure rate comparing with normal traffic condition in that cell. The traffic in cells are extremely different to the normal traffic condition in regular days. However, thanks to predetermination of these events (schedules and locations are predefined), pseudonym servers in involving cells can be

Figure 8. Stochastic model. The arrival rate: arrival rate of demands for pseudonyms from vehicles in a cell. The departure rate: expiring rate of pseudonyms. (Yan et al., 2013; Arif et al., 2012).


Departure: pseudonym expiring rate
Pseudonym Sever
Departure: pseudonym requesting rate

prepared for these phenomenal events. From the engineering perspective, we can request full capacity of pseudonyms and divide the involving cells into sub-cells or micro cells. Each of them has full capacity of pseudonyms. The cell's size can be decreased to a few hundred meters so that the server can feed each vehicle a pseudonym. Although it can be expensive, technically, it can be handleable.

7.2 Cell Size Analysis

We are interested in the following problem. Consider a cell with finite capacity *N*. At time *t*=0, the cell contains $n_0 \geq 0$ cars. After that, cars arrive and depart at time-dependent rates as described next. If the cell contains *k*, (0≤*k*≤*N*), cars at time *t*, then the car arrival rate is

$$a_k(t) = \frac{N-K}{N}\lambda(t) \tag{2}$$

and the car departure rate $\beta_k(t)$ is

$$\beta_k(t) = k\mu(t) \tag{3}$$

where for all *t*≥0, λ(*t*) and μ(*t*) are *integrable* on [0,*t*]. It is worth noting that both $a_k(t)$ and $\beta_k(t)$ are functions of both *t* and *k*. In particular, itmaywellbethecasethatfor $t_1 \neq t_2, a_k(t_1) \neq a_k(t_2)$ and similarly for $\beta_k(t_1)$ and $\beta_k(t_2)$, giving mathematical expression to the fact that at different times of the day, say, the departure rate depends not only on the number of cars present in the cell but also on time-dependent factors.

Consider the counting process {*X*(*t*) | *t*≥0} of continuous parameter *t*, where for every positive integer *k*, (1≤*k*≤*N*), the event {*X*(*t*)=*k*} occurs if the cell contains *k* cars at time *t*. We let $P_k(t)$ denote the probability that the event {*X*(*t*)=*k*} occurs. In other words,

$$P_k(t) = \Pr\left[\{X(t) = k\}\right]$$

In addition to $P_k(t)$, of interest are the expected number *E*[*X*(*t*)] and the variance *Var*[*X*(*t*)] of the number of cars in the cell at time *t*>0, as well as the limiting behavior of these parameters as $t \to \infty$, whenever such a limit exists and/or makes sense.

7.3 Deriving a Closed Form for $P_k(t)$

To make the mathematical derivations more manageable, we set $P_k(t) = 0$ for *k*<0 and *k*>*N*. Thus, $P_k(t)$ is well defined for all integers *k*∈(−∞,∞) and for all *t*≥0. In particular, the assumption about the cell containing n_0 cars at *t*=0 translates into $P_k(0) = 1$ if $k = n_0$ and 0 otherwise.

Let *t*, (*t*≥0), be arbitrary and let *h* be sufficiently small that in the time interval [*t*,*t*+*h*] the probability of two or more arrivals or departures, or of a simultaneous arrival and departure, is *o*(*h*). With *h* chosen as stated, the probability $P_k(t+h)$ that the cell contains *k*, (0≤*k*≤*N*), cars at time *t*+*h* has the following components:

$$P_k(t)\left[1 - h\frac{N-k}{N}\lambda(t) - kh\mu(t) + o(h)\right]$$

$$P_{k-1}(t)\left[1 - h\frac{N-k}{N}\lambda(t) - kh\mu(t) + o(h)\right]$$

$$P_{k+1}(t)\left[(k+1)h\mu(t) + o(h)\right]$$

Visibly: (See Box 1) with the initial condition $P_k(0) = 1$ for $k = n_0$ and 0 otherwise.

Let

$$G(z,t) = \sum_k P_k(t)z^k \tag{5}$$

Box 1.

$$P_k(t+h) = P_k(t)\left[1 - h\frac{N-k}{N}\lambda(t) - kh\mu(t)\right] + P_{k-1}(t)h\frac{N-k+1}{N}\lambda(t) + P_{k+1}(t)(k+1)h\mu(t) + o(h)$$

$$= P_k(t)\left[1 - h\frac{N-k}{N}\lambda(t)\right] + P_{k-1}(t)h\frac{N-k+1}{N}\lambda(t) + (k+1)h\mu(t) + o(h)$$

After transposing $P_k(t)$ and dividing by h we have:

$$\frac{P_k(t+h) - P_k(t)}{(t+h)-t} = \left[\frac{N-k}{N}\lambda(t) + k\mu(t)\right]P_k(t) + \frac{N-k+1}{N}\lambda(t)P_{k-1}(t) + P_{k+1}(t)\mu(t) + \frac{o(h)}{h}$$

Taking limits on both sides as $h \to 0$ yields the differential equation:

$$\frac{{}^0dP_k(t)}{{}^0dt} = -\left[\frac{N-k}{N}\lambda(t) + k\mu(t)\right]P_k(t) + \frac{N-k+1}{N}\lambda(t)P_{k-1}(t) + (k+1)\mu(t)P_{k+1}t \qquad (4)$$

be the probability generating function of $P_k(t)$. Recall that since $P_k = 0$ for $k<0$ and $k>N$, there is no harm working with $k\in(-\infty,\infty)$. Upon multiplying (4) by z^k and upon summing over $k\in(-\infty,\infty)$ we obtain (see Box 2) with $G(z,0) = z^n 0$ and auxiliary equations

$$\frac{{}^0dt}{1} = \frac{{}^0dz}{(z-1)\left[\frac{\lambda(t)}{N}z + \mu(t)\right]} = \frac{{}^0dG}{(z-1)\lambda(t)G} \qquad (7)$$

After the change of variable $z - 1 = \cdot\frac{1}{y}$, the differential equation

$$\frac{{}^0dt}{1} = \frac{{}^0dz}{(z-1)\left[\frac{\lambda(t)}{N}z + \mu(t)\right]}$$

becomes

$$\frac{{}^0dy}{{}^0dt} + \left[\frac{\lambda(t)}{N} + \mu(t)\right]y = -\frac{\lambda(t)}{N} \qquad (8)$$

Using standard techniques, equation (8) yields

$$e^{h(t)}y + \int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du = cons\tan t$$

or, equivalently,

$$\frac{e^{h(t)}}{z-1} + \int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du = c_1 \qquad (9)$$

where c_1 is an arbitrary constant and the function $h : \left[0,\infty\right) \to \left[0,\infty\right.$ is such that for all non-negative x,

$$h(x) = \int_0^x \left[\frac{\lambda(s)}{N} + \mu(s)\right]^0 ds \qquad (10)$$

For later reference, we now state and prove the following technical result.

Box 2.

$$\frac{\partial G(z,t)}{\partial t} = \sum_{k} \frac{dP_k(t)}{dt} =$$

$$-\sum_{k}\left|\frac{N-k}{N}\lambda(t) + k\mu(t)\right| P_k(t)z^k + \lambda(t)\sum_{k}\frac{N-k+1}{N}P_{k-1}(t)z^k + \mu(t)\sum_{k}(k+1)P_{k+1}(t)z^k$$

$$= -\lambda(t)\sum_{k}\cdot P_k(t)z^k - z\mu(t)\sum_{k}\cdot kP_k(t)z^{k-1} + z\lambda(t)\sum_{k}P_{k-1}(t)z^{k-1} + \mu(t)\sum_{k}(k+1)P_{k+1}(t)z^k$$

$$= -\lambda(t)G(z,t) + \left[\frac{\lambda(t)}{N} - \mu(t)\right] z\frac{\partial G(z,t)}{\partial z} + \lambda(t)zG(z,t) - \frac{\lambda(t)}{N}z^2\frac{\partial G(z,t)}{\partial z} + \mu(t)\frac{\partial G(z,t)}{\partial z}$$

$$-(z-1)\left[\frac{\lambda(t)}{N}z + \mu(t)\right]\frac{\partial G(z,t)}{\partial z} + \lambda(t)(z-1)G(z,t)$$

Thus, we have obtained the partial differential equation

$$\frac{\partial G(z,t)}{\partial t} + (z-1)\left[\frac{\lambda(t)}{N} + \mu(t)\right]\frac{\partial G(z,t)}{\partial z} = \lambda(t)(z-1)G(z,t) \qquad (6)$$

$$c_1 = -1 + \frac{z}{z-1}e^{h(t)} - \int_0^t \mu(u)e^{h(u)0}du \qquad (11)$$

By simple manipulations, (9) yields

$$c_1 = \frac{e^{h(t)}}{z-1} + \int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du$$

$$= \frac{e^{h(t)}}{z-1} + \int_0^t \left[\frac{\lambda(u)}{N} + \mu(u)\right]e^{h(u)0}du - \int_0^t \mu(u)e^{h(u)0}du$$

$$= \frac{e^{h(t)}}{z-1} + e^{h(t)} - 1 - \int_0^t \mu(u)e^{h(u)0}du$$

$$= -1 + \frac{z}{z-1}e^{h(t)} - \int_0^t \mu(u)e^{h(u)0}du$$

where we have used the fact that

$$\int_0^t \left[\frac{\lambda(u)}{N} + \mu(u)\right]e^{h(u)0}du = e^{h(t)} - 1$$

which is implied by the Fundamental Theorem of Calculus. Thus (11) holds, as claimed.

Returning to the auxiliary equations (7), we observe that by selecting the multiplicands x_1, x_2, x_3 as

$$x_1 = -(z-G)\left[\frac{\lambda(t)}{N} + \mu(t)\right]$$

$$x_2 = G$$

$$x_3 = -\frac{z-1}{N}$$

the ratio

$$R = \frac{x_1{}^0 dt + x_2{}^0 dz + x_3{}^0 dG}{x_1 + x_2(z - 1 _- \left[\frac{\lambda(t)}{N} z + \mu(t)\right] + x_3(z-1)\lambda(t)G}$$

$$= \frac{-\left[\frac{\lambda(t)}{N} + \mu(t)\right]^0 dt + \frac{{}^0 dz}{z-1} - \frac{{}^0 dG}{NG}}{0}$$

implying that

$$-\int_0^t \left[\frac{\lambda(u)}{N} + \mu(u)\right]^0 du + 1n(z-1) - 1nG\,\frac{1}{N} = Cons\tan t$$

which, in turn, yields

$$-ht + 1n\,\frac{z-1}{G(z,t)\,\frac{1}{n}} = cons\tan t$$

whereupon, by exponentiation, we obtain

$$\exp\left[-h(t) + 1n\,\frac{z-1}{G(z,t)\,\frac{1}{N}}\right] = \frac{z-1}{G(z,t)\,\frac{1}{N}}\,e^{-h(t)} \tag{12}$$

$$= c_2 \tag{13}$$

for some constannt. c_2 The two constants c_1 and c_2 are related by

$$c_2 = \psi\left[c_1\right] \tag{14}$$

where Ψ is an arbitrary function.

As it turns out, (12), (14), along with condition $G(z,0) = z^n 0$ can be used to determine Ψ. For this purpose, we first find an explicit closed form for Ψ. It is easy to confirm that for an arbitrary real *x*,

$$\psi\left[x\right] = \frac{1}{x}\left[\frac{x}{x+1}\right]^{\frac{n_0}{N}} \tag{15}$$

Now, (9), (11), (12) and (15), combined, allow us to write (see Box 8)

$$G(z,t) = (z-1)^N e^{-Nh(t)} (1 + c_1)^{n_0} c_1{}^{N-n_0}$$

$$= \left[e^{-h(t)}(z-1)(1+c_1)\right]^{n_0} \left[e^{-h(t)}(z-1)c_1\right]^{N-n_0}$$

$$= \left[z(1 - e^{-h(t)} \int_0^t \mu(u) e^{h(u)0} du) + e^{-h(t)} \int_0^t \mu(u) e^{h(u)0} du\right]^{n_0} \tag{16}$$

In spite of its complexity, (16) reveals a whole cell about the structure of the process $\{X(t) \mid t \geq 0\}$. To see this, observe that $G(z,t)$ is the product of the following two factors:

$$\left[z(1 - e^{-h(t)} \int_0^t \mu(u) e^{h(u)0} du) + e^{-ht} \int_0^t \mu(u) e^{h(u)0} du\right]^{n_0}$$

Box 8.

$$= \left[z(1 - e^{-h(t)}(1 + \int_0^t \mu(t) e^{h(u)0} du)) + e^{-h(t)}(1 + \int_0^t \mu(t) e^{h(u)0} du)\right]^{N-n_0} \tag{17}$$

which is the probability generating function of a binomial random variable with parameter n_0 and success probability

$$p(t) = 1 - e^{-h(t)}\int_0^t \mu(u)e^{h(u)0}du; \qquad (18)$$

$$\left[ze^{-h(t)}\int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du + (1 - e^{-h(t)}\int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du)\right]^{N-n_0}$$

which is the probability generating function of a binomial random variable with parameter $N - n_0$ and success probability

$$q(t) = e^{-h(t)}\int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du \qquad (19)$$

Define two additional counting processes

- {$R(t)$ | $t \geq 0$} that keeps track of the number of the n_0 cars present at time t=0 that are still in the cell at time t; it is clear that the success probability $p(t) = 1 - e^{-h(t)}\int_0^t \mu(u)e^{h(u)0}du$ is precisely the probability that a generic such car is still in the cell at time t;
- {$S(t)$ | $t \geq 0$} that keeps track of the number of cars in the cell at time t that were not in the cell at time t=0; this is also a binomial process with parameters $N - n_0$ and success probability

$$e^{-h(t)}\int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du$$

It is immediate that for all t, $R(t)$ and $S(t)$ are independent random variables. Further, the expression of $G(z,t)$ as a product implies that for all $t \geq 0$, $N(t)$ is the convolution of $R(t)$ and $R(t)$ and so

$$X(t) = R(t) + S(t). \qquad (20)$$

Next, we turn to the task of computing a closed form for the expected number, $E[X(t)]$, of cars in the cell at time t and its variance $Var[X(t)]$. Observe that by (20) and the linearity of expectation we can write the following equations (see Box 9-10).

$$E\left[X(t)\right] = E\left[R(t)\right] + E\left[S(t)\right]$$

$$= n_0 e^{-h(t)} + e^{-h(t)}\int_0^t \lambda(u)e^{h(u)0}du$$

$$= e^{-h(t)}\left[n_0 + \int_0^t \lambda(u)e^{h(u)0}du\right] \qquad (21)$$

Similarly, since as noted for every t>0, the random variables $R(t)$ and $S(t)$ are independent, and thus, uncorrelated, we can write

$$Var\left[X(t)\right] = Var\left[R(t)\right] + Var\left[S(t)\right]$$

$$= n_0 p(t)\left[1 - p(t)\right] + \Lambda(t)$$

$$= n_0 p(t)\left[1 - p(t)\right] + \frac{\int_0^1 \lambda(u)e\int_0^u \mu(s)^0 ds^0 du}{e\int_o^t \mu(u)^0 du}$$

$$= p(t)\left[n_0\left[1 - p(t)\right] + \int_0^t \lambda(u)e\int_0^u \mu(s)^0 ds^0 du\right] \qquad (22)$$

Box 9.

$$= n_0(1 - e^{-h(t)}\int_0^t \mu(u)e^{h(u)0}du) + (N - n_0)e^{-h(t)}\int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du$$

Box 10.

$$n_0 + Ne^{-h(t)}\int_0^t \frac{\lambda(u)}{N}e^{h(u)0}du - n_0e^{-h(t)}\int_0^t \left[\frac{\lambda(u)}{N} + \mu(u)\right]e^{h(u)0}du$$

7.4 Pseudonym Collision

Although we assume that the PSs and sub-PSs are generally synchronized, there may occasionally have situations (e.g. network congestion) that will cause the delays of updating pseudonyms. Thereafter, there will have a certain probability that the pseudonyms that are collided. We are interested to check the probability of the collision of pseudonyms in the worst cases. In each cell, the total number of pseudonyms is *n* in a pool. Each vehicle is given *k* pseudonyms from the cell CA. Given a vehicle *a*, define event *c* as another randomly selected vehicle *b* does not adopt any pseudonyms that the vehicle *a* does, i.e. no two vehicles share the same set of pseudonyms. We are interested in the probability *P*(*c*).

$$P(c) = \frac{C_{n-k}^{k}}{C_{k}^{n}}$$

$$= \frac{(n-k)!^2}{n!(n-2k)!}$$

According to Stirling's approximation,

$$n! \approx \sqrt{2\pi(n^{n+0.5}e^{-n})}$$

We read

$$P(c) \approx \frac{(n-k)^{2n-2k+1}e^{-2n+2k}}{n^{n+0.5}e^{-n}(n-2k)^{n-2k+0.5}e^{-n+2k}}$$
$$= \frac{(1-\frac{k}{n})^{2(n-k+0.5)}}{(1-\frac{2k}{n})^{n-2k+0.5}}$$

When $x \to 0,\ 1 - x \approx e^{-x}$. Since *n*≫*k*, we write

$$= \frac{(1-\frac{k}{n})^{2(n-k+0.5)}}{(1-\frac{2k}{n})^{n-2k+0.5}}$$

8. SIMULATION RESULTS

In this section we evaluate the analytical results derived in Section 7.2 by comparing the theoretical predictions with numerical results. The numerical results were obtained by mathematically simulating the cell. We assume the capacity of the cell is fairly large (e.g., 1000), even though it is considered unbounded in the analytical derivations.

8.1 Simulation Setup

We assumed that, at the beginning of the simulation, i.e., at time t=0, there were $n_0 = 500$ vehicles in the cell. The vehicles were assumed to arrive into and and depart from the cell at certain time-varying rates. To evaluate the analytical results, we considered three scenarios, as described next:

- **Scenario 1:** the first set of results was designed for *constant* arrival rate and departure rates. For illustration purposes, we have chosen λ=800 and μ=2;
- **Scenario 2:** the second set of results was designed for time-dependent arrival and departure rates but such that limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ does not exist;
- **Scenario 3:** the third set of results was designed for time-dependent arrival and departure rates such that limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ exists.

To set the stage for explaining our design decisions, imagine a typical long-term cell of a mid-size cell. A glance at the flight arrivals and departure schedule will convince us, first, that the flight arrival and departure rates are time-dependent stochastic phenomena; and, second, that flights depart and arrive on a 24-hour periodic schedule. It is, consequently, clear the car arrival and departure rates from the cell(s) will mirror fairly closely the flight departure and arrival rates. It follows that the car arrival and departure rates into/from the long-term cell and departure rates should also periodic functions of time.

While many periodic functions could possibly be employed, we have decided to adopt as generic arrival and departure rates

$$\lambda(t)=a+b\sin\Theta(t) \tag{23}$$

and

$$\mu(t)=c+d\sin\Theta(t). \tag{24}$$

where a, b, c, d are constants. Observe that b and d control the fluctuation of the arrival and departure rates, respectively. Indeed, from (23) it is clear that the larger b, the larger the fluctuation of λ(t) as a function of time. Similarly for d. Having settled on this choice, there were two further problems that needed attention. The first was the most appropriate simulation granularity: after some trials we have decided that the most appropriate time unit model the car arrival and departure is one hour. With this is mind, for arbitrary t≥0, we have decided to take. $\Theta(t) = \frac{\pi t}{12}$

Assuming the average the number of pseudonyms of a mid-size cell to be between 400 and 500, we have chosen the constants a=1300,b=500,c=3,d=1. With this in mind, the arrival and departure rates used in Scenario 2 were

$$\lambda(t) = 1300 + 5\sin(\frac{\pi t}{12})$$

and the departure-rate of vehicles is

$$\mu(t) = 3 + 1\sin(\frac{\pi t}{12})$$

It is easy to confirm that, in this case, the limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ does not exist.

For Scenario 3, we adopted quasi-periodic time-dependent arrival and departure rates λ(t) and μ(t), with a period close to 24 hours, in such a way that the limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ existed. Again many quasi-periodic functions could possibly be employed. But we have decided to adopt as generic arrival and departure rates as

$$\lambda(t) = 800 + 400\left[1 + 2\exp(-0.3t\right]\sin(\frac{\pi t}{12})$$

and

$$\mu(t) = 2 + \left[1 + \exp(-0.2t)\right]\sin(\frac{\pi t}{12})$$

8.2 Detailed Discussion of Simulation Results

We begin by investigating the expected number *E*[*N*(*t*)] of cars in the cell. The comparisons between the analytical results and the numerical results have been performed for the above three scenarios. For each scenario, we set the simulation time to 60 hours and we computed the expected number of existing vehicles at time *t*. Figure 9(a) shows *E*[*N*(*t*)] plotted against time in Scenario 1 with arrival rate λ=800 and the departure rate μ=2. We notice that *E*[*N*(*t*)] stabilizes at $\frac{\lambda}{\mu} = 400$. In fact, even though the initial number of vehicles was 500, *E*[*N*(*t*)] dropped sharply to 400 to within 3 hours into the simulation. This is because the vehicles initially existing have all left the cell in the first 3 hours. The cell becomes stabilized after that, due chiefly to the constant arrival and departure rates.

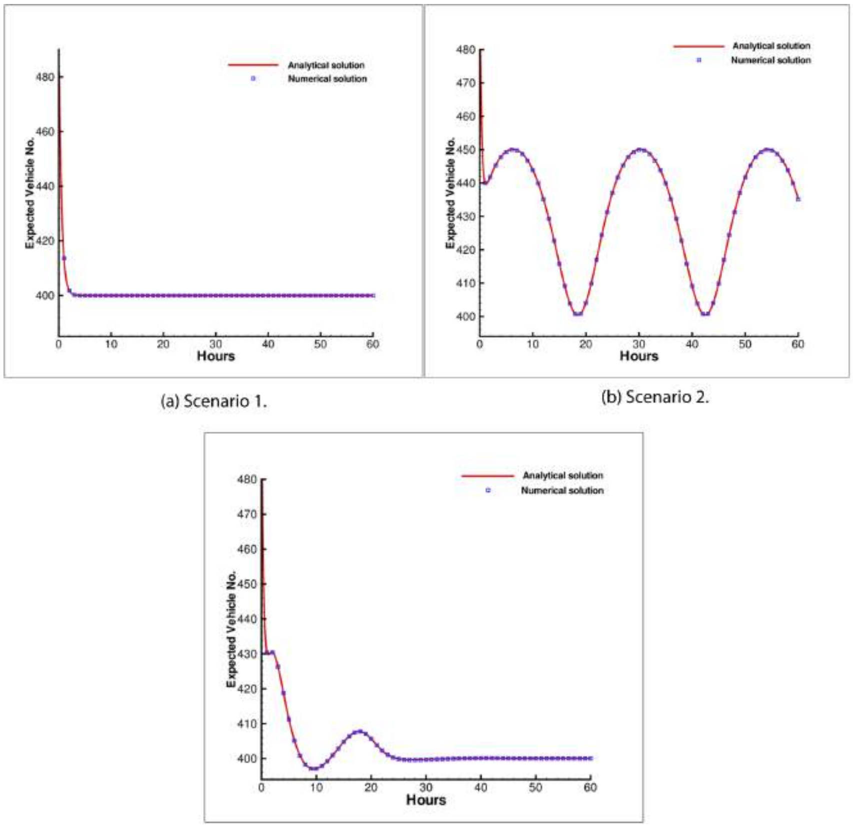
The effect of the initial conditions on *E*[*N*(*t*)] is similar in Scenarios 2 and 3: due to the departure of the initially existing vehicles, the expected number of existing vehicles is dropping sharply as illustrated in Figures 9(b) and 9(c). However, once the effect of the initial conditions has worn off, Scenarios 2 and 3 are vastly different. As expected, in Scenario 2 we see a periodic fluctuation of the expected number of existing vehicles as shown in Figure 9(b). In addition, Figure 9(b) clearly shows, as predicted by our analytical results, that *E*[*N*(*t*)] is bounded by 400 and 450 after *t*>3. In the case of Scenario 3, Figure 9(c) shows, as expected, that in the long-run *E*[*N*(*t*)] settles down to a constant value near the limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$. Before stabilization, fluctuating values are shown but fluctuation becomes weaker as *t* increases. The duration of the fluctuation actually depends on the exponential parameter. The exponential component of arrival rate in this result is exp(−0.3*t*) and the one of departure rate is exp(−0.2*t*). The bigger values of the exponential components are, the faster the system stabilizes.

Next, we turned our attention to evaluating *Var*[*N*(*t*)] versus time. Three sets of comparisons (corresponding to the three scenarios discussed in Subsection 8.1) between the analytical results and the numerical results have been performed. For each of them, the simulation time *t* was 60 hours and, as before, the number of the initial existing vehicles is $n_0 = 500$. Figure 10(a) shows that in Scenario 1, *Var*[*N*(*t*)] stabilizes to a constant value after fluctuation in the first few hours. Fluctuation in the first few hours is caused by the departure of the initially existing vehicles, explained earlier. As expected, the variance of the expected number of existing vehicles settles down to $\frac{\lambda}{\mu}$, i.e., 400 when *t*>3.

By contrast, Figure 10(b) captures the behavior of *Var*[*N*(*t*)] in Scenario 2 where the limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ does not exist. Just as predicted by our analytical derivations, Figure 10(b) shows that *Var*[*N*(*t*)] fluctuates with a period of 24 hours. In addition, a noticeable variance range [400, 450] can be read from Figure 10(b).

The situation is vastly different in Scenario 3 as illustrated in Figure 10(c). Here, *Var*[*N*(*t*)] stabilizes at 400 after *t*>25 hours of simulation.

Figure 9. The expected the number of pseudonyms vs time.(Yan et al., 2013; Arif et al., 2012).



(a) Scenario 1.

(b) Scenario 2.

(c) Scenario 3.

The probability of the departure of initially existing vehicles and the exponential component in the arrival rate and departure rate can affect the pattern of the unstable fluctuation.

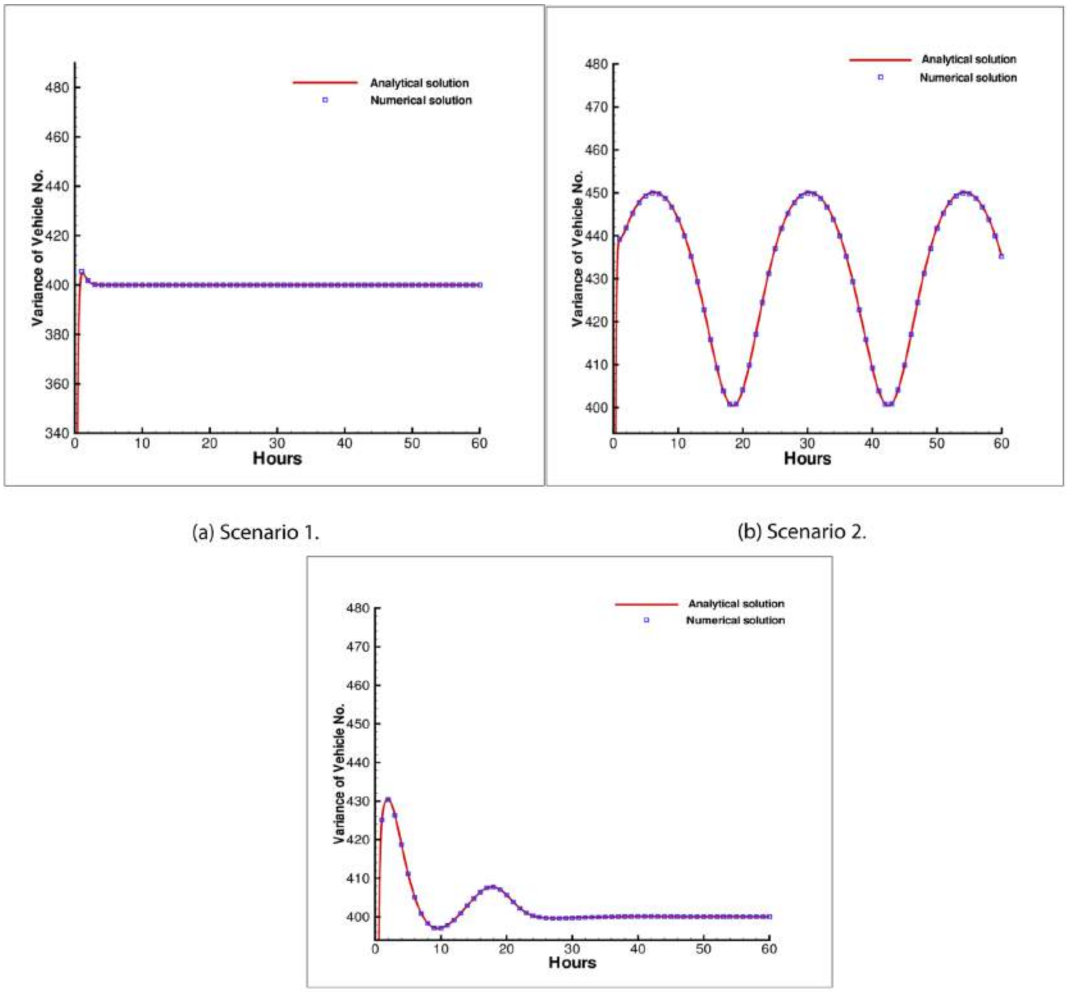
It is observed that the variance is stabilized at 400 for both Figure 10(a) and 10(c) and at a range [400, 450] in Figure 10(b). These values are rather large, considering that the expected number of existing vehicles $E[N(t)]$ is also 400. Therefore, we need to show the in spite of the large variance of $N(t)$, the probability that the number of vehicles in the cell is low is negligeable. We have decided to investigate, experimentally, $Pr[\{N(t)<400\}]$ and $Pr[\{N(t)<300\}]$. Three sets of comparisons (one for each scenario) have been performed. For each set of comparison, we adopt the same simulation settings as the previous ones and compute the probability values. Figure 11(a) shows the probability $Pr[\{N(t)<400\}]$ in the case of Scenario 1. The probability $Pr[\{N(t)<400\}]$ tends to 0.5 and $Pr[\{N(t)<300\}]$ tends to be 0 after an initial unstable fluctuation induced by the initial conditions. In Scenario 2, where both arrival rate and departure rates are periodical function of time and the limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ does not exist, we expect to see the periodical probability values. Figure 11(b) shows, as expected, that the probability $Pr[\{N(t)<400\}]$ is periodical and bounded by [0, 0.5]. The probability $Pr[\{N(t)<300\}]$ remains close to 0.

In the case of Scenario 3, Figure 11(c) shows that, as expected, the probability $Pr[\{N(t)<400\}]$ stabilizes at 0.5 after a certain fluctuation due mostly to the effect of initial conditions. The duration of the fluctuation is controlled by the initially existing vehicles and by exponential components in the arrival and departure rates.

In summary, the probability of the event that there are at least 300 vehicles in the cell is 100%. In other words, it is guaranteed that there are at least 300 vehicles existing in the cell at any time for our utilization.

To test the clustering of the probability mass around *E*[*N*(*t*)], we have performed a three-dimensional plots of $P_k(t)$, i.e. *Pr*[{*N*(*t*)<*k*}], time *t* and the number of existing vehicles *k* versus time. The goal was to find a direct view of relationships of the probability $P_k(t)$ at time *t* that there are *k* vehicles existing in the cell. We varied both time *t* and the number *k* of existing vehicles and calculate the probability values. Two cases have been investigated: Scenario 2 and Scenario 3. Figure 12 shows a shaded surface and as a contour plot of both cases. It clearly shows that in Scenario 2, the probability varies periodically with time *t* and *k*. By contrast, Figure 12(b) shows, as expected, that in Scenario 3 the probability eventually stabilizes after some initial fluctuation. To see better these trends, we also plotted the logarithm of the probability at various times *t* and

Figure 10. The variance of the the number of pseudonyms vs time. (Yan et al., 2013; Arif et al., 2012).



(a) Scenario 1.

(b) Scenario 2.

(c) Scenario 3.

the number of existing vehicles k. Figure 13(a) shows these results in the case of Scenario 2, while Figure 13(b) shows the logarithm of the probability at various times in the case of Scenario 3.

We are interested of the expected number of pseudonyms that the cell server will receive. The numerical results were obtained by mathematically simulating the cell. We assume the capacity of the cell is fairly large (e.g., 4000).We assumed that, at the beginning of the simulation, i.e., at time t=0, there were $n_0 = 500$ vehicles in the cell. The vehicles were assumed to arrive into and depart from the cell at certain time-varying rates. To evaluate the analytical results, we considered the following scenario: normal traffic conditions are designed with time-dependent arrival and departure rates but the limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ does not exist.

To set the stage for explaining our design decisions, imagine a typical cell of a city. Obviously, traffic arrival and departure rates are time-dependent stochastic phenomena; and, second, the vehicles depart and arrive on a week periodic schedule for Scenario 1 and 2 according to the statistic data (Federal Highway Administration, 2012).

With this in mind, the arrival and departure rates used in this simulation scenario were

$$\lambda(t)=347.19+80.90*\sin(0.8913t-2.8913)$$

and

$$\mu(t)=50.48+0.25\sin(0.8913t-2.8913).$$

It is easy to confirm that, in this case, the limit $\lim_{t \to \infty} \frac{\lambda(t)}{\mu(t)}$ does not exist.

8.3 Detailed Discussion of Simulation Results

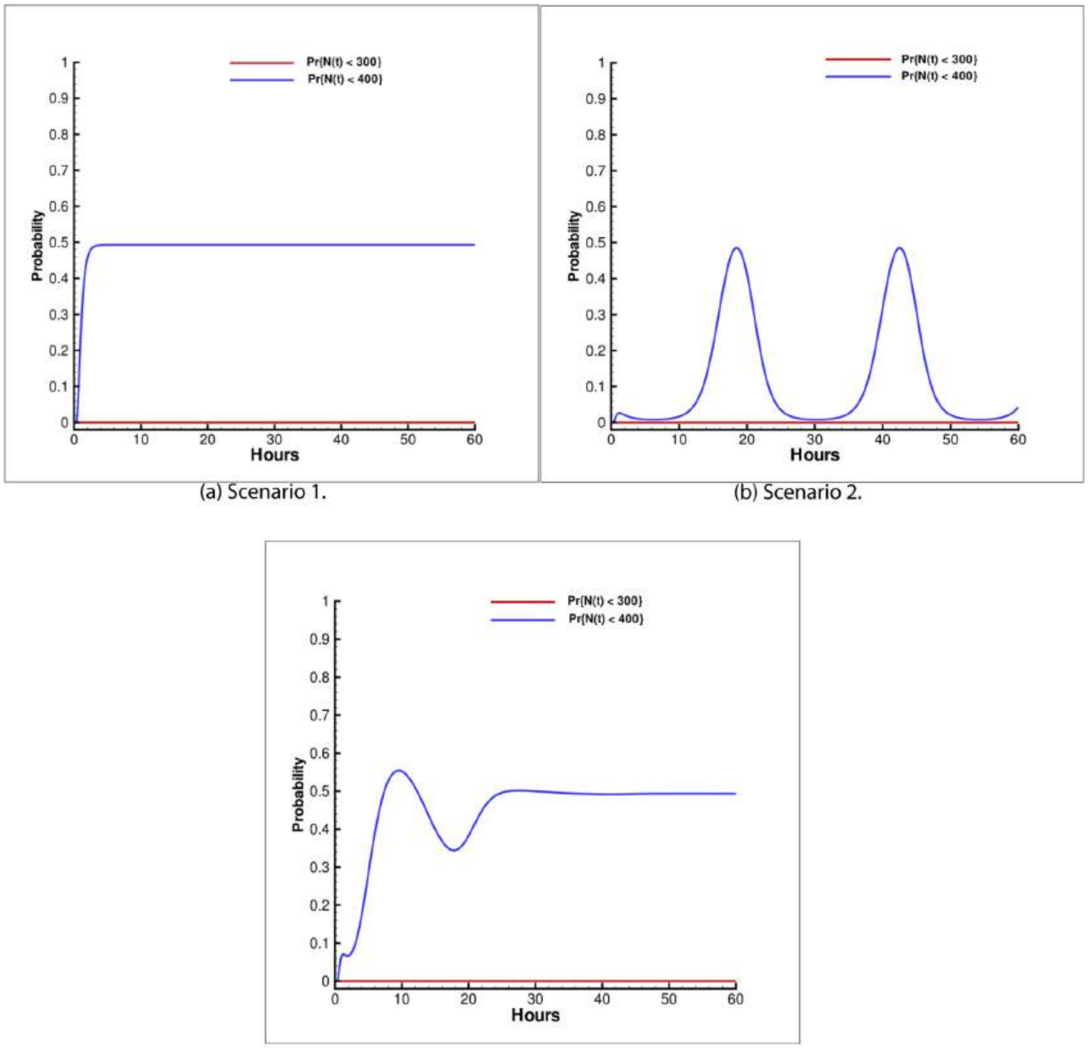
The value of the initial conditions on P_j is small: the arrivals of the initially pseudonym requests will need a while to build up, as illustrated in Figures 14 and 15. However, once the effect of the initial conditions has worn off, simulation 1, 2 and 3 are stabilized. As expected, in simulations we see a periodic fluctuation of the probability P_j of pseudonym requests as shown in Figure 14, 15, and 16. In addition, Figure 14 clearly shows that P_j is bounded by 0 and 0.12. In the case of Scenario 2, Figure 15 shows, as expected, that the value of j significantly affect the value of P_j. Figure 16, compared with Figure 15, shows that the value of N does not significantly affect the value of P_j.

In addition, we were interested to investigate the relationship among three variables: N, t, and P_j. We varied the value of N from 10 to 500 and the value of t from 0 to 24 which stands for one day. The result, shown in Figure 17, clearly presents periodic fluctuation and stabilized the vales of P_j.

8.4 Network Simulation

We were interested to investigate the network performance of our proposed method. We first applied SUMO (Krajzewicz et al., 2002) to generate a mobility trace file and then fed the trace file to NS-2 (ns-, 2001) where the corresponding wireless network was simulated. We chose SUMO and NS-2.30 not only because they are publicly available, but also because they are both well maintained and well accepted in the research community. We assumed a 700m x 700m area of city streets to represent a cell, shown in Figure 18. The pseudonym server is placed at the center (350m, 350m) of the cell. Vehicles entered the

Figure 11. The probability Pr[N(t)].(Yan et al., 2013; Arif et al., 2012).

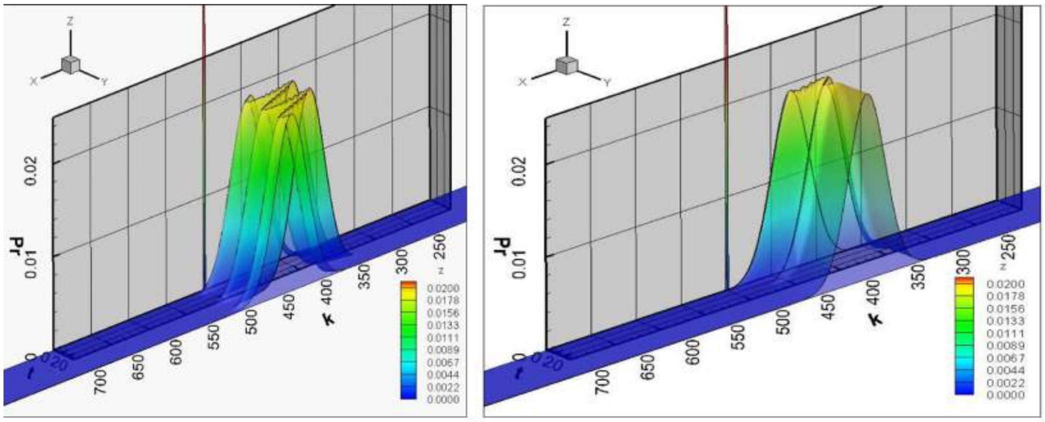


(a) Scenario 1.

(b) Scenario 2.

(c) Scenario 3.

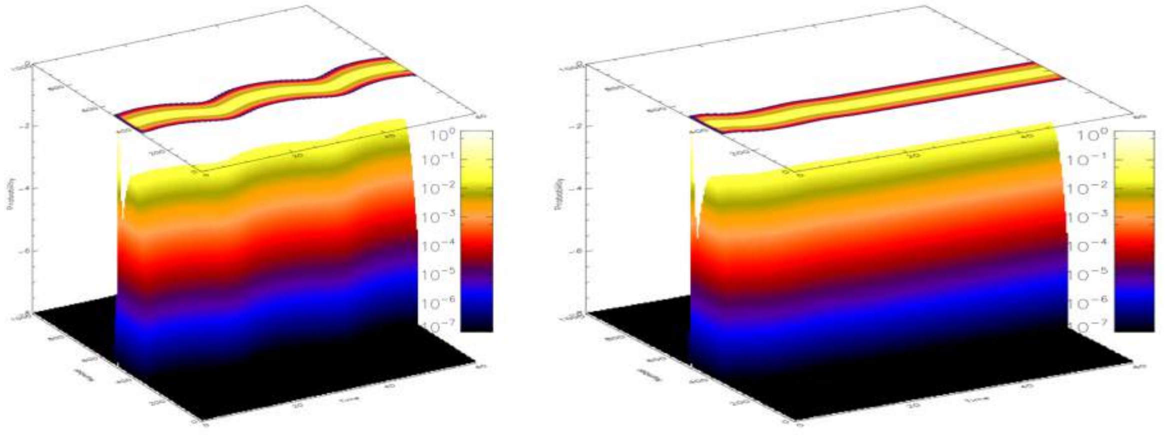
Figure 12. The number of existing vehicles k vs time t vs Pr[N(t)]. (Yan et al., 2013; Arif et al., 2012).



(a) Scenario 2.

(b) Scenario 3.

Figure 13. The number of existing vehicles k vs time t vs lnPr[N(t)].(Yan et al., 2013; Arif et al., 2012).



(a) Scenario 2.

(b) Scenario 3.

cell from the border streets and then randomly moved on streets in the cell.

We applied SUMO random traffic generator to choose a path between source and destination. We initially placed 150 vehicles. Vehicles make random turning decisions at each intersection. The speed limits on the streets range from 5 to 20 m/s (11-45 mile per hour), and the vehicles are constrained by the speed limit. Traffic lights are randomly simulated as well. At time *t*, a street can be green light or red. SUMO generates a mobility trace file that can be imported into NS-2. Then, NS-2 is executed and nodes in NS-2 follow the nodes in SUMO respectively. For the car following model, we used the Krauβ Model. Out of the initial 150 vehicles in the simulation, some of them are chosen (at random) as marked cars which will send pseudonym request randomly. The pseudonym server served as a data sink, i.e. service provider. There are 31 traffic flows, each corresponding to a street. Each traffic flow sends UDP packets (512 bytes for each packet).

We compared two scenarios. The first scenario sets the transmission range (TR) at 350m for each car, while the second scenario sets the transmission range as 700m for each car. The antenna height, the CSThresh and RXThresh values in NS-2 can be configured to determine the communication range. Packets are routed from source to destination if there is no direct route. Each vehicle will buffer packets (in a finite queue) until a route has been found to the destination.

8.4.1 The Macroscopic Perspective

The average throughput of each street requests is of interest in network simulations. We varied the transmission range (TR) in three scenarios: 233 meters, 350 meters and 700 meters. For each scenario, we collected the throughput of each traffic flow and then computed the average throughput. Each traffic flow stands for an individual street. The result is shown in Figure 19. As expected, the throughput value of 700m TR is about 50% higher than for a TR of 350m. This is because cars in the scenario two can directly communicate with the pseudonym server but the cars in the scenario one will need to relay request to the pseudonym server when the direct connection is unavailable. It is interesting to notice that the throughput of 233m TR is similar to the one of 350m TR. This is because as long as relay is needed, two-hop communication does not make significant difference to three-hop communication. According

Figure 14. Result of simulation 1.(Yan et al., 2013; Arif et al., 2012).

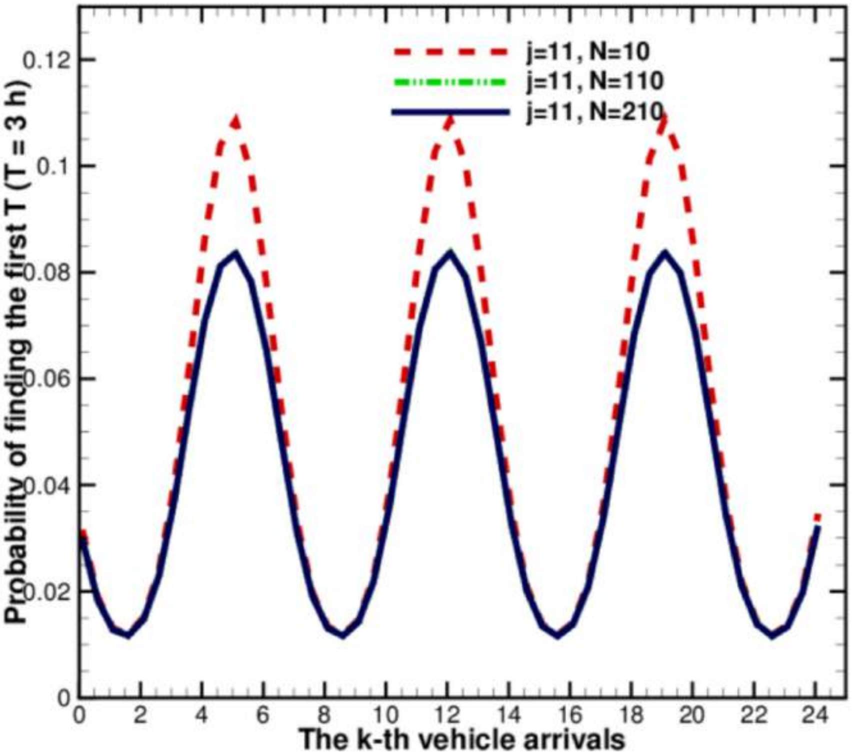


Figure 15. Result of simulation 2.(Yan et al., 2013; Arif et al., 2012).

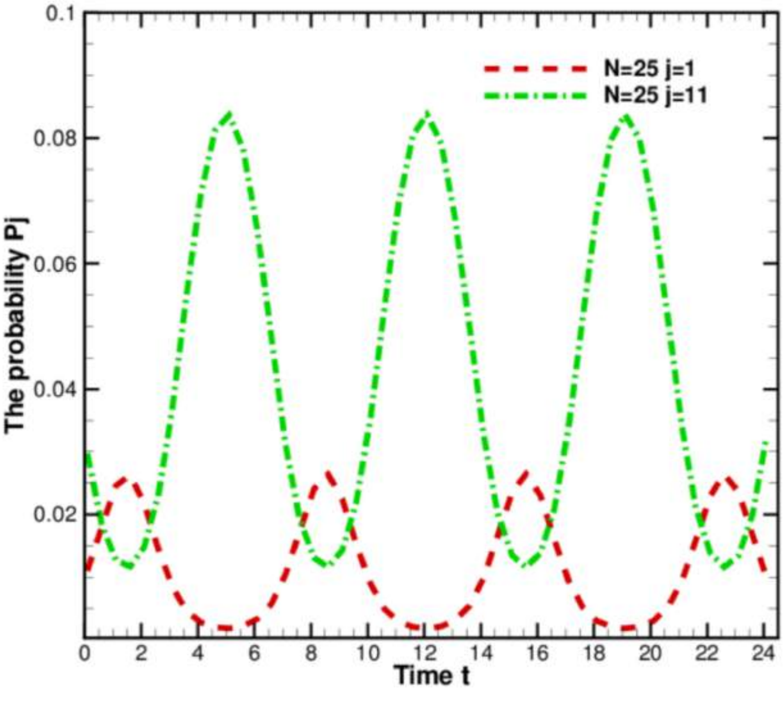


Figure 16. Result of simulation 3.(Yan et al., 2013; Arif et al., 2012).

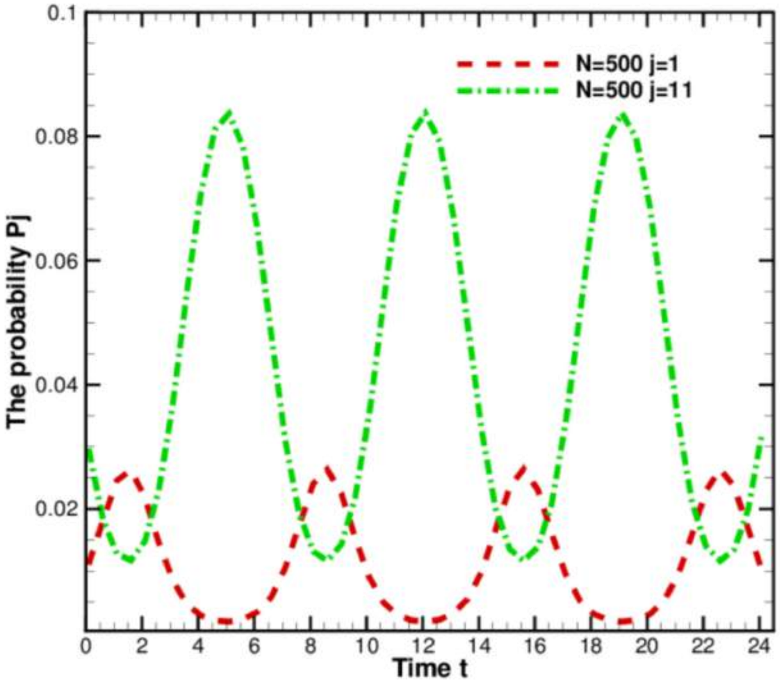


Figure 17. Result of simulation 4.(Yan et al., 2013; Arif et al., 2012).

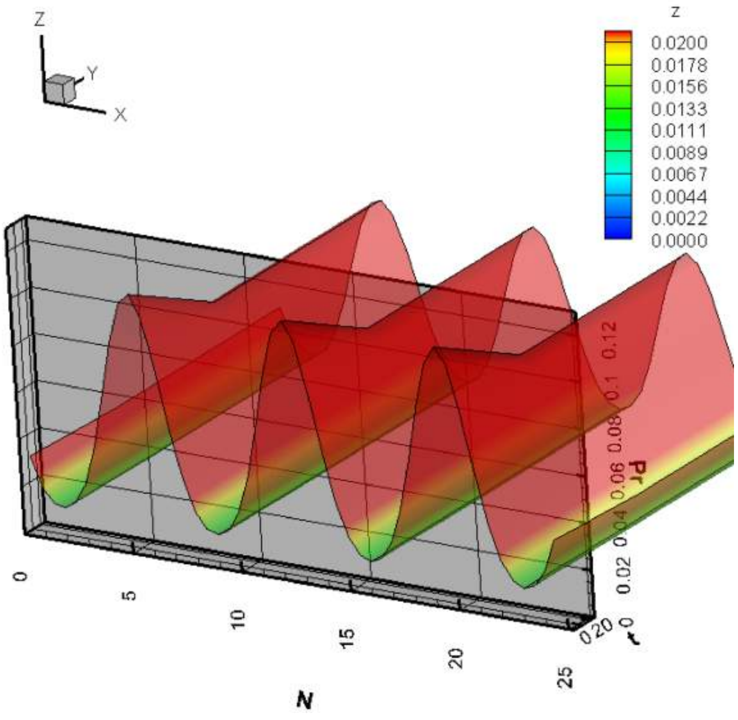
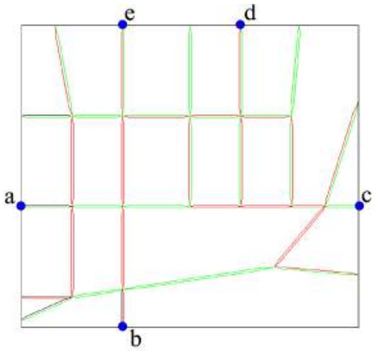


Figure 18. Illustrating our assumed map topology. The points (a,b,c,d,e) are the initial major entries of traffic. The green and red colored edges show the traffic lights at time t.(Yan et al., 2013; Arif et al., 2012).



to Figure 19, the peak value of request, about 1.1 request per second, i.e. 3960 request per hour, closely matches the 4000 peak value.

We then computed the average loss rate of requests for every street and show the results in Figure 20. Figure 20 shows that the drop rate in the scenario one *TR*=350*m* fluctuates from 0.2 to 0.9. This high loss rate in the worst case is because vehicles moving to the borders of the map need intermediate vehicle to relay packets and the relay vehicles are not always available. Once vehicles move to the center of the map, more routing paths are available, and the loss rate begins to decrease. Figure 20 shows that the case *TR*=700*m* is better than the case *TR*=350*m* and the case *TR*=233*m* in terms of drop rate.

We were also interested in the request response delays. The packet response delays of each flow were collected and computed. The result is shown in Figure 21. As expected, the delay values of requests of the case *TR*=233*m* and the case *TR*=350*m* are slightly larger than the case *TR*=700*m*. The reason is obviously because of the unreliable vehicular networks communication. The more hops in communication, the bigger delay values will be. In our proposed scheme, our assumption that the communication can directly reach to the cell shows both theoretical value and empirical meaning in this simulation.

8.4.2 The Microscopic Perspective

We also presented results from a microscopic perspective. One street was randomly selected to display more detailed network communication information. We investigated the jitter of requests in the selected street. The definition of jitter is as follows:

1. Jitter 1: $jitter(i+1)=jitter(i)+[|(R(i+1)-S(i+1))-(R(i)-S(i))|-jitter(i)]/16$
2. Jitter 2: $jitter(i+1)=jitter(i)+[|(R(i+1)-R(i))-(R(i)-R(i-1))|-jitter(i)]/16$

where *jitter*(*i*) is the jitter value of packet '*i*'; *S*(*i*) is the time at which packet '*i*' was transmitted from the sender; *R*(*i*) is the time at which packet '*i*' was received by the destination. The results of both Jitter 1 and 2 were collected and computed, as shown in Figures 22(a) and 22(b). As expected, the jitter values (both Jitter 1 and 2) shows that the jitter in scenarios *TR*=233*m* and *TR*=350*m* has a significantly larger amplitude and fluctuation than the one in scenario *TR*=700*m*. The reason lies in the mobility of vehicles. It is fairly interesting to note that the jitter values are higher at the middle of the day and the end of the day. For middle of day, more cars are on street and the wireless channels become more crowded and more likely to collide. So the jitter values increase. Towards the end of the day, the population of vehicles is greatly decreased. Vehicles in scenario one could fail to connect the pseudonym server because no intermediate cars can be used as communication relay nodes. Comparing Figure 22(a) and Figure 22(b), we note that jitter values will be different if the jitter is defined differently.

Figure 19. Throughput of server.(Yan et al., 2013; Arif et al., 2012).

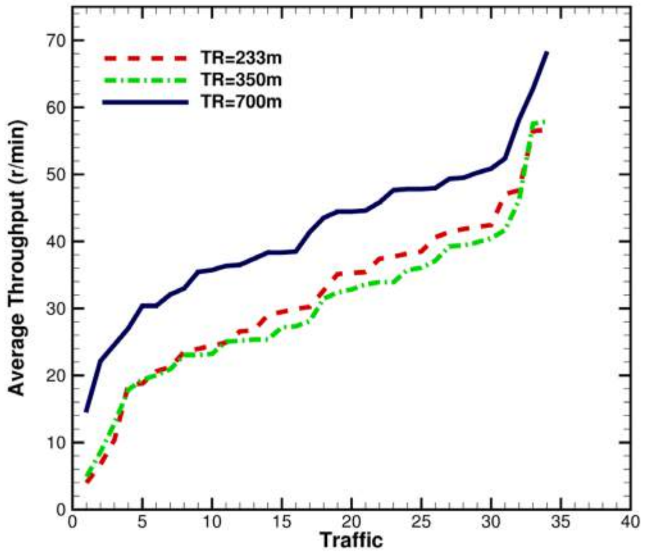
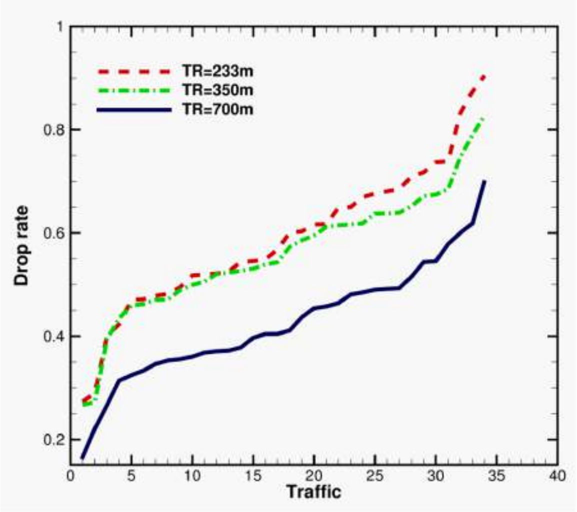


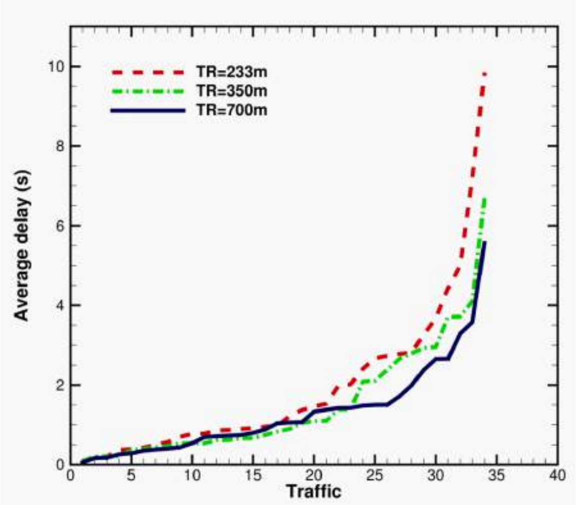
Figure 20. Drop rate of packets.(Yan et al., 2013; Arif et al., 2012).



9. CONCLUDING REMARKS AND DIRECTIONS FOR FUTURE WORK

To accommodate increasing demand from the driving public, car and truck manufacturers are offering more and more sophisticated on-board devices, including powerful computers, a large array of sensors, radar devices, cameras, and wireless transceivers. The powerful on-board devices support new applications, including location-specific services, on-line gaming, delivering multimedia content and various forms of mobile infotainment made possible by the emergence, in the past decade, of vehicular networks. However,

Figure 21. Delay of requests.(Yan et al., 2013; Arif et al., 2012).

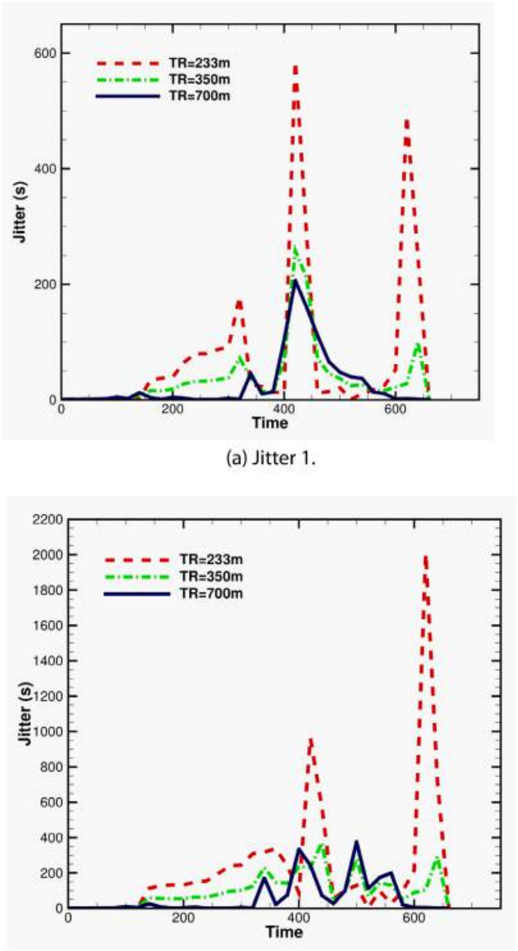


the increased Internet presence that enables the above applications invites various forms of privacy attacks. Invariably, these privacy attacks exploit the various forms of correlation that exist between the identity of a vehicle and that of its driver.

Virtually all published papers in the recent literature ignore the issues of scalability and robustness in the context of privacy protection. In this work we took a non-trivial step towards providing a robust and scalable solution to privacy protection in vehicular networks. To promote scalability and robustness we employ two strategies. First, we viewed vehicular networks as consisting of non-overlapping subnetworks each local to a geographic area referred to as a cell. Second, instead of issuing pseudonyms to vehicles proactively (as virtually all existing schemes do) we issue pseudonyms only to those vehicles that request them. This strategy is suggested by the fact that, in a typical scenario, only a fraction of the vehicles in an area will engage in communication with other vehicles and/or with the infrastructure and, therefore, do not need pseudonyms. Our second main contribution was to model analytically the time-varying request for pseudonyms in a given cell. This is important for capacity planning purposes since it allows managers to predict the probability that a given number of pseudonyms will be required at a certain time as well as the expected number of pseudonyms in use in a cell at a certain time. Empirical results obtained by detailed simulation confirmed the accuracy of our analytical predictions.

It is important to point out that there are some special cases where our model does not apply. One example is that of special events in a cell, such as super-bowl of football game, air show, etc., will create exceptional arrival rates and departure rates when compared to normal traffic conditions in the cell. During these special events, the traffic patterns in the cell are very different from normal traffic conditions. However, since these events are planned for carefully and well ahead of time the pseudonym server can be prepared for the extra load imposed by these events. From an engineering perspective, we can request full capacity of pseudonyms and divide the involving cells into sub-cells or microcells. Each of them has full capacity of pseudonyms. The cell size can be decreased to a few hundred meters so that the server can feed each vehicle a pseudonym. Although it can be expensive, technically, it can be handled. However, while interesting and challenging in its own right, this aspect is well beyond the scope of the chapter and will be looked at in future work.

Figure 22. Jitters from a street.(Yan et al., 2013; Arif et al., 2012).



(a) Jitter 1.

(b) Jitter 2.

As future work, we propose to combine the proposed privacy scheme with security and communication systems. In addition, the analytical model can be extended to more practical cases. For example some events such as football game will cause a large amount of vehicles parked in a cell and they will need to request pseudonyms at the same time when the game is over. Another example, vehicle accidents often result traffic arrival rate change. We will need to study how to work with the sudden traffic events.

REFERENCES

Arapinis, M., Chothia, T., Ritter, E., & Ryan, M. (2010). Analysing unlinkability and anonymity using the applied pi calculus. In *Proceedings of CSF*, (pp. 107–121). CSF.

Arif, S., Olariu, S., Wang, J., Yan, G., Yang, W., & Khalil, I. (2012). Datacenter at the airport: Reasoning about time-dependent parking lot occupancy. *IEEE Transactions on Parallel and Distributed Systems*, 99.

Blanchet, B., Abadi, M., & Fournet, C. (2008). Automated verification of selected equivalences for security protocols. *Journal of Logic and Algebraic Programming*, *75*(1), 3–51. doi:10.1016/j.jlap.2007.06.002.

Brusò, M., Chatzikokolakis, K., & den Hartog, J. (2010). Formal verification of privacy for rfid systems. In *Proceedings of CSF*, (pp. 75–88). CSF.

Choi, J. Y., Golle, P., & Jakobsson, M. (2006). Tamper-evident digital signatures: Protecting certification authorities against malware. In *Proceedings of the IEEE International Symposium on Dependable, Autonomic and Secure Computing (DASC)*, (pp. 37–44). IEEE.

Dahl, M., Delaune, S., & Steel, G. (2010). Formal analysis of privacy for vehicular mix-zones. In *Proceedings of the 15th European Conference on Research in Computer Security* (ESORICS'10), (pp. 55–70). ESORICS.

Delaune, S., Kremer, S., & Ryan, M. (2010). Verifying privacy-type properties of electronic voting protocols: A taster. In *Towards Trustworthy Elections* (pp. 289–309). Academic Press. doi:10.1007/978-3-642-12980-3_18.

Dok, H., Fu, H., Echevarria, R., & Weerasinghe, H. (2010). Privacy issues of vehicular ad-hoc networks. *International Journal of Future Generation Communication and Networking*, *3*(1), 17–32.

Federal Highway Administration. (2012). *Traffic congestion and reliability: Trends and advanced strategies for congestion mitigation*. Retrieved from http://www.ops.fhwa.dot.gov/congestion_report/chapter2.htm

Guo, J., Baugh, J. P., & Wang, S. (2007). A group signature based secure and privacy-preserving vehicular communication framework. In *Proceedings of the 2007 Mobile Networking for Vehicular Environments*. IEEE.

Huang, D., Misra, S., Xue, G., & Verma, M. (2011). Pacp: An efficient pseudonymous authentication based conditional privacy protocol for vanets. *IEEE Transactions on Intelligent Transportations*, *12*(3), 736–746. doi:10.1109/TITS.2011.2156790.

Hubaux, J.-P., Capkun, S., & Luo, J. (2004). The security and privacy of smart vehicles. *IEEE Security and Privacy Magazine*, *2*(3), 49–55. doi:10.1109/MSP.2004.26.

Krajzewicz, D., Hertkorn, G., Rössel, C., & Wagner, P. (2002). SUMO (simulation of urban mobility) — An open-source traffic simulation. In M. Al-Akaidi (Ed.), *MESM 2002, 4th Middle East Symposium on Simulation and Modelling*, (pp. 183–187). Erlangen, Germany: IEEE.

Le, Z., Ouyang, Y., Chen, G., & Makedon, F. (2011). Dynamic mix zone: Location data sanitizing in assisted environments. *Universal Access in the Information Society*, *10*(2), 195–205. doi:10.1007/s10209-010-0198-4.

Li, L., Song, J., Wang, F.-Y., Niehsen, W., & Zheng, N. (2005). New developments and research trends for intelligent vehicles. *IEEE Intelligent Systems*, *20*(4), 10–14. doi:10.1109/MIS.2005.73.

Lin, X., Lu, R., Liang, X., & Shen, X. (2011). Stap: A social-tier-assisted packet forwarding protocol for achieving receiver-location privacy preservation in vanets. In *Proceedings of the 30th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies,* (pp. 2147–2155). IEEE.

Lu, H., & Poellabauer, C. (2010). Balancing broadcast reliability and transmission range in vanets. *SIGMOBILE Mob. Comput. Commun. Rev.*, *14*(4), 25–27. doi:10.1145/1942268.1942278.

Lu, R., Lin, X., Liang, X., & Shen, X. S. (2012a). A dynamic privacy-preserving key management scheme for location-based services in vanets. *IEEE Transactions on Intelligent Transportation Systems*, *13*(1), 127–139. doi:10.1109/TITS.2011.2164068.

Lu, R., Lin, X., Luan, T., Liang, X., Li, X., Chen, L., & Shen, X. (2012b). Prefilter: An efficient privacy-preserving relay filtering scheme for delay tolerant networks. In *Proceedings of the 31th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies.* IEEE.

Lu, R., Lin, X., & Shen, X. (2010). Spring: A social-based privacy-preserving packet forwarding protocol for vehicular delay tolerant networks. In *Proceedings of the 29th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies,* (pp. 632–640). IEEE.

Lu, R., Lin, X., Zhu, H., Ho, P.-H., & Shen, X. (2008). Ecpp: Efficient conditional privacy preservation protocol for secure vehicular communications. In *Proceedings of 27th IEEE International Conference on Computer Communications, Joint Conference of the IEEE Computer and Communications Societies,* (pp. 1229–1237). IEEE.

National Highway Traffic Safety Administration. (2012). *An examination of driver distraction as recorded in NHTSA databases*. Retrieved from http://www-nrd.nhtsa.dot.gov/Pubs/811216.pdf

Ns. (2001). *The network simulator ns-2 (v2.1b8a)*. Retrieved from http://www.isi.edu/nsnam/ns/

Palanisamy, B., & Liu, L. (2011). Mobimix: Protecting location privacy with mix-zones over road networks. In *Proceedings of the 27th International Conference on Data Engineering (ICDE 2011)*, (pp. 494–505). Hannover, Germany: ICDE.

Rawat, D. B., Popescu, D., Gongjun, Y., & Olariu, S. (2011). Enhancing vanet performance by joint adaptation of transmission power and contention window size. *IEEE Transactions on Parallel and Distributed Systems*, *22*(9), 1528–1535. doi:10.1109/TPDS.2011.41.

Raya, M., Papadimitratos, P., & Hubaux, J.-P. (2006). Securing vehicular communications. *IEEE Wireless Communications Magazine*, 8–15.

Ribagorda-Garnacho, A. (2010). Authentication and privacy in vehicular networks. *Journal of UPGRADE*, *11*(1), 72–79.

Sampigethaya, K., Huang, L., Li, M., Poovendran, R., Matsuura, K., & Sezaki, K. (2005). Caravan: Providing location privacy for vanet. In Proceedings of Embedded Security in Cars. ESCAR.

Sampigethaya, K., Li, M., Huang, L., & Poovendran, R. (2007). Amoeba: Robust location privacy scheme for vanet. *IEEE Journal on Selected Areas in Communications*, *25*(8), 1569–1589. doi:10.1109/JSAC.2007.071007.

Song, J., Wong, V. W. S., & Leung, V. C. M. (2009). Wireless location privacy protection in vehicular ad-hoc networks. *Mobile Networks and Applications*, *15*(1), 160–171. doi:10.1007/s11036-009-0167-4.

Studer, A., Shi, E., Bai, F., & Perrig, A. (2009). Tacking together efficient authentication, revocation, and privacy in vanets. In *Proceedings of the 6th Annual IEEE Communications Society Conference on Sensor, Mesh and Ad Hoc Communications and Networks*, (pp. 484–492). IEEE.

Sun, J., Zhang, C., Zhang, Y., & Fang, Y. M. (2010a). An identity-based security system for user privacy in vehicular ad hoc networks. *IEEE Transactions on Parallel and Distributed Systems*, *21*, 1227–1239. doi:10.1109/TPDS.2010.14.

Sun, Y., Su, X., Zhao, B., & Su, J. (2010b). Mix-zones deployment for location privacy preservation in vehicular communications. In *Proceedings of the 10th IEEE International Conference on Computer and Information Technology (CIT 2010)*, (pp. 2825–2830). West Yorkshire, UK: IEEE.

Toyota. (2007). *Pre-crash safety*. Retrieved from http://www.toyota.co.jp/en/about_toyota/in_the_world/pdf2007/safety.pdf

US Department of Transporation, Research and Innovative Technology Association. (2011). *National transportation statistics*. Retrieved from http://www.bts.gov/publications/national_transportation_statistics/

Wang, F.-Y. (2010). Parallel control and management for intelligent transportation systems: concepts, architectures, and applications. *IEEE Transactions on Intelligent Transportation Systems*, *11*(3), 630–638. doi:10.1109/TITS.2010.2060218.

Wen, D., Yan, G., Zheng, N., Shen, L., & Li, L. (2011). Towards cognitive vehicles. *IEEE Intelligent Systems Magazine*, *26*(3), 76–80. doi:10.1109/MIS.2011.54.

Xi, Y., Sha, K., Shi, W., Schwiebert, L., & Zhang, T. (2007). Enforcing privacy using symmetric random key-set in vehicular networks. In *Proceedings of the Eighth International Symposium on Autonomous Decentralized Systems*, (pp. 344–351). IEEE.

Xie, H., Kulik, L., & Tanin, E. (2010). Privacy-aware traffic monitoring. *IEEE Transactions on Intelligent Transportation Systems*, *11*(1), 61–70. doi:10.1109/TITS.2009.2028872.

Yan, G., & Olariu, S. (2011). A probabilistic analysis of link duration in vehicular ad hoc networks. *IEEE Transactions on Intelligent Transportation Systems*, *12*(4), 1227–1236. doi:10.1109/TITS.2011.2156406.

Yan, G., Olariu, S., & Popescu, D. (2012). *Notice: An architecture for the notification of traffic incidents*. IEEE Intelligent Transportation Systems Magazine.

Yan, G., Olariu, S., Wang, J., & Arif, S. (2013). Towards providing scalable and robust privacy in vehicular networks. *IEEE Transactions on Parallel and Distributed Systems*. doi:10.1109/TPDS.2013.142.

Yan, G., Olariu, S., & Weigle, M. (2009a). Providing location security in vehicular ad hoc networks. *IEEE Wireless Communications*, *16*(6), 48–55. doi:10.1109/MWC.2009.5361178.

Yan, G., Olariu, S., & Weigle, M. C. (2008). Providing VANET security through active position detection. *Computer Communications*, *31*(12), 2883–2897. doi:10.1016/j.comcom.2008.01.009.

Yan, G., Olariu, S., & Weigle, M. C. (2009b). Providing location security in vehicular ad-hoc networks. *IEEE Wireless Communications*, *16*(6), 48–55. doi:10.1109/MWC.2009.5361178.

Yan, G., Yang, W., Rawat, D. B., & Olariu, S. (2011). Smartparking: A secure and intelligent parking system. *IEEE Intelligent Transportation Systems Magazine*, *3*(1), 18–30. doi:10.1109/MITS.2011.940473.