



ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

School of Computer & Communication Sciences
Laboratory for computer Communications and Applications
(LCA1)

Master of Science

Location Privacy amidst Local Eavesdroppers

by

Mathias Humbert

Supervisors

Prof. Jean-Pierre Hubaux

Mr. Julien Freudiger

Dr. Mohammad Hossein Manshaei

Lausanne, June 2009

Abstract

Mobile social networks and location-aware applications are becoming more and more widespread in current wireless networks. Thanks to Wifi or Bluetooth-enabled devices, everyday mobile users can enjoy new services. However, the deployment of these mobile technologies leads to many concerns for privacy, especially location privacy. In wireless networks, external malicious parties can monitor pseudonyms used for identification to learn mobile users' location and track their movements. A common technique for protecting location privacy consists in changing pseudonyms in regions called mix zones. In this report, we present a game-theoretic approach to evaluate the interaction and behaviors of an attacker aiming to jeopardize mobile nodes' location privacy and nodes willing to thwart adversary's spiteful plans. Assuming that such an attacker is armed with local eavesdropping devices, he must deploy them in an efficient way to track as many nodes as possible. On the other hand, mobile users have to define where are the best places to locate their mix zones. In order to evaluate their potential benefit, mobile nodes need to know the mixing effectiveness of possible mix zone locations. We propose a simplified metric based on the mobility profiles to determine the location privacy achieved with mix zones. We also build a payoff model to deal with costs, as well as benefits, led by this extended hide-and-seek game. By means of analytical and numerical results, we show the existence of Nash equilibria for all values of players' costs and mobility parameters. We prove that the adversary's best behavior evolves regarding the benefit he gets from traffic sniffing and the cost led by deployment of eavesdropping stations. On the other hand, we show that, when the mobility profile is homogeneous, the mobile nodes' best response is independent of the adversary's deployment of sniffing stations. Furthermore, the nodes' best response is only dependent on the number of mix zones they deploy, not on their particular locations.

Acknowledgements

First of all, I would like to thank Jean-Pierre Hubaux without whom I would never have come to such an interesting Master thesis subject. He provided me a good insight into the project and, more generally, tremendous motivation to pursue research in wireless privacy. I am also very grateful to Julien Freudiger and Mohammad Hossein Manshaei for supervising me during the whole semester. They both provided me with excellent theoretical support and relevant feedback. I would also like to thank Reza Shokri who helped me developing a simplified measure of mix zone effectiveness. Finally, a special note of thanks to Igor Bilogrevic for his game-theoretic and writing expertise, as well as our stimulating and friendly discussions.

Contents

Contents	vii
1 Introduction	1
2 Related work	5
3 Preliminaries	9
3.1 System model	9
3.2 Mix zones effectiveness	12
4 Game model	17
4.1 Game-theoretic background	17
4.2 Benefits	19
4.2.1 Mobile nodes	19
4.2.2 Adversary	20
4.3 Costs	20
4.3.1 Mobile nodes	20
4.3.2 Adversary	21
4.4 Payoffs	21
5 Game results	25
5.1 Equality game: G^e	25
5.1.1 Nodes' best response	26
5.1.2 Adversary's best response	28
5.1.3 Nash equilibrium	30
5.1.4 Discussion and simulations	31
5.2 General game: G^g	35
5.2.1 Adversary's best response	35
5.2.2 Nodes' best response	37
5.2.3 Discussion and simulations	38
5.3 Discussion	42

6 Conclusion and Future Work	43
Bibliography	45
Appendices	49
A Game results	51
A.1 Equality game	51
A.1.1 Nodes' best response	52

Chapter 1

Introduction

Over the last three decades, the number of cell phones, laptops and other electronic mobile devices has tremendously increased. In the last few years, these mobile devices have started to be equipped, besides cellular interfaces (GSM), with peer-to-peer communication technologies, such as WiFi or Bluetooth. They enable new services and applications such as *context-aware* applications. For example, in France, INRIA, in collaboration with JCDecaux (second largest outdoor advertising corporation in the world), is currently developing a service that will allow a person passing in front of a movie advertising board to download the trailer on his mobile phone via Bluetooth or Infrared. In front of a car billboard, he will receive the address of the closest car dealer by SMS. These services would require user's prior consent and registration of personal information (age, leisure, ...) on his phone [23]. Another benefit provided by these new communication technologies is the possibility given to friends to automatically detect and exchange information through mobile social networks [1, 2].

However, the deployment of these technologies leads to serious privacy issues. As Buttyán and Hubaux show with RFID tags [6], without privacy protecting measures, our wireless world “could easily degenerate into the one described by Orwell in his book 1984”. Nowadays, it is easy for an ordinary person to monitor all Bluetooth-enabled devices and process gathered information using an automated system [11]. In order to prevent these new threats on privacy, an information report written by two French senators reaffirms the right to privacy as a fundamental right of our democratic countries. They warn us of the danger led by new technologies (Bluetooth, RFID tags, geolocation, ...), the surge of social networks and their consequences on privacy. They make fifteen recommendations in order to better protect the right to privacy in our new information age [10]. Daniel Solove also asserts that it is imperative to have a theory about what privacy is and why it is valuable [32]. This suggests that it is also ex-

tremely imperative to develop novel technical mechanisms in order to protect this privacy. In this project, we concentrate on privacy-preserving mechanisms in wireless networks.

In mobile wireless networks, a malicious third party can monitor pseudonyms used for identification to learn mobile nodes' locations. Mobile nodes (e.g. pedestrian or vehicular) can be tracked by monitoring their MAC (Medium Access Control) addresses (link layer identifiers), IP addresses (network layer identifiers) or even wireless fingerprints (physical layer identifiers). An adversary aiming to track mobile devices is also able to eavesdrop the public keys (application layer identifiers) that have been either preloaded into mobile devices or directly generated by mobile nodes. An external party can compromise the *location privacy*, defined in [21] as the ability to prevent other parties from learning one's movement.

One common technique for improving the location privacy consists in using *multiple pseudonyms* [3] and coordinate their changes in regions called *mix zones* [4]. Other popular techniques developed to prevent nodes' tracking will be detailed in chapter 2.

In contrast with previous works that assume the presence of a *global passive adversary (GPA)* being able to eavesdrop all communications, we study location privacy amidst *local eavesdroppers*. We keep the assumption of a passive adversary (the adversary cannot inject messages into a target's neighborhood) and thus call the malicious external party a *local passive adversary (LPA)*. We think that global coverage assumption is too simplistic for several reasons. First, no network has a complete coverage, even a phone network. Second, cost might be too expensive for an adversary to build and maintain a global eavesdropping network. Third, computational complexity is as well too high to sort and process all the received signals. Finally, the adversary gets more flexibility and can change the place of stations dynamically.

The first main contribution lies in the use of a game-theoretic approach to model the interaction between mobile nodes' and adversary's strategies. As we made the LPA assumption, the attacker has a limited number of eavesdropping stations and, thus, must optimize their placement to efficiently track mobile nodes. Moreover, these devices are not free and incur a non-negligible cost. On the other hand, we assume that mobile nodes use active (or silent) mix zones to protect their location privacy. We define an active mix zone as a precise place where mobile devices deliberately shut their transmitters down. As mobile nodes must both change pseudonyms and stop transmitting during the time they are within the mix zones, mix zones are costly for mobile nodes. Trade-off between benefits and costs, as well as interaction and conflicting goals between mobile nodes' and attacker's strategies led us to tackle this problem using game theory.

Second, we develop a simplified *flow-based* measure of the mix zone's effec-

tiveness. This measurement, in terms of the location privacy it provides, depends on the adversary's ability to correlate nodes that enter and leave the mix zone. Previous works on mix zones' effectiveness was either *event-based* [3] or *flow-based* [14]. We propose to base our analysis on entering and exiting traffic intensities, and on the variances of nodes' sojourn time. It keeps the interesting feature of flow-based model, which is to be able to compute mixing effectiveness prior to mix zones' deployment, and remains simple and intuitive.

The remainder of this Master thesis is structured as follows. In next chapter, we will briefly discuss current privacy-preserving mechanisms and metrics measuring the level of location privacy. Chapter 3 is dedicated to the system model. We also explain, in chapter 3, how we derive the simplified *flow-based* effectiveness measure. In chapter 4, we derive and clarify both the nodes' and adversary's payoff functions. Chapter 5 provides the analytical and simulated results. Finally, in chapter 6 we conclude and present the remaining challenges.

Chapter 2

Related work

In order to achieve location privacy, mobile nodes can first merely add noise to their location [18], or report their location as a region instead of a point [33]. Unfortunately, these techniques are insufficient to protect location privacy since peer-to-peer wireless communications between nodes unveil their locations.

Freudiger *et al.* develop a new paradigm based on ring signatures [13]. Ring signatures schemes are simplified group signatures, which are themselves a “generalization” of the credential/membership authentication schemes. Group signatures [8] represent a signature scheme that allow a group member to sign on behalf of a group without revealing the identity of the signer. They have the following three properties. First, only members of the group can sign messages. Second, the receiver can verify that it is a valid group signature, but cannot discover which group member made it. Third, in case of dispute later on, the signature can be “opened” to reveal the identity of the signer. Unlike group signatures, ring signatures [27] need no group managers, no setup procedures, no revocation procedures, and no coordination. The members of a ring can thus change over time without central coordination. Ring signatures are finally useful when the members do not want to cooperate, contrarily to group signatures that require cooperation of members. For both ring and group signatures, the size of the group respectively the ring determines the privacy of its members.

Besides cryptographic constructions, mobile nodes can also protect their location privacy using spatial/temporal mixing. To efficiently mix their trajectories, nodes must first change their pseudonyms over time. But this is not sufficient. Indeed, only changing pseudonyms does not help since most of time mobile nodes have different speeds and trajectories and, thus, attacker can jeopardize their location privacy using spatial and temporal correlation.

In order to make correlation harder, Huang *et al.* propose to stop transmitting for a variable length silent period during which at least two nodes close together

are changing their pseudonyms [20]. This leads us to mix zone's concept [4]. Mix zones are pre-determined regions within which mobile nodes are forced to change their pseudonyms. However, important drawback of mix zones is that they are inefficient when nodes' density is low and can be costly in terms of pseudonym management.

Another method enhancing location privacy is Swing & Swap [24], which is user-centric. Swing enables nodes to loosely synchronize updates when changing their velocity. Swap, an extension of Swing, provides even more location privacy. It assumes cooperation of nodes that enables exchange of their identifiers.

We need to define appropriate metrics in order to measure the level of location privacy in mix zones, i.e. mix zones' efficiency. In [26], Pfitzmann and Köhntopp define anonymity as state of being not identifiable within a set of subject, the *anonymity set*. In mobile networks, the anonymity set is typically the set of nodes within the mix zone.

Serjantov and Danezis [29] point out some issues associated with the anonymity set and proposes an information-theoretic metric based on entropy and anonymity probability distribution. Using mix-based anonymity systems [7] in which all network communication is observable by the attacker, they discuss the vulnerability of anonymity set metric against an attacker's additional knowledge (the probabilities of A communicating with B and C for instance). In mobile networks, the distribution of nodes entering and leaving the mix zone may well represent the metric developed by Serjantov and Danezis.

Freudiger *et al.* propose a novel metric based on nodes' mobility profiles [14]. They compute the mixing effectiveness prior to operations of mobile network using an information-theoretic distance measure: the Jensen-Shannon divergence. They provide a lower and an upperbound on the probability of error made by an adversary aiming to track mobile nodes. These bounds on probability of error take into account the relative flow intensities and the sojourn time distributions, modeled as Normal distributions, within the mix zones.

Finally, Shokri *et al.* present in [31] formal representations of four among the most relevant categories of location privacy metrics. They analyse these metrics based on a set of criteria such as probability of error, tracking error and actual trace. In addition, they propose a *distortion-based* metric for measuring location privacy. In this metric, they estimate how distorted is the reconstructed *actual trace* of each user at any time epoch. The location privacy of a mobile nodes is then directly proportional to the distortion level in reconstructing his actual trace.

Even though nodes transmitting and receiving data over a mobile wireless network are vulnerable to location tracking and other privacy-leaking attacks, they also get important advantages such as location-based applications, which use node's current location to provide services (e.g., restaurant or gas station

close to nodes, road traffic information). The data quality providing by location-aware services depends mainly on the accuracy of location information. Mix zones, Swing & Swap and all other techniques based on silent period need to stop transmitting during a variable length of time. This leads to a decrease in quality of location-based services, which is a non-negligible cost for nodes. In [19], Hoh and Gruteser propose a QoS metric based on mean location error for a set of N different users' paths of length K . In order to enhance QoS while still protecting location privacy, [22] develops a protocol called silent cascade. This model achieves anonymity without violating QoS by trading off the length of mix-network for anonymity.

Chapter 3

Preliminaries

3.1 System model

We consider a part of a road network composed of N mobile nodes equipped with peer-to-peer communication technologies, such as WiFi or Bluetooth. We model this road network as a 2-by-2 grid, representing four roads crossing at four intersections. This simplified model does not encompass the whole complexity of a real geographical space but can be viewed as a small part of a greater map. Indeed, by assembling many 2-by-2 grids, we could almost reach typical road networks, such as Manhattan's one. Moreover, we must have a simple model in order to keep level of complexity acceptable.

The adversary installs eavesdropping stations at crossroads. Therefore, in our model, he has at most four different places where he can sniff the network, which matches with our *local adversary* assumption. Even though the attacker places four sniffing stations, he still does not have a complete coverage since the adversary is not able to eavesdrop the communications of nodes outside the range of his radio receivers.

On the other hand, mobile nodes assign the responsibility to establish security and privacy in the network to a trusted central authority (e.g., in vehicular networks, the vehicle registration authority). This authority deploys a limited number of *active* mix zones at crossroads. In order to mix their trajectories, mobile nodes stop transmitting during the time they spend within mix zones (what we call *active* in contrast with *passive* mix zone that we define as the region where the adversary has no coverage). We assume that mix zones are a bit smaller than adversary's coverage radius. Each mix zone i is traversed by incoming roads $r_j \in R_i \subseteq R$. Previous work on mix zones' effectiveness was either *event-based* [3] or *flow-based* [14]. We propose to base our analysis on entering and exiting traffic intensities, as well as variances of mobile nodes' sojourn time. As we ex-

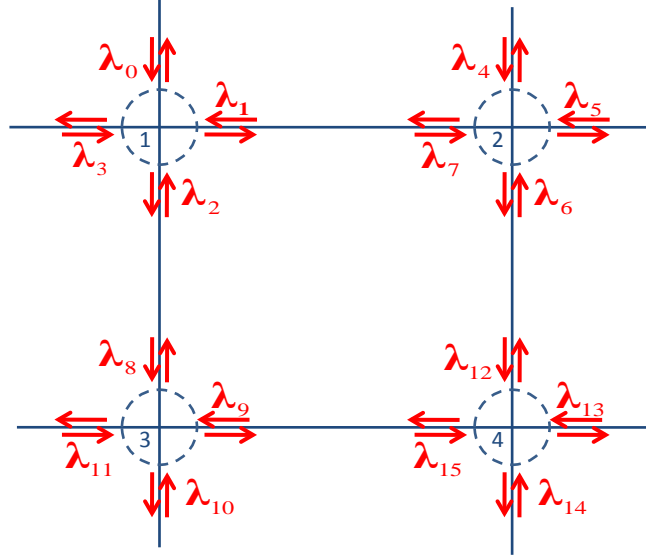


Figure 3.1: System model. 2-by-2 grid that represents 4 roads crossing at 4 points of junction. Intersections are numbered from 1 to 4, from top left to bottom right.

plain in section 3.2, this is a simplified version of flow-based model derived in [14]. It keeps the interesting feature of flow-based model, which is to be able to compute mixing effectiveness prior to mix zones' deployment.

Figure 3.1 shows our system model, with traffic intensity expressed as λ_i . We assume that intensities entering and exiting the intersections at a precise point are equal. Note that λ_2 is equal to λ_8 since nodes exiting intersection 1 are going directly to intersection 3. In the same way, $\lambda_1 = \lambda_7$, $\lambda_6 = \lambda_{12}$ and $\lambda_9 = \lambda_{15}$. The traffic intensities are formally defined in a square matrix F :

$$F = \begin{bmatrix} \lambda_0 & \lambda_1 & \lambda_2 & \lambda_3 \\ \lambda_4 & \lambda_5 & \lambda_6 & \lambda_7 \\ \lambda_8 & \lambda_9 & \lambda_{10} & \lambda_{11} \\ \lambda_{12} & \lambda_{13} & \lambda_{14} & \lambda_{15} \end{bmatrix}$$

where the lines represent the four road intersections $C = \{C_1, C_2, C_3, C_4\}$ and the columns the point of entrance of nodes in the intersection. First column means that nodes come from the North, second one from the East, third one from the South and fourth from the West. Moreover, we define a vector $\vec{\mu}$ which represents the expected time the nodes spend within each intersection (sojourn time) that

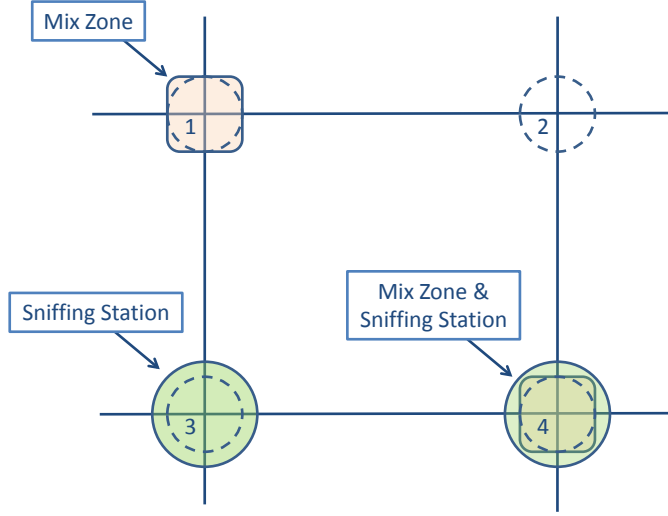


Figure 3.2: Example with 2 mix zones and 2 eavesdropping stations:
 $\vec{m} = [1 \ 0 \ 0 \ 1]^T$, $\vec{a} = [0 \ 0 \ 1 \ 1]^T$.

will be useful to compute the cost due to disconnected time:

$$\vec{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \\ \mu_4 \end{bmatrix}$$

Finally, a matrix V of sojourn time variances represents variances of delays for each possible entries at each intersection:

$$V = \begin{bmatrix} \sigma_0^2 & \sigma_1^2 & \sigma_2^2 & \sigma_3^2 \\ \sigma_4^2 & \sigma_5^2 & \sigma_6^2 & \sigma_7^2 \\ \sigma_8^2 & \sigma_9^2 & \sigma_{10}^2 & \sigma_{11}^2 \\ \sigma_{12}^2 & \sigma_{13}^2 & \sigma_{14}^2 & \sigma_{15}^2 \end{bmatrix}$$

Based on the intensity matrix F , and variances matrix V (mobility profile), we would like to compute the effectiveness provided by placing a mix zone for each intersection. In this way, we will be able to compute the best placement of mix zones for the nodes, i.e., get the best strategy. Vector \vec{m} represents the presence (entry equal to 1) or absence (entry equal to 0) of a mix zone. The adversary

will be as well able to optimize the placement of his sniffing stations given the mobility profile of nodes. His strategic vector is \vec{a} , with entries equalling 1 if he places an eavesdropping station at the intersection and 0 otherwise. Figure 3.2 is an example of possible nodes' and adversary's behavior.

We link effectiveness of mix zones directly to the uncertainty it leads to for the adversary trying to track nodes. Based on work done in [14], we derived a metric that takes into account the flows' delay, the relative and absolute intensities of flows. In next subsection, we formally derive this value and explain how we are getting it.

3.2 Mix zones effectiveness

We base the derivation of the effectiveness measure essentially on the work done in [14] and the conclusions they reached. In this paper, they take four mobility factors into account in order to capture the whole uncertainty, described as a probability of error, provided by a mix zone. These four factors come from two different features of flows: their intensities and their sojourn time distribution. The first factor used to compute the probability of error is given by the relative intensities of flows, what is called the *a priori* probability. The second one involved at the end of the computation of this probability of error is given by the absolute intensities of flows. The last two factors are included in the sojourn time distributions. They are the mean and variance of sojourn times. These are used to model the sojourn times of different flows, that are distributed according to Normal distributions.

Although the difference of sojourn time means between two different flows is helpful to distinguish between the two distributions, we decide here, in order to keep a succinct effectiveness measure, to take only the intensities and the sojourn time variances into account. Indeed, as nodes entering the mix zone via different roads have independent random rates of arrival, the difference δ between an arrival from one road and another is completely random. Hence, even with a large difference between two expected sojourn times Δ , nodes might still be very difficult to distinguish at the exit if their difference of arrival times δ is approximately equal to Δ . As arrival rates are independent, the probability to have $\delta \approx \Delta$ should not be smaller than $\delta \approx 0$ if Δ is not too big. Therefore, we assume that, among all entering nodes and not only two specific arrivals, the intensities and the sojourn time variances encompass all the main information needed to evaluate global uncertainty provided by a mix zone.

First, we focus on the relative intensities of flows and compute the entropy led by the probabilities of incoming flows. Entropy plays a central role in information theory as a measure of information, choice and uncertainty [30]. It has been proposed as a metric for anonymity in [29]. Let us assume that n entering roads

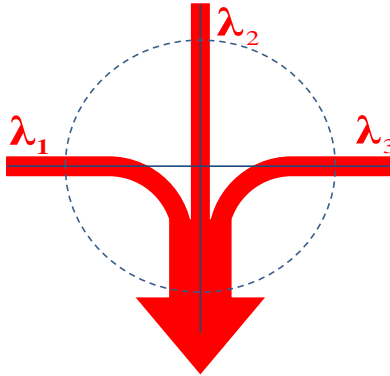


Figure 3.3: Model with $n = 3$ entering roads leaving at the same exit.

are converging to one exiting road (Figure 3.3). We get the highest possible value of entropy whether all incoming road flows have the same probability, thus equal to $1/n$. The more different the probabilities are, the less is the uncertainty for the attacker, and so the smaller the entropy. This value contains uncertainty (or certainty) led by the relative intensity of flows.

Second, we need to take the absolute intensities of flows into account. The uncertainty increases when the number of entering nodes is bigger, i.e. when the sum of intensities of entering roads is higher. This is quite intuitive since the more nodes there are within the crossroads, the more dense is the output and thus the more difficult is the monitoring process for the adversary. This value includes the uncertainty (or certainty) led by the absolute intensity of flows.

Finally, the adversary is using as well the information about sojourn times (delays) of the different entering flows. We assume that delays are distributed as Gaussian distributions, with possibly different means and variances. What is of first importance in delays distribution for our metric is the variance of each sojourn time distribution. The greater are the variances, the harder it is for the adversary to match entering with leaving nodes. This value takes thus into account the uncertainty (or certainty) led by the sojourn time distributions.

Note that we assume for simplicity, as in [14] that the sojourn time distribution is independent of the flows' intensity. The model can be extended to capture the dependency between mobile nodes' arrivals, intensities, and their effect on the sojourn time distributions [16].

We are now able to express this effectiveness measure formally. Let us call it E_{tot} . Then, assuming we have n entering road segments r_1, \dots, r_n , total uncer-

tainty is equal to the multiplication of all these three mixing factors:

$$E_{tot} = H(\vec{p})\bar{\sigma}^2 \sum_{i=1}^n \lambda_i \quad [bits \cdot nodes] \quad (3.1)$$

where p_i 's are the relative intensity probabilities:

$$p_i = \frac{\lambda_i}{\sum_{j=1}^n \lambda_j}, \quad (3.2)$$

the entropy [9] is:

$$H(\vec{p}) = - \sum_{i=1}^n p_i \log p_i \quad (3.3)$$

and the average variance is:

$$\bar{\sigma}^2 = \sum_{i=1}^n p_i \sigma_i^2 \quad (3.4)$$

In order to test this new metric, we plot three different graphs providing the uncertainty with respect to two of the three factors. Note here that we use $n = 2$ entering flows. Figure 3.4 shows how our uncertainty measure is influenced by the sum of intensities $\lambda_1 + \lambda_2$ and by the ratio λ_2/λ_1 of flows' intensities, with fixed average variance equal to $\bar{\sigma}^2 = 0.6$. Figure 3.5 displays the values of our uncertainty measure for different average variances $\bar{\sigma}^2$ and different ratios λ_2/λ_1 , with fixed absolute intensities $\lambda_1 + \lambda_2 = 1$. Figure 3.6 shows how the measure varies with absolute values of intensities $\lambda_1 + \lambda_2$ and the average variance $\bar{\sigma}^2$, with a fixed ratio $\lambda_2/\lambda_1 = 1$. First, we observe that if either $\lambda_1 + \lambda_2$ or $\bar{\sigma}^2$ increases, then E_{tot} increases as well, showing that if intensity of nodes entering and exiting the mix zone is higher it becomes much more difficult to track every nodes accurately. It also shows that if variances of sojourn time distributions increase, then uncertainty raises proportionally since delay probability distribution function is less narrow and thus exit time distributed over a wider range with higher probability. In addition, we observe in figures 3.4 and 3.5 that if λ_2/λ_1 increases and absolute intensities and delay variances remain constant, uncertainty decreases. The intuition is that as the ratio between two flows' intensities increases, the flow with higher intensity dominates the exit of the considered mix zone and becomes easier to track. The uncertainty, and thus the mixing effectiveness, is maximal when both flows have the same intensity ($\lambda_2/\lambda_1 = 1$) and have the highest possible intensities and sojourn time variances.

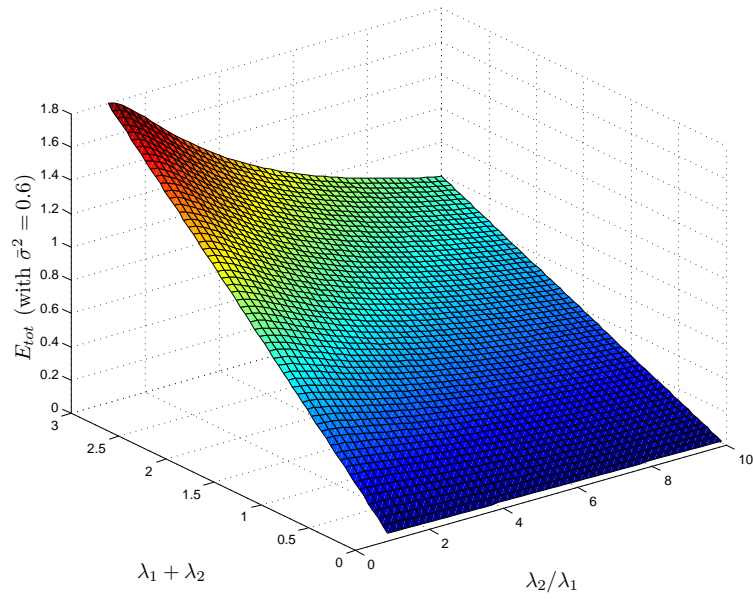


Figure 3.4: Uncertainty measure E_{tot} [bits · nodes] led by two entering roads r_1 and r_2 with fixed average variance $\bar{\sigma}^2 = 0.6s$, $\lambda_2/\lambda_1 \in [1, 10]$, and $\lambda_1 + \lambda_2 \in [0, 3]$. As $\lambda_1 + \lambda_2$ increases, nodes' density within the mix zone increases as well and it becomes more difficult to correlate exiting events to entering ones. The increase in E_{tot} is slower if ratio λ_2/λ_1 is bigger.

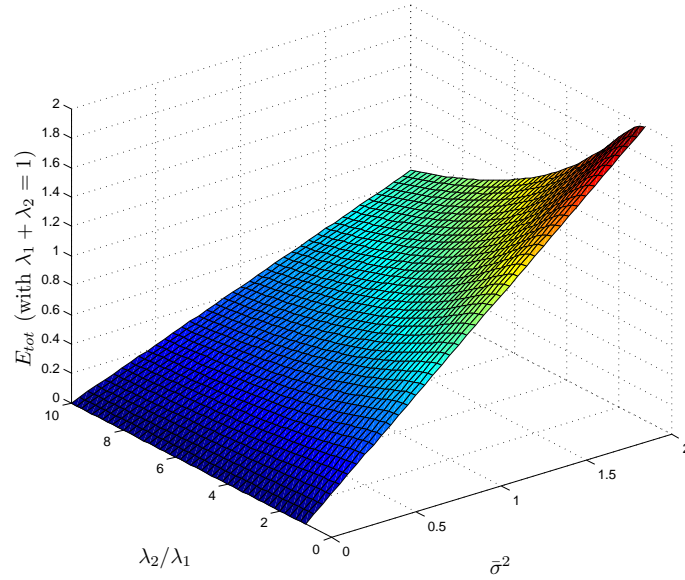


Figure 3.5: Uncertainty measure E_{tot} [bits · nodes] with fixed sum of absolute intensities $\lambda_1 + \lambda_2 = 1$, $\bar{\sigma}^2 \in [0, 2]$, and $\lambda_2/\lambda_1 \in [1, 10]$. As average variance $\bar{\sigma}^2$ raises, uncertainty on time of exit increases and thus correlation becomes harder. The increase in E_{tot} is slower if ratio λ_2/λ_1 is bigger.

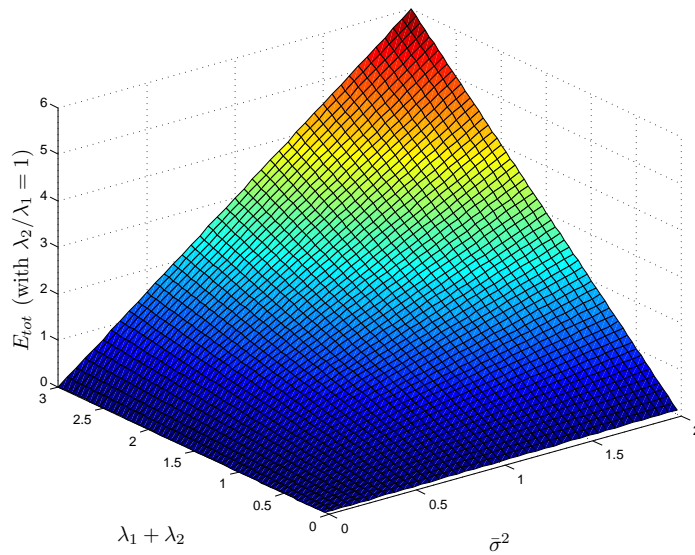


Figure 3.6: Uncertainty measure E_{tot} [bits · nodes] with fixed ratio $\lambda_2/\lambda_1 = 1$, $\bar{\sigma}^2 \in [0, 2]$, and $\lambda_1 + \lambda_2 \in [0, 3]$. As $\bar{\sigma}^2$ and $\lambda_1 + \lambda_2$ increase, uncertainty increase proportionally to both parameters.

Chapter 4

Game model

4.1 Game-theoretic background

In this chapter, we define and explain a game that models the conflict between mobile nodes willing to protect their location privacy and the adversary trying to jeopardize it.

Game theory allows for modeling such conflicting situations and for predicting the behavior of participants. Roger B. Myerson defines it as the “study of mathematical models of conflict and cooperation between intelligent rational decision-makers” [25]. In our analysis, as in all game-theoretic models, we assume that participants are rational, which means that they make decisions consistently in pursuit of their own payoff’s maximization. Finally, it should be remembered here that game theory is the logical fulfillment of decision theory and has been first used in economics and political science to model human behaviors.

We first define a game G as a triplet $(\mathcal{P}, \mathcal{S}, \mathcal{U})$, where \mathcal{P} is the set of players, \mathcal{S} is the set of strategies and \mathcal{U} is the set of payoff functions:

- **Players:** There are two of them. The first one is represented by the aggregation of mobile nodes, that are cooperating together in order to achieve location privacy. Freudiger *et al.* have already studied protection of location privacy, but in a user-centric manner assuming non-cooperative behavior of the mobile nodes [12]. We call this first player n . The second player is the adversary or attacker, called a , whose aim is to track mobile nodes. Hence, the set of players is $\mathcal{P} = \{n, a\}$ where n is itself the set of all mobile nodes within the network.
- **Strategies:** Each player has 16 possible strategies. Mobile nodes use active mix zones and must decide whether to make them run or not, for each crossroads. The strategies $s_n^x \in \mathcal{S}_n$ are represented by vector \vec{m}_n , of length

4 (number of intersections in the network) with $m_i = \{0 \text{ (no mix zone), } 1 \text{ (mix zone)}\}$. As we have two alternatives, “place” a mix zone ($m_i = 1$) or not ($m_i = 0$), at each crossroads, we get $2^4 = 16$ strategies, from zero (s_n^1) to four mix zones (s_n^{16}). On the other hand, the adversary deploys eavesdropping stations at crossroads to efficiently track the nodes. His strategies $s_a^x \in \mathcal{S}_a$ are represented by vector \vec{a} , of length 4 as well, with $a_i = \{0 \text{ (no sniffing station), } 1 \text{ (sniffing station)}\}$. The attacker also has the choice to place an eavesdropping station ($a_i = 1$) or not ($a_i = 0$) at each crossroads. He then also has $2^4 = 16$ different strategies, from zero (s_a^1) to four sniffing stations (s_a^{16}). The set of strategies in the game is thus $\mathcal{S} = \{\vec{m}, \vec{a}\}$.

- **Payoff Functions:** Generally speaking, payoff is expressed as benefit minus cost. Mobile nodes’ payoff function is thus define as $u_n = b_n - c_n$ and $u_a = b_a - c_a$. In next section, we will explain in more details our benefits and costs, for both players.

Before concentrating on the payoffs, let us introduce game-theoretic concepts, such as *best response* and *Nash equilibrium*, and define the type of game we are dealing with.

We assume a complete information game, which means that each player’s payoff function is common knowledge among all the players [17]. We then assume that the adversary is aware of mobile nodes’ mobility profile, i.e., that he knows the traffic intensities and the sojourn time distributions for all entries of crossroads. In addition, our game is a non-cooperative one. Although mobile nodes are cooperating together to form a coalition and thus one unique player (just like Luxembourg, Switzerland and Austria are cooperating to protect bank secrecy against EU complaints), they are not cooperating with the adversary since nodes get no incentive to it. Their main goal is to thwart attacker’s tracking and thus cooperation with the adversary is irrelevant here. Finally, we approach this model with a simultaneous (or static) game, i.e. the players simultaneously choose actions and then receive payoffs that depend on the combination of actions just chosen.

Let us now introduce the concept of *best response*. We can write $br_i(s_j)$, the best response of player i to the opponent’s strategy s_j .

Definition 4.1. *The best response function of player i to the profile of strategies s_j is a function $br : s_j \rightarrow s_i$ that maximizes player i ’s payoff given other player’s strategies s_j . We can express it formally as:*

$$br_i(s_j) = \arg \max_{s_i \in \mathcal{S}_i} u_i(s_i, s_j) \quad (4.1)$$

If two strategies are mutual best responses to each other, then no player has the motivation to deviate from the given strategy profile. This leads us to the concept of Nash Equilibrium.

Definition 4.2. A strategy profile s^* constitutes a Nash equilibrium if, for each player i ,

$$u_i(s_i^*, s_j^*) \geq u_i(s_i, s_j^*), \forall s_i \in \mathcal{S}_i \quad (4.2)$$

where s_i^* and s_j^* are the Nash equilibrium strategies of player i and j , respectively.

In a Nash equilibrium, none of the players can unilaterally change his strategy to increase his payoff.

4.2 Benefits

4.2.1 Mobile nodes

Nodes obtain benefit only at the intersections where they place a mix zone. Then, either the adversary places a sniffing station at the same crossroads, or he does not, and benefit will depend on the attacker's strategy as well.

If there are both a sniffing station and a mix zone at intersection i , then each node going through this point of junction yields a benefit of $\sum_{j:r_j \in R_i} \frac{E_j}{4}$ where E_j is the uncertainty for the adversary at exit road r_j . We assume that nodes are entering through road r_k , leaving through road r_j , following an homogeneous Poisson process. The rate of arrival is λ_k , and of departure λ_j . From stochastic processes theory, we know that the expectation of a Poisson process $N(t)$ of intensity λ is

$$E[N_j(T)] = \lambda_j T. \quad (4.3)$$

Hence, for all $j : r_j \in R_i$, we need to multiply our sum by $\lambda_j t_{tot}$ where t_{tot} is the total time during which the nodes are moving in the network. The benefit for nodes with presence of both a mix zone and a sniffing station at an intersection is:

$$\sum_{\text{roads } j \in i} \frac{E_j}{4} \quad (4.4)$$

Regarding the case when there is a mix zone and no sniffing station at the intersection i , we assume that the mixing power of the mix zone is maximum. Therefore, we use an upperbound on E_j that we call E_{max}^i and we get a benefit equal to

$$\sum_{\text{roads } j \in i} \frac{\lambda_j T E_{max}^i}{4} \quad (4.5)$$

We can now express the nodes' benefit.

$$b_n = \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j T E_j}{4} + (1 - a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j T E_{max}^i}{4} \right] \quad (4.6)$$

4.2.2 Adversary

The adversary will get a benefit only at the crossroads where he is sniffing the network. As for the nodes, the adversary's benefit also depends on the nodes' strategy. The payoff will be higher whether there is no mix zone bothering attacker's tracking.

In this case, the adversary will get the same benefit as the nodes when there is a mix zone without any sniffing station. We consider the adversary being able to sniff all the nodes going through the intersection i and thus getting a benefit of E_{max}^i per node.

$$\sum_{\text{roads } j \in i} \frac{\lambda_j T E_{max}^i}{4} \quad (4.7)$$

If the nodes' strategy is to place a mix zone at the same intersection as the adversary's sniffing station, then his benefit will be proportional to $\sum_{j:r_j \in R_i} \frac{E_{max}^i - E_j}{4}$. When the uncertainty led by the mix zone is maximal, the benefit is very small or even zero and when the uncertainty tends to be small, the tracking benefit will be much higher. Hence, total benefit for an adversary sniffing at crossroads i with a mix zone's presence is

$$\sum_{\text{roads } j \in i} \frac{\lambda_j T (E_{max}^i - E_j)}{4} \quad (4.8)$$

Let us now express the adversary's total benefit.

$$b_a = \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j T (E_{max}^i - E_j)}{4} + (1 - m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j T E_{max}^i}{4} \right] \quad (4.9)$$

4.3 Costs

Let us discuss now the nodes' and adversary's costs. Placing a mix zone or an eavesdropping station is not free, neither for the nodes nor for the attacker.

$$c_n^{silence} = \frac{\text{disconnected time}}{\text{connected time}} = f(F, \vec{\mu}, \vec{m}, T) \quad (4.10)$$

4.3.1 Mobile nodes

The first cost of the nodes is related to the price of a pseudonym. As suggested in [12], pseudonym change is costly because it requires contacting a central authority, or self-generating a new pseudonym. Moreover, routing [28] becomes more difficult and requires frequent updates of routing tables. The cost of changing one pseudonyms is equal to γ . We then get a global cost of

$$c_n^{\text{pseudo}} = \gamma T (F \cdot \vec{1})^t \vec{m} \quad (4.11)$$

The second cost is involved by the time when the nodes stop transmitting and receiving data. It has for instance great impact on quality of service for location-aware applications. We compute this cost as the ratio of disconnected time over connected time. Total disconnected time is equal to $t_{tot}(F \cdot \vec{1})^T \vec{\mu}_m$ where vector $\vec{\mu}_m$ is an aggregation of vectors $\vec{\mu}$ and \vec{m} :

$$\vec{\mu}_m = \begin{bmatrix} \mu_1 \cdot m_1 \\ \mu_2 \cdot m_2 \\ \mu_3 \cdot m_3 \\ \mu_4 \cdot m_4 \end{bmatrix}$$

Total connected time is equal to total time minus total disconnected time. In order to ease the analysis, we express total time t_{tot} as four times the time spent in the crossroads. We thus assume that nodes spend on average one quarter of time within the crossroads and three quarters outside them. Factor 4 can be replaced by another one if needed. Total cost is

$$c_n^{\text{silence}} = \frac{t_{tot}(F \cdot \vec{1})^T \vec{\mu}_m}{t_{tot}(F \cdot \vec{1})^T 4\vec{\mu} - t_{tot}(F \cdot \vec{1})^T \vec{\mu}_m} = t_{tot} \frac{(F \cdot \vec{1})^T \vec{\mu}_m}{(F \cdot \vec{1})^T 4\vec{\mu} - (F \cdot \vec{1})^T \vec{\mu}_m} \quad (4.12)$$

Finally, global nodes' cost is

$$c_n = \gamma t_{tot}(F \cdot \vec{1})^T \vec{m} + t_{tot} \frac{(F \cdot \vec{1})^T \vec{\mu}_m}{(F \cdot \vec{1})^T 4\vec{\mu} - (F \cdot \vec{1})^T \vec{\mu}_m} \quad (4.13)$$

4.3.2 Adversary

Attacker's cost is highly related to the cost of managing the sniffing stations' operations, monitoring the network and processing eavesdropped signals. It is thus proportional to the total time and the number of eavesdropping stations deployed by the adversary, weighted by a constant factor β :

$$c_a = \frac{\beta T \sum_i a_i}{4} \quad (4.14)$$

4.4 Payoffs

Mixing both previous subsections, we can now express the complete payoff formulas:

$$u_n = \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j t_{tot} E_j}{4} + (1 - a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j t_{tot} E_{max}^i}{4} \right] - \gamma t_{tot}(F \cdot \vec{1})^T \vec{m} - t_{tot} \frac{(F \cdot \vec{1})^T \vec{\mu}_m}{(F \cdot \vec{1})^T 4\vec{\mu} - (F \cdot \vec{1})^T \vec{\mu}_m} \quad (4.15)$$

$$u_a = \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j t_{tot} (E_{max}^i - E_j)}{4} + (1 - m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j t_{tot} E_{max}^i}{4} \right] - \frac{\beta t_{tot} \sum_{i=1}^4 a_i}{4} \quad (4.16)$$

We can already simplify both u_n and u_a by t_{tot} and then get payoff functions independent of the total time of network operations/simulation:

$$u_n = \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_j}{4} + (1 - a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} \right] - \gamma (F \cdot \vec{1})^T \vec{m} - \frac{(F \cdot \vec{1})^T \vec{\mu}_m}{(F \cdot \vec{1})^T 4\vec{\mu} - (F \cdot \vec{1})^T \vec{\mu}_m} \quad (4.17)$$

$$u_a = \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (E_{max}^i - E_j)}{4} + (1 - m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} \right] - \frac{\beta \sum_{i=1}^4 a_i}{4} \quad (4.18)$$

It remains to express E_j and E_{max}^i mathematically. From section 3.2, uncertainty E_j is the multiplication of three factors: the absolute values of traffic intensities λ_i 's, the entropy led by the relative difference between these traffic intensities and the mean of traffic delays' variances.

$\forall i \in C = \{C_1, C_2, C_3, C_4\}$ and $\forall j$ such that $r_j \in R_i \subseteq R$,

$$E_j = (\lambda_{k_j} + \lambda_{l_j} + \lambda_{m_j}) H(\alpha_j \lambda_{k_j}, \alpha_j \lambda_{l_j}, \alpha_j \lambda_{m_j}) \bar{\sigma}_j^2 \quad (4.19)$$

with

$$\begin{aligned} k_j &= j + 1 \pmod{4} + 4(i - 1) \\ l_j &= j + 2 \pmod{4} + 4(i - 1) \\ m_j &= j + 3 \pmod{4} + 4(i - 1) \\ \alpha_j &= \frac{1}{\lambda_{k_j} + \lambda_{l_j} + \lambda_{m_j}} \end{aligned}$$

and

$$\bar{\sigma}_j^2 = \alpha_j \left(\lambda_{k_j} \sigma_{k_j}^2 + \lambda_{l_j} \sigma_{l_j}^2 + \lambda_{m_j} \sigma_{m_j}^2 \right)$$

Now, regarding the computation of E_{max}^i , we would like to get an upperbound on our uncertainty measure. We thus maximize all of the three different factors we are taking into account.

First, from theorem 2.3.1 of [15], entropy is upperbounded by $\log K$ where K is the number of elements for our sample space, and equal to 3 in our case. Hence, our maximal entropy is equal to $\log 3$.

Second, we must use an upperbound on the absolute traffic intensities value. Let us first define

$$\lambda_{i,max} = \max_{j:r_j \in R_i} \lambda_j \forall i \quad (4.20)$$

Then, $\forall i$,

$$(\lambda_{k_j} + \lambda_{l_j} + \lambda_{m_j}) \leq (\lambda_{i,max} + \lambda_{i,max} + \lambda_{i,max}) = 3\lambda_{i,max} \quad (4.21)$$

We thus choose $3\lambda_{i,max}$ as upperbound. Finally, defining

$$\sigma_{i,max}^2 = \max_{j:r_j \in R_i} \sigma_j^2 \forall i \quad (4.22)$$

we get

$$\bar{\sigma}_j^2 \leq \sigma_{i,max}^2 \quad (4.23)$$

and the upperbound is $\sigma_{i,max}^2 \forall i$. Putting everything together, the overall upperbound on uncertainty is

$$E_{max}^i = 3\lambda_{i,max}\sigma_{i,max}^2 \log 3 \quad (4.24)$$

Rewriting (4.17) and (4.18) and replacing E_j and E_{max}^i by their actual values,

$$\begin{aligned} u_n &= \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_j}{4} + (1-a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} \right] \\ &\quad - \gamma (F \cdot \vec{1})^T \vec{m} - \frac{(F \cdot \vec{1})^T \vec{\mu}_m}{(F \cdot \vec{1})^T 4\vec{\mu} - (F \cdot \vec{1})^T \vec{\mu}_m} \\ &= \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (\lambda_{k_j} + \lambda_{l_j} + \lambda_{m_j}) H(\alpha_j \lambda_{k_j}, \alpha_j \lambda_{l_j}, \alpha_j \lambda_{m_j}) \bar{\sigma}_j^2}{4} \right. \\ &\quad \left. + (1-a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (3\lambda_{i,max}\sigma_{i,max}^2 \log 3)}{4} \right] - \gamma (F \cdot \vec{1})^T \vec{m} - \frac{(F \cdot \vec{1})^T \vec{\mu}_m}{(F \cdot \vec{1})^T 4\vec{\mu} - (F \cdot \vec{1})^T \vec{\mu}_m} \end{aligned} \quad (4.25)$$

$$\begin{aligned} u_a &= \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (E_{max}^i - E_j)}{4} + (1-m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} \right] - \frac{\beta \sum_{i=1}^4 a_i}{4} \\ &= \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j \left(3\lambda_{i,max}\sigma_{i,max}^2 \log 3 - (\lambda_{k_j} + \lambda_{l_j} + \lambda_{m_j}) H(\alpha_j \lambda_{k_j}, \alpha_j \lambda_{l_j}, \alpha_j \lambda_{m_j}) \bar{\sigma}_j^2 \right)}{4} \right. \\ &\quad \left. + (1-m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (3\lambda_{i,max}\sigma_{i,max}^2 \log 3)}{4} \right] - \frac{\beta \sum_{i=1}^4 a_i}{4} \end{aligned} \quad (4.26)$$

Chapter 5

Game results

The payoff functions (4.25) and (4.26) we have defined in chapter 4 are complex. In order to get a first insight into the game's behavior (players' best responses and Nash equilibrium), these payoff functions need to be simplified. Therefore, we define a game G^e , where all mobility profile parameters are equal in all crossroads. This scenario, studied in section 5.1, is a special case of the general game G^g examined in section 5.2, after having got intuition about how the game behaved with equal parameters. For both cases, G^e and G^g , we will formally analyse the players' behavior, i.e. their best responses and possible Nash equilibria, check our analytical results with simulations and finally discuss them.

5.1 Equality game: G^e

$$f(\lambda, \sigma^2) \tag{5.1}$$

In this particular scenario, we first show that nodes' best response is defined by the number of mix zones the mobile nodes decide to deploy, regardless of their placement and of the adversary's strategy. Moreover, we prove that until adversary's cost does not go beyond a certain threshold, which is a function of the mobility parameters, attacker's best response is to deploy his eavesdropping stations wherever there is no mix zone. Finally, we show the existence of at least one Nash equilibrium in this specific game G^e .

Assuming all mobility profile parameters equal, we have $\lambda_i = \lambda$, $\sigma_i^2 = \sigma^2$, and

$\mu_i = \mu$ for all i . Note that, in the following, we have:

$$\begin{cases} \lambda_{i,max} = \lambda \quad \forall i \\ \sigma_{i,max}^2 = \sigma^2 \quad \forall i \\ \alpha_j = \frac{1}{3\lambda} \quad \forall j \\ \bar{\sigma}_j^2 = \sigma^2 \quad \forall j \end{cases}$$

Replacing these values in equations (4.25) and (4.26), we can simplify nodes' and adversary's payoff functions (the derivations are presented in A.1):

$$u_n = \sum_{i=1}^4 m_i \left[3\lambda^2 \sigma^2 \log 3 - 4\gamma\lambda - \frac{1}{16 - \sum_{j=1}^4 m_j} \right] \quad (5.2)$$

$$u_a = \sum_{i=1}^4 a_i \left[3\lambda^2 \sigma^2 \log 3 (1 - m_i) - \frac{\beta}{4} \right] \quad (5.3)$$

5.1.1 Nodes' best response

We first state an important theorem defining the mobile nodes' best response, which depends only on the number of mix zones, $N_n = \sum_i m_i$, they deploy and not on their particular placement.

Theorem 5.1. *The nodes' best response is independent of the adversary's strategy, depending only on the mobility parameters λ , σ^2 and the cost of a pseudonym γ . More formally, the nodes' best response is expressed as the number of deployed mix zones*:*

$$s_n^* = N_n^* = br_n(\lambda, \sigma^2, \gamma) = \left\lceil 16 - \frac{4}{\sqrt{3\lambda^2 \sigma^2 \log 3 - 4\gamma\lambda}} \right\rceil \quad (5.4)$$

Note that N_n must be between 0 and 4.

Proof. First of all, as the parameters within the sum do not depend on i , we can slightly change the payoff function, by expressing the sum over m_i 's as one single variable x , which is the number of mix zones the nodes deploy:

$$u_n = x \left(3\lambda^2 \sigma^2 \log 3 - 4\gamma\lambda - \frac{1}{16 - x} \right) \quad (5.5)$$

Maximizing (5.5) with respect to x , results in:

$$\frac{\partial u_n}{\partial s_n} = \frac{\partial u_n}{\partial x} = 3\lambda^2 \sigma^2 \log 3 - 4\gamma\lambda - \frac{1}{16 - x} - \frac{x}{(16 - x)^2} = 0 \quad (5.6)$$

*[.] represents the nearest integer function.

Rearranging the terms of (5.6), the number of mix zones that maximizes the payoff can be computed by solving the following quadratic expression:

$$x^2 - 32x + 16^2 - \frac{16}{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda} = 0 \quad (5.7)$$

Equation (5.7) has two solutions: $x_{1,2}^* = 16 \pm \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}}$. The only acceptable solution is $x_2^* = 16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}}$. Indeed, $x_1^* = 16 + \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} > 16$ cannot be a maximum since the second derivative of u_n is equal to $-\frac{32}{(16-x)^3} > 0$ when $x > 16$, which shows that it is a local minimum. \square

In other words, since all the traffic intensities of entries at each crossroads are equal, entropy in the uncertainty measure is maximal. Therefore, mobile nodes do not care about sniffing stations' placement since, even with an eavesdropping station at the same intersection, a mix zone would still mix nodes with high efficiency. Moreover, since all crossroads have the same uncertainty value, the nodes' payoff function does not depend on the placement of mix zones, but only on their number N_n .

Corollary 5.1. *We notice in theorem 5.1 that nodes' best response is independent of the mix zones' location and, thus, depends only on the number of mix zones we deploy. Therefore, $\forall s_a \in \mathcal{S}_a$, the number of nodes' best responses is equal to all the possible combinations of N_n^* mix zones amid 4 crossroads:*

$$\binom{4}{N_n^*} = \begin{cases} 1 & \text{if } N_n^* = 0 \\ 4 & \text{if } N_n^* = 1 \\ 6 & \text{if } N_n^* = 2 \\ 4 & \text{if } N_n^* = 3 \\ 1 & \text{if } N_n^* = 4 \end{cases} \quad (5.8)$$

We are now able to write up nodes' best response, i.e. the best number of deployed mix zones, using threshold values on the pseudonym price γ , on the traffic intensity λ or on the sojourn time variance σ^2 . It is interesting to express the nodes' best strategy with respect to those mobility and cost parameters since mobile users' decision (to deploy a mix zone or not) is directly based on them. Explanations on how we derive these results can be found in A.1.1.

First, we express N_n^* with fixed λ and σ^2 , and varying γ :

$$s_n^* = N_n^*(\gamma) = \begin{cases} 0 & \text{if } \gamma > \gamma_1 \\ \left[16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} \right] & \text{if } \gamma_1 \leq \gamma \leq \gamma_2 \\ 4 & \text{if } \gamma < \gamma_2 \end{cases} \quad (5.9)$$

where $\gamma_1 = \frac{3\lambda^2\sigma^2 \log 3 - (\frac{8}{31})^2}{4\lambda}$ and $\gamma_2 = \frac{3\lambda^2\sigma^2 \log 3 - (\frac{8}{25})^2}{4\lambda}$.

This result is the most important of our thresholds since it allows us to compare cost led by deployment of a mix zone with respect to mobility parameters. Since nodes change their pseudonyms whenever they go through a mix zone, deployment of a mix zone has a cost directly related to the price γ of a pseudonym. Mobile nodes are then able to base their strategic decision, to add a mix zone or not, on the values of γ . We furthermore notice that bounds on γ are growing linearly with respect to σ^2 but not with respect to λ .

We can also express mobile nodes' best response with respect to bounds on mobility parameters λ and σ^2 .

For fixed values of γ and σ^2 , nodes' best response expressed as a function of λ is

$$s_n^* = N_n^*(\lambda) = \begin{cases} 0 & \text{if } \lambda < \lambda_1 \\ \left[16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} \right] & \text{if } \lambda_1 \leq \lambda \leq \lambda_2 \\ 4 & \text{if } \lambda > \lambda_2 \end{cases} \quad (5.10)$$

where $\lambda_1 = \frac{2\gamma + \sqrt{4\gamma^2 + 3\sigma^2 \log 3 (\frac{8}{31})^2}}{3\sigma^2 \log 3}$ and $\lambda_2 = \frac{2\gamma + \sqrt{4\gamma^2 + 3\sigma^2 \log 3 (\frac{8}{25})^2}}{3\sigma^2 \log 3}$. In addition, for fixed values of γ and λ , and varying σ^2 ,

$$s_n^* = N_n^*(\sigma^2) = \begin{cases} 0 & \text{if } \sigma^2 < \sigma_1^2 \\ \left[16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} \right] & \text{if } \sigma_1^2 \leq \sigma^2 \leq \sigma_2^2 \\ 4 & \text{if } \sigma^2 > \sigma_2^2 \end{cases} \quad (5.11)$$

where $\sigma_1^2 = \frac{(\frac{8}{31})^2 + 4\gamma\lambda}{3\lambda^2 \log 3}$ and $\sigma_2^2 = \frac{(\frac{8}{25})^2 + 4\gamma\lambda}{3\lambda^2 \log 3}$.

We remark that the greater is the cost γ , the greater have to be λ and σ^2 to prompt mobile nodes to deploy mix zones. In both cases, λ and σ^2 must grow linearly with respect to γ to keep the same best response.

5.1.2 Adversary's best response

The second main result we derived is related to the adversary's best response. We first explain it briefly and write it up more formally in theorem 5.2.

The adversary's best response behaves in two different ways, depending on whether the benefit (function of λ and σ^2) is greater than sniffing stations' cost β or not. In the latter case, his best response is to deploy no eavesdropping stations, independently of the mobile nodes' strategy. Since benefits are too small with respect to costs, the adversary gets no incentive to sniff the network, regardless of what mobile nodes do.

In the other case, when λ or σ^2 are high enough with respect to β , the adversary's best response is to deploy its sniffing stations at every crossroads

where the mobile nodes have not placed a mix zone. If the adversary placed a sniffing station at the same intersection as a mix zone, then high nodes' mixing effectiveness due to uncertainty would prevent him to make any benefit. To summarize, the adversary adopts a complementary strategy to the mobile nodes' one.

We provide now the detailed steps for the adversary's best response derivation. We also clarify and formally define the threshold on cost that influences his best response. As the adversary's best response depends on nodes' strategy, we first consider two extreme scenarios in the following lemmas.

Lemma 5.1. *If $\vec{m} = [0 \ 0 \ 0 \ 0]^T = s_n^1$, then $u_a = N_a \left[3\lambda^2\sigma^2 \log 3 - \frac{\beta}{4} \right]$ where $N_a = \sum_i a_i$. Therefore, u_a is a linear function of the number of sniffing stations and the best response is:*

$$s_a^* = N_a^* = br_a(\lambda, \sigma^2, \beta) = \begin{cases} 4 & \text{if } 3\lambda^2\sigma^2 \log 3 > \frac{\beta}{4} \\ 0 & \text{if } 3\lambda^2\sigma^2 \log 3 \leq \frac{\beta}{4} \end{cases} \quad (5.12)$$

Proof. Replacing the m_i 's by 0 in (5.3), we get:

$$\sum_{i=1}^4 a_i \left[3\lambda^2\sigma^2 \log 3 - \frac{\beta}{4} \right] = N_a \left[3\lambda^2\sigma^2 \log 3 - \frac{\beta}{4} \right] \quad (5.13)$$

As (5.13) is a linear function of N_a , we just care about whether its multiplicative constant is negative or positive. If it is strictly positive, then u_a is increasing linearly with N_a and so $N_a^* = 4$. If $(3\lambda^2\sigma^2 \log 3 - \frac{\beta}{4})$ is negative, the adversary's payoff is decreasing linearly with N_a . The best response is then $N_a^* = 0$. \square

This theorem shows us the importance of cost β for the adversary to get his best response. With respect to the bound on β , he can change his best response from four sniffing stations (if cost is cheap) to none of them (if cost is expensive).

Lemma 5.2. *If $\vec{m} = [1 \ 1 \ 1 \ 1]^T = s_n^{16}$, then $u_a = -\frac{\beta}{4}N_a$. Since $\beta > 0$, the best response is:*

$$s_a^* = N_a^* = br_a = 0 \quad (5.14)$$

Proof. Replacing the m_i 's by 1 in (5.3), we get:

$$\sum_{i=1}^4 a_i \left[0 - \frac{\beta}{4} \right] = -\frac{\beta}{4}N_a \quad (5.15)$$

As (5.15) is decreasing with N_a , the best response obviously is $N_a^* = 0$. \square

Using results from lemmas 5.1 and 5.2, we can now state the theorem on the adversary's best response, with respect to a threshold defined by the mobility parameters λ and σ^2 and the adversary's cost β .

Theorem 5.2. *For fixed values of λ and σ^2 , the adversary's best response is a function of β :*

$$s_a^* = br_a(s_n) = \vec{a}^*(\beta) \quad (5.16)$$

where

$$a_i^* = \begin{cases} 1 - m_i & \text{if } \beta < \beta_1 \\ 0 & \text{if } \beta \geq \beta_1 \end{cases} \quad (5.17)$$

where $\beta_1 = 12\lambda^2\sigma^2 \log 3$.

Proof. Rewriting (5.3) and sorting terms with respect to nodes' strategy, we get:

$$u_a = (3\lambda^2\sigma^2 \log 3 - \frac{\beta}{4}) \sum_{i:m_i=0} a_i - \frac{\beta}{4} \sum_{j:m_j=1} a_j \quad (5.18)$$

From lemma 5.2, we know that the adversary does not get any incentive to place a sniffing station where there already is a mix zone. Formally, $\forall j$ s.t. $m_j = 1$, $a_j^* = 0$. Using results of lemma 5.1, we can assert that $\forall i$ s.t. $m_i = 0$, $a_i^* = 1$ if $3\lambda^2\sigma^2 \log 3 > \frac{\beta}{4}$, i.e. $\beta < 12\lambda^2\sigma^2 \log 3$ and, $a_i^* = 0$ if $3\lambda^2\sigma^2 \log 3 \leq \frac{\beta}{4}$, i.e. $\beta \geq 12\lambda^2\sigma^2 \log 3$. \square

5.1.3 Nash equilibrium

We are now able to state a theorem about the Nash equilibria.

Theorem 5.3. *For given λ and σ^2 in game G^e , there always exists at least one Nash equilibrium. However, uniqueness is not always satisfied and depends on γ and β .*

Proof. Considering best response results of theorems 5.1 and 5.2, corollary 5.1 and equation (5.9), we can derive NE cases for all bounds on γ and β with respect to γ_1 , γ_2 and β_1 . We provide here an example.

If $\gamma_1 \leq \gamma \leq \gamma_2$ and $N_n^* = 2$ following theorem 5.1, there are 6 Nash equilibria according to corollary 5.1 and two groups of Nash equilibria can be identified. If $\beta > \beta_1$, the Nash equilibria are all the possible combinations of 2 mix zones among 4 crossroads (there are 6 combinations) for the nodes' strategies, and no sniffing station for the adversary. If $\beta \leq \beta_1$, the Nash equilibria are still all the combinations of 2 mix zones among 4 intersections for the mobile nodes, and, for the adversary's strategies, placement of eavesdropping stations wherever there is no mix zone yet (complementary deployment).

Note finally that number of Nash equilibria depend on number of nodes' best responses defined in corollary 5.1. For instance, if $N_n^* = 0$ or $N_n^* = 4$, there is one and only one Nash equilibrium. \square

What we can notice regarding the Nash equilibria is that, first, the number of Nash equilibria is growing like the number of nodes' best responses. Thus, when best number of mix zones' is equal to 1, 2 or 3, we get multiple Nash equilibria. On the other hand, as the adversary's best response varies with respect to the threshold on β , being either to place no eavesdropping station or the complementary of mix zones' placement, the Nash equilibria vary as well with respect to this threshold. We provide some examples in the following in order to show how best responses and Nash equilibria are changing with respect to mobility and cost parameters.

5.1.4 Discussion and simulations

In this subsection, we do simulations using Matlab in order to verify whether our analytical derivations are correct or not. We then discuss our simulated results with respect to the formal ones.

We provide below some plots showing both players' best responses and Nash equilibria on the same graph. We first simulate with variable values of λ and other parameters fixed. Afterwards, we display game's behavior with respect to varying values of β . Both of these simulations encompass more or less all interesting results and behaviors that we could expect.

Let us set up our fixed values in the first scenario. Regarding the mobility parameters, as we said, traffic intensity λ is varying and we fix the sojourn time variance to $\sigma^2 = 0.5s$. We also fix the players' cost: adversary's cost of sniffing stations management $\beta = 1$ (thus, one sniffing station costs 1/4), and mobile nodes' cost of one pseudonym $\gamma = 0.1$. We display some of the results for $\lambda \in (0, \infty)$ in the following 3 figures. Note that $s_n^1 = [0 \ 0 \ 0 \ 0]^T$ is the nodes' strategy to deploy no mix zone and $s_n^{16} = [1 \ 1 \ 1 \ 1]^T$ is the nodes' strategy to activate four mix zones (all of them). Strategies s_n^2 to s_n^5 represent the use of one mix zone, s_n^6 to s_n^{11} , two mix zones, and s_n^{12} to s_n^{15} three mix zones. In the same way, $s_a^1 = [0 \ 0 \ 0 \ 0]^T$ is the adversary's strategy to place no eavesdropping station and $s_a^{16} = [1 \ 1 \ 1 \ 1]^T$ is the adversary's strategy to place sniffing stations at all crossroads. Strategies s_a^2 to s_a^5 represent the placement of one eavesdropping station, s_a^6 to s_a^{11} , of two eavesdropping stations, and s_a^{12} to s_a^{15} of three eavesdropping stations.

For $\lambda \in (0, 0.271]$, nodes' and adversary's best responses are to deploy nothing, neither mix zones nor sniffing stations (Figure 5.1). 0.271 should correspond to threshold λ_1 we have defined after equation 5.10. Using our numerical values, we get a theoretical threshold of $\lambda_1 = 0.2714$, which matches very well with our simulation result. When $\lambda \in [0.272, 0.281]$, nodes get an incentive to place one mix zone, at one of the crossroads. Adversary's best response does

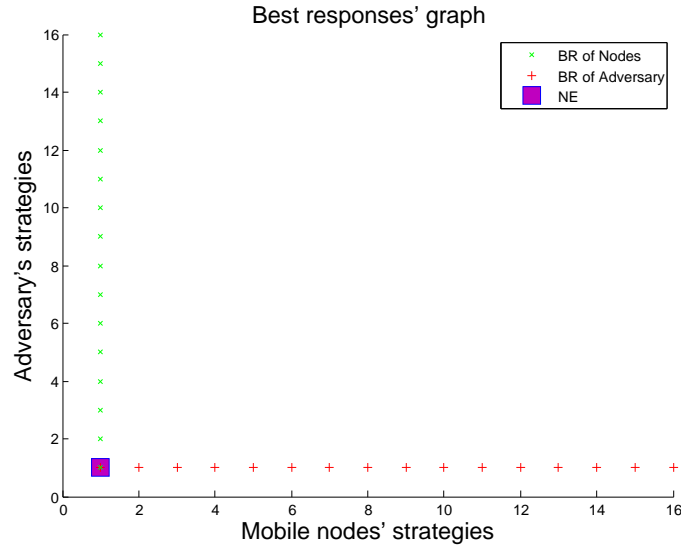


Figure 5.1: Simulation result with $\sigma^2 = 0.5s$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda \in (0, 0.271]$, nodes' best response is to deploy no mix zone and adversary's best response is to place no sniffing station. Thus, we get one Nash equilibrium, at (s_n^1, s_a^1) .

not change. If $\lambda \in [0.282, 0.293]$, nodes increase their mix zones number to two. Nodes' best responses are thus all possible combinations of two mix zones among four crossroads (Figure 5.2). Adversary's best response remains to do nothing. For $\lambda \in [0.294, 0.308]$, nodes place three mix zones and adversary still does nothing. 0.308 represent here threshold λ_2 defined after equation 5.10. After replacing our numerical values in the formal expression, we find $\lambda_2 = 0.3081$, which verifies the correspondence between our analytical and simulated results. If $\lambda \in [0.309, 0.324]$, mobile nodes place a mix zone at every crossroads and adversary's best response does not change. Finally, for $\lambda \in [0.325, \infty)$, nodes' best response is still to deploy four mix zones but, adversary's one is to sniff crossroads wherever there is no mix zone yet (Figure 5.3).

The second scenario fixes mobility parameters $\lambda = 0.3$ and $\sigma^2 = 0.5s$, as well as pseudonym's cost $\gamma = 0.1$, and makes β vary from zero to infinity. In this case, game behaves like in figure 5.4 for $\beta \in (0, 0.856)$, and like in figure 5.5 for $\beta \in [0.856, \infty)$. In the latter case, as adversary's cost β is quite big, attacker's best response is to place no eavesdropping stations. Mobile nodes' best response is to deploy three mix zones since mobility parameters λ and σ^2 are high enough with respect to cost γ . In the second case, when β is small, then adversary's best response is to place sniffing stations where there is no mix zone yet.

We would like now to compare these empirical results with the analytical ones we get in previous subsections. We can notice first that Figure 5.1 represents the

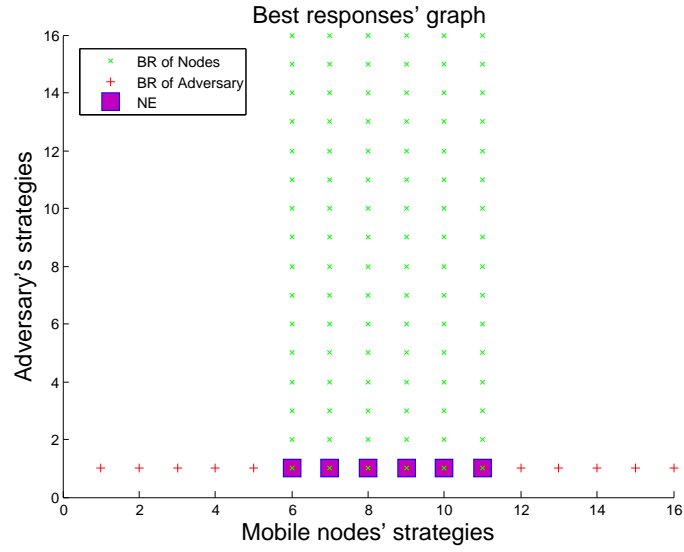


Figure 5.2: Simulation result with $\sigma^2 = 0.5s$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda \in [0.282, 0.293]$, nodes' best responses are to deploy two mix zones and adversary's best response is to place no sniffing station. Thus, we get six Nash equilibria, at (s_n^6, s_a^1) , (s_n^7, s_a^1) , (s_n^8, s_a^1) , (s_n^9, s_a^1) , (s_n^{10}, s_a^1) and (s_n^{11}, s_a^1) .

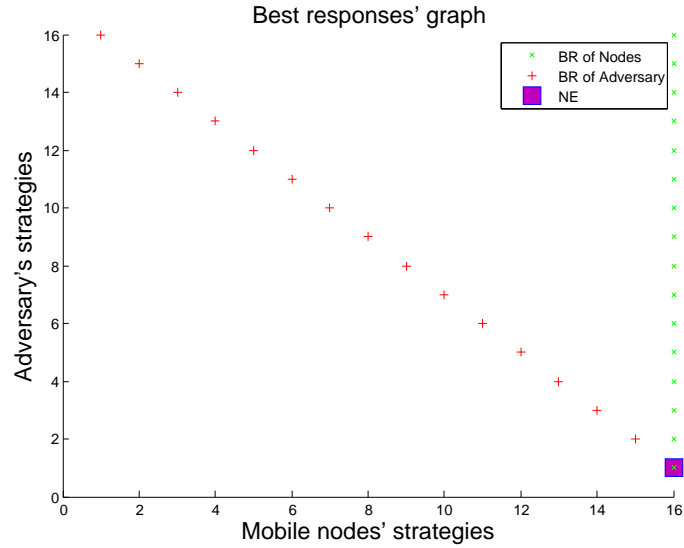


Figure 5.3: Simulation result with $\sigma^2 = 0.5s$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda \in [0.325, \infty[$, nodes' best response is to deploy four mix zones and adversary's best response is to place sniffing stations where there is no mix zone. Thus, we get one Nash equilibrium, at (s_n^{16}, s_a^1) .

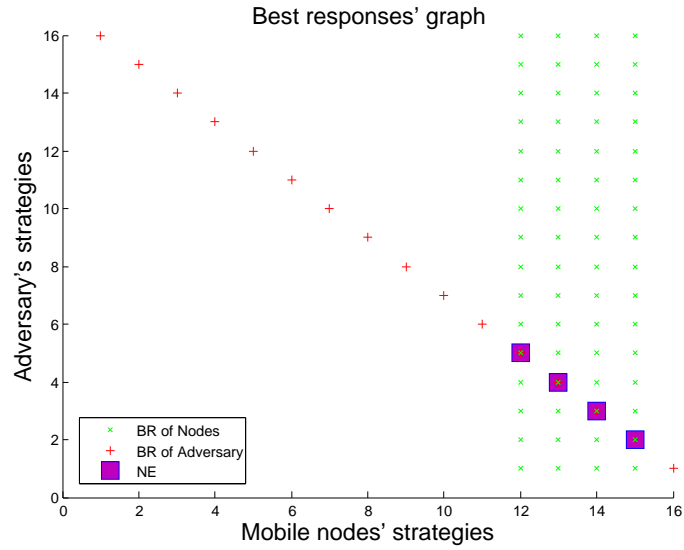


Figure 5.4: Simulation result with $\sigma^2 = 0.5s$, $\lambda = 0.3$ and $\gamma = 0.1$. For $\beta \in (0, 0.856)$, nodes' best response is to deploy three mix zones (regardless of values of β) and adversary's best response is to place sniffing stations where there is no mix zone. Thus, we get four Nash equilibria, at (s_n^{12}, s_a^5) , (s_n^{13}, s_a^4) , (s_n^{14}, s_a^3) and (s_n^{15}, s_a^2) .

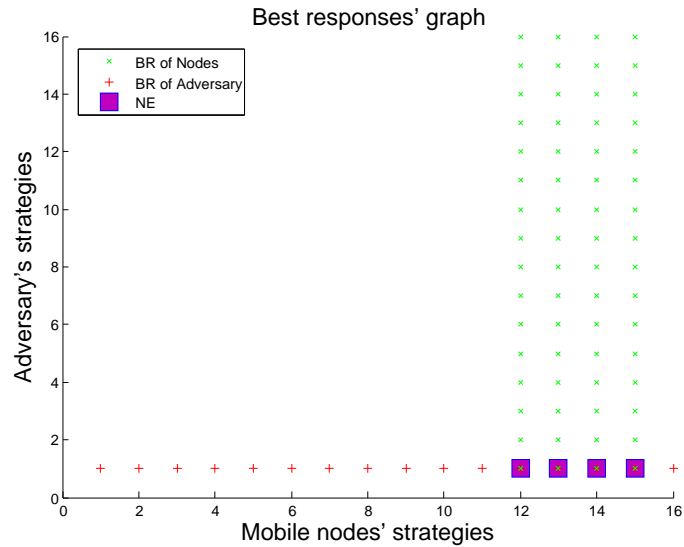


Figure 5.5: Simulation result with $\sigma^2 = 0.5s$, $\lambda = 0.3$ and $\gamma = 0.1$. For $\beta \in (0.856, \infty)$, nodes' best response is to deploy three mix zones (regardless of values of β) and adversary's best response is to place no sniffing stations. Thus, we get four Nash equilibria, at (s_n^{12}, s_a^1) , (s_n^{13}, s_a^1) , (s_n^{14}, s_a^1) and (s_n^{15}, s_a^1) .

Nash equilibrium (s_n^1, s_a^1) . So let us check if the empirical values we use to plot this first graph correspond to the conditions on γ and β defined in equations (5.9) and (5.17). For the values used in Figure 5.1, we have $\gamma_1 \in (0, 0.0996]$ and γ_2 even smaller. Hence, as $\gamma = 0.1$, it is greater than γ_1 . Moreover, $\beta_1 \in (0, 0.6984]$, which means that, since $\beta = 1$, it is also greater than threshold on β_1 . Since we get the Nash equilibrium coming from the same best responses, this shows that analytical and empirical results match very well. We check correspondence between analytical Nash equilibria and simulations in the same way for other figures (5.2, 5.3, 5.4 and 5.5), leading us to conclude to the accuracy of Nash equilibria and best responses derived in game G^e .

This special scenario when mobility profile of the network is totally homogeneous helps us to get a good feeling of how the more general game G^g might behave. We can be quite confident of the existence of Nash equilibria in G^g as well. That is what we are going to look at now.

5.2 General game: G^g

Even though general game G^g has complex expressions of both mobile nodes' and adversary's payoffs, we manage to find some interesting results. We show that strategic decision of the adversary to place an eavesdropping station at crossroads C_i is dependent on mobility parameters at this crossroads, and only dependent on these parameters. In other words, the adversary's decision to place a sniffing station at C_i is independent of parameters at C_j , for all $j \neq i$ (others than those coming from crossroads C_i). Moreover, the adversary's global best response is equivalent to the union of all local best responses (at each crossroads). In contrary, nodes' decision to deploy a mix zone at C_i does depend on other placements of mix zones and on mobility parameters at different intersections. Mobile nodes cannot split their strategic problem into smaller local ones at intersections and have to take it as a whole. Finally, even though we are not able to prove it formally, we reach one or more Nash equilibria in all our simulations.

5.2.1 Adversary's best response

The main interesting result of G^g game is provided by adversary's best response. First, the decision to place a sniffing station or not at some crossroads to get a best response is independent of other crossroads mobility parameters. Second, the global adversary's best response is equal to the union of the local best responses at each crossroads. In order to show that, let us first rewrite equation (4.26) in

its compact form:

$$u_a = \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (E_{max}^i - E_j)}{4} + (1 - m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} \right] - \frac{\beta \sum_i a_i}{4} \quad (5.19)$$

We can then include the cost within the parenthesis and bring out a_i :

$$u_a = \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (E_{max}^i - E_j)}{4} + (1 - m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} - \frac{\beta}{4} \right] \quad (5.20)$$

After a few steps, we can simplify our payoff and keep m_i in front of only one summation over j :

$$u_a = \sum_{i=1}^4 a_i \left[\sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} - m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_j}{4} - \frac{\beta}{4} \right] \quad (5.21)$$

Note that (5.21) is linear with respect to the a_i 's. Therefore, the adversary, in order to get his best response, is just concerned by the sign of the multiplicative factor

$$\sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} - m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_j}{4} - \frac{\beta}{4} \quad (5.22)$$

In order to be beneficial to the adversary, this equation must be positive. We can now derive a theorem on the adversary's best response.

Theorem 5.4. *The adversary's best response is*

$$s_a^* = br_a(s_n) = \vec{a}^* \quad (5.23)$$

where

$$a_i^* = \begin{cases} 0 & \text{if } \sum_{j=4(i-1)}^{4i-1} \lambda_j E_{max}^i < \beta \\ 1 - m_i & \text{if } \beta < \sum_{j=4(i-1)}^{4i-1} \lambda_j E_{max}^i < \sum_{j=4(i-1)}^{4i-1} \lambda_j E_j + \beta \\ 1 & \text{if } \sum_{j=4(i-1)}^{4i-1} \lambda_j E_{max}^i > \sum_{j=4(i-1)}^{4i-1} \lambda_j E_j + \beta \end{cases} \quad (5.24)$$

Proof. All thresholds in (5.24) follow from sign evaluation of multiplicative factor (5.22). \square

Developing E_{max}^i in the first condition, we get the following threshold on β :

$$\sum_{j:r_j \in R_i} \lambda_j 3 \lambda_{i,max} \sigma_{i,max}^2 \log 3 < \beta \quad (5.25)$$

or

$$\lambda_{i,max}\sigma_{i,max}^2\bar{\lambda}_i < \frac{\beta}{3\log 3} \quad (5.26)$$

where $\bar{\lambda}_i$ is the sum of all traffic intensities entering the intersection C_i . Hence, if the multiplication of the maximal traffic intensity, the maximal delay variance and the sum of traffic intensities entering the intersection is too small with respect to the sniffing station's cost, it prompts the adversary to do nothing, to deploy no sniffing station at the specific crossroads C_i . In addition to the main results described at the beginning of the section, theorem 5.4 also shows us that the adversary's best response depends on the mobile nodes' strategy only under a certain condition (second case of equation (5.24)). Under this condition, the adversary's best response is to place a sniffing station if there is no mix zone at the same intersection, just like in theorem 5.2 of section 5.1.

5.2.2 Nodes' best response

Mobile nodes' payoff function (4.25) is very complex to analyse in the general case. In the following, we attempt to provide the simplest possible expression of u_n and get some interesting results. However, due to high complexity of the derivative of u_n with respect to s_n , analytical derivation of mobile nodes' best response is not feasible in this report. We will thus use simulations to see how nodes behave in game G^g . To show why this payoff function is too complicated to be formally analysed, let us first rewrite equation (4.25) in its compact form:

$$u_n = \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_j}{4} + (1 - a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} \right] - \gamma(F \cdot \vec{1})^T \vec{m} - \frac{(F \cdot \vec{1})^T \vec{\mu}_m}{(F \cdot \vec{1})^T 4\vec{\mu} - (F \cdot \vec{1})^T \vec{\mu}_m} \quad (5.27)$$

We then express costs as summations instead of matrices operations:

$$u_n = \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_j}{4} + (1 - a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j E_{max}^i}{4} \right] - \gamma \sum_{i=1}^4 m_i \sum_{j=4(i-1)}^{4i-1} \lambda_j - \frac{\sum_{i=1}^4 \mu_i m_i \sum_{j=4(i-1)}^{4i-1} \lambda_j}{4 \sum_{i=1}^4 \mu_i \sum_{j=4(i-1)}^{4i-1} \lambda_j - \sum_{i=1}^4 \mu_i m_i \sum_{j=4(i-1)}^{4i-1} \lambda_j} \quad (5.28)$$

We can now bring m_i out and keep a_i in front of only one summation over j :

$$u_n = \sum_{i=1}^4 m_i \left[\frac{\bar{\lambda}_i E_{max}^i}{4} - a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda_j (E_{max}^i - E_j)}{4} - \gamma \bar{\lambda}_i - \frac{\mu_i \bar{\lambda}_i}{4 \sum_{k=1}^4 \bar{\lambda}_k \mu_k - \sum_{k=1}^4 \bar{\lambda}_k \mu_k m_k} \right] \quad (5.29)$$

where $\bar{\lambda}_i$ is the sum of all traffic intensities entering the intersection C_i . We notice in this last equation that u_n is not linear with respect to the m_i 's since the m_i 's are also included in the denominator of last cost. Hence, it is very complicated to get the best response using derivatives with respect to the m_i 's and we would rather simulate the game in order to see how the mobile nodes behave in the general game. Before simulating, we can already assert that mobile nodes' best response at intersection C_i must depend on mobility parameters and costs at other intersections.

5.2.3 Discussion and simulations

We provide here simulations results first, and then discuss them and verify whether they seem to match with analytical results or not. We provide in the following some plots showing, as in section 5.1, both players' best responses and Nash equilibria on the same graph. We first vary λ_0 , which represents the traffic intensity of the road in the North of the first intersection. We fix other mobility and cost parameters:

$$\begin{cases} \lambda_i = 0.2 \quad \forall i \neq 0 \\ \mu_i = 2 \quad \forall i \\ \sigma_i^2 = 0.5 \quad \forall i \\ \beta = 1 \\ \gamma = 0.1 \end{cases} \quad (5.30)$$

We display the results for $\lambda_0 \in (0, \infty)$ in the following 5 figures. For $\lambda_0 \in (0, 0.304]$, nodes' and adversary's best responses are to deploy nothing, neither mix zones nor eavesdropping stations (Figure 5.6). We thus observe a unique Nash equilibrium, at (s_n^1, s_a^1) . Even though λ_0 goes beyond 0.2, neither the mobile nodes nor the adversary get enough benefit to deploy a mix zone or a sniffing station. For $\lambda_0 \in [0.305, 0.414]$, mobile nodes get incentive to deploy a mix zone at the first intersection if the adversary's strategy is to do nothing (Figure 5.7). The nodes' best response depends on the adversary's strategy a_1 and shows us that, contrary to game G^e , in game G^g , the nodes' best strategy is not independent of the attacker's strategy. Note that the Nash equilibrium here is to deploy zero eavesdropping station (s_a^1) and one mix zone at intersection 1 (s_n^2). For $\lambda_0 \in [0.415, 0.595]$, nodes' best response remains as in previous case but the attacker's best response changes from doing nothing to place an eavesdropping device at C_1 whether there is no mix zone yet (Figure 5.8). The players' best responses are thus symmetric and the game reaches two Nash equilibria, at (s_n^1, s_a^2) and (s_n^2, s_a^1) . From $\lambda_0 = 0.596$, the adversary changes again his best response to the deployment of one sniffing station at intersection 1, regardless of the mobile nodes' strategy (Figure 5.9). We then get a Nash equilibrium at (s_n^1, s_a^2) . To be complete, note that for values of λ_0 greater than 4, the mobile nodes' best

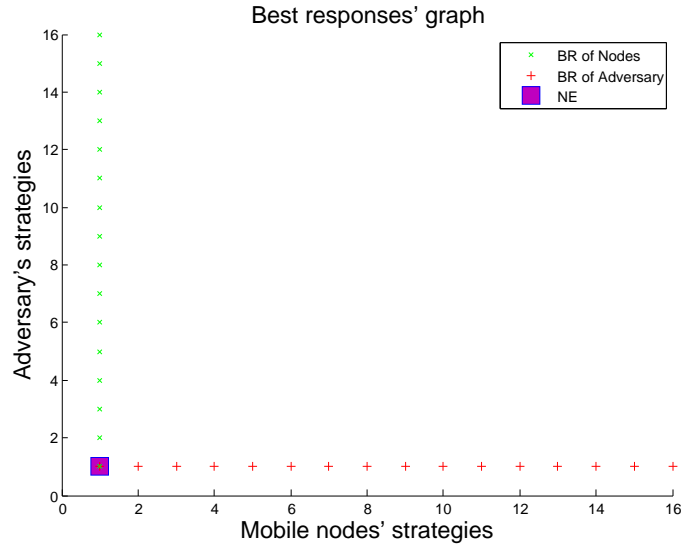


Figure 5.6: Simulation result with $\lambda_i = 0.2 \forall i \neq 0$, $\mu_i = 2$ and $\sigma_i^2 = 0.5s \forall i$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda_0 \in (0, 0.304]$, nodes' best response is to deploy no mix zone and adversary's best response is to place no sniffing station. Thus, we get one Nash equilibrium, at (s_n^1, s_a^1) .

response is also to deploy a mix zone at C_1 , regardless of the adversary's strategy (Figure 5.10). However, first, in reality traffic intensities are never so high and, second, if 4 nodes per second enter through one road, it is completely improbable to have only 0.6 leaving nodes per second in total via the three other roads.

Note that, for the four realistic ranges of λ_0 values, the game's equilibria evolve, for increasing λ_0 , in an interesting way: first, Nash equilibrium is (s_n^1, s_a^1) ; then it is (s_n^2, s_a^1) ; there are next (s_n^1, s_a^2) and (s_n^2, s_a^1) ; and finally (s_n^1, s_a^2) . Even though, at the beginning, mix zones seem to be more "powerful" than eavesdropping stations (when $NE = (s_n^2, s_a^1)$), adversary's devices become the ones remaining at equilibrium when λ_0 reaches 0.595 (s_n^1, s_a^2) , with a transition phase in-between (with both NE).

We would like now to verify whether thresholds derived in (5.24) for the adversary's best response match with the simulations. By replacing the numerical values we have fixed for our simulations in (5.26), we get, for $\lambda_0 = 0.414$, $3 \log 3\lambda_{1,max}\sigma_{1,max}^2\bar{\lambda}_1 = 0.998 < \beta = 1$ which matches well with the adversary's best response shown in figures 5.6 and 5.7 ($s_a^* = s_a^1$). If λ_0 becomes equal to 0.415, $3 \log 3\lambda_{1,max}\sigma_{1,max}^2\bar{\lambda}_1 = 1.0014 > \beta = 1$ and the adversary places a sniffing station if there is no mix zone yet, like in figure 5.8. We verify in the same way the upper threshold in (5.24), which matches with results shown in figure 5.9 ($s_a^* = s_a^2$).

We can extend the results achieved for C_1 to other crossroads C_2 , C_3 and C_4 ,

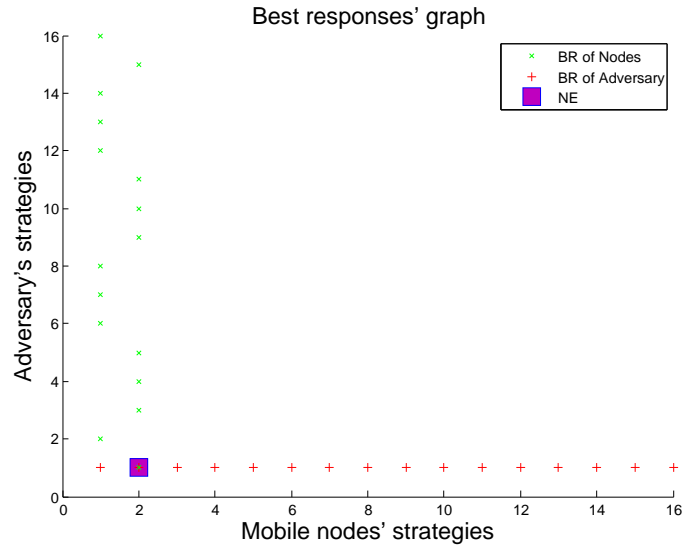


Figure 5.7: Simulation result with $\lambda_i = 0.2 \forall i \neq 0$, $\mu_i = 2$ and $\sigma_i^2 = 0.5s \forall i$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda_0 \in [0.305, 0.414]$, nodes' best response is to deploy one mix zone at C_1 if there is no sniffing station yet, and adversary's best response is to place no sniffing station. Thus, we get one Nash equilibrium, at (s_n^2, s_a^1) .

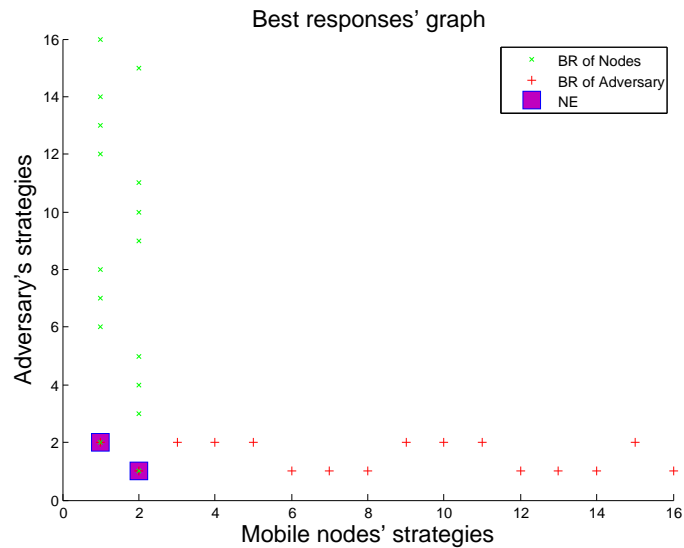


Figure 5.8: Simulation result with $\lambda_i = 0.2 \forall i \neq 0$, $\mu_i = 2$ and $\sigma_i^2 = 0.5s \forall i$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda_0 \in [0.415, 0.595]$, nodes' best response is to deploy one mix zone at C_1 if there is not a sniffing station yet, and adversary's best response is to place one sniffing station at C_1 if there is not a mix zone yet. Thus, we get two Nash equilibria, at (s_n^2, s_a^1) and (s_n^1, s_a^2) .

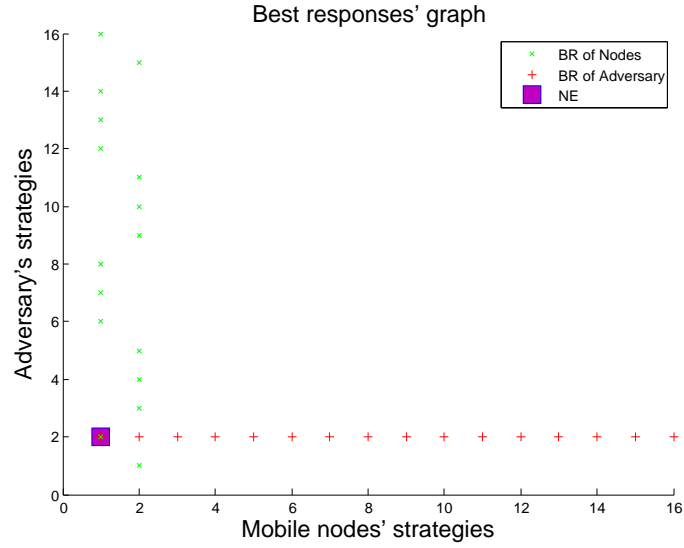


Figure 5.9: Simulation result with $\lambda_i = 0.2 \forall i \neq 0$, $\mu_i = 2$ and $\sigma_i^2 = 0.5s \forall i$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda_0 \in [0.596, 4.499]$, nodes' best response is to deploy one mix zone at C_1 if there is not already a sniffing station, and adversary's best response is to place one sniffing station at C_1 , regardless of the nodes' strategy. Thus, we get one Nash equilibrium, at (s_n^1, s_a^2) .

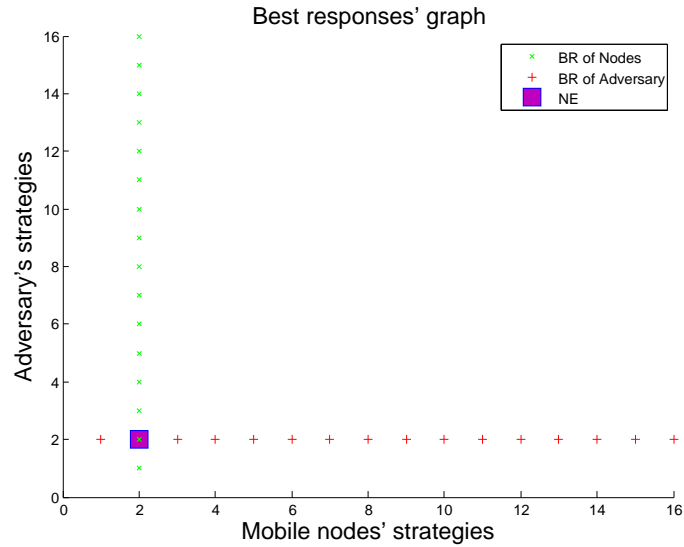


Figure 5.10: Simulation result with $\lambda_i = 0.2 \forall i \neq 0$, $\mu_i = 2$ and $\sigma_i^2 = 0.5s \forall i$, $\beta = 1$ and $\gamma = 0.1$. For $\lambda_0 \in [4.5, \infty)$, nodes' best response is to deploy one mix zone at C_1 , regardless of the adversary's strategy, and adversary's best response is to place one sniffing station at C_1 , regardless of the nodes' strategy. Thus, we get one Nash equilibrium, at (s_n^2, s_a^2) .

at least for the adversary's strategic decisions since these decisions are dependent only on the specific intersection we are looking at. Moreover, as we have already seen, the global best response of the adversary is the union of the local ones.

Finally, we simulate for other varying values of λ_i 's, $i \neq 0$. We notice that, when changing values of λ_i 's for different crossroads than the first one, it changes the threshold values on λ_0 for which the mobile nodes' best response is to deploy a mix zone at the first intersection or not. This shows that our claim at the beginning of section 5.2, that the nodes' decision to deploy a mix zone at a particular intersection depends on mobility parameters at other intersections, is correct.

We have not simulated for other varying mobility profile parameters, such as sojourn time variables σ_i^2 's or μ_i 's. It would be interesting to do so in future work, as well as changing players' costs for the general game G^g .

5.3 Discussion

The purpose of G^e was first to give us an insight into the general game G^g . Simple though it was, this equality game provided us very interesting results and allowed us to get intuition for the existence of Nash equilibria in our model. This intuition about the general game was confirmed by our simulations. Contrary to G^e game, G^g simulations showed us only one or two Nash equilibria. In addition, even though the mobile nodes' best response did not depend on the adversary's strategy in G^e , it happened to be highly dependent on the attacker's placement of sniffing stations in the general case G^g . On the other hand, the adversary's best response in G^e was either to do nothing or to deploy sniffing stations wherever there were no mix zone yet. Hence, in the "best" case, the adversary's best strategy was still dependent on the mobile nodes' strategy. In G^g , a third case appeared for the attacker's best response: to always deploy a sniffing station at an intersection, regardless of nodes' strategy, if benefit at this intersection was high enough with respect to sniffing stations' cost.

Chapter 6

Conclusion and Future Work

Preventing third parties to track mobile users' movements is one of the main topics of research in wireless privacy. The wireless nature of mobile communications allow malicious individuals to easily eavesdrop important personal information, such as identity. In order to thwart this threat on location privacy, many privacy-preserving mechanisms have already been developed.

In this thesis, we chose active mix zones to protect mobile users' location privacy. We furthermore assumed that a local passive adversary tries to jeopardize it with sniffing devices. On the one hand, mobile nodes want to maximize their payoff, i.e. their location privacy. On the other hand, the attacker attempts to optimize its own utility, i.e. his tracking power. Assuming both players being rational, we analyzed the interplay between the mobile nodes' strategic decisions and the adversary strategic choice. We developed a payoff model that takes into account most of important game's parameters. We modeled benefit of mobile nodes as the total mixing effectiveness of their deployed mix zones. We included the cost of changing pseudonyms, as well as the cost of being silent within the mix zones. On the other side, benefit of the attacker is proportional to his tracking power but weakened by mixing uncertainty due to mix zones placement. In order to let players decide on their strategies prior to the operation of the mobile network, we presented a simplified flow-based metric.

By studying our model using a game-theoretic approach, we showed, analytically in a specific scenario (homogeneous mobility profile), and with simulations in the general case, that our game always reaches at least one Nash equilibrium. We furthermore studied mobile nodes' and adversary's best responses and proved formally what they are when mobility profile is homogeneous. We showed that, in this particular case, the nodes' best response does not depend on the attacker's strategy. In addition, nodes' payoff is proportional to the number of mix zones they deploy, regardless of where they locate their mix zones. On the contrary, in

the general game, the nodes' best response highly depends on the eavesdropping stations deployment and the mobile users rarely get benefit to locate a mix zone where there already is an adversary's station. We also showed that the adversary's best response depends on the mobile nodes' deployment of mix zones, in both equality and general games. However, under certain conditions on sniffing stations' cost, the attacker gets no incentive to deploy eavesdropping devices, regardless of what the nodes do. Finally, by means of simulations, we verified the accuracy of our analytical results.

In future work, it would be interesting to develop a more complex mobility parameters model. We could for instance relax our assumption of independency between the traffic intensities and sojourn time statistics. In real life, expected sojourn time within crossroads is dependent on the traffic congestion. Another possible improvement could be to extend our model to an incomplete information game. Indeed, the adversary may not have all nodes' payoff information, especially their mobility profile, and vice versa, the nodes may not know all adversary's payoff. Last but not least, mobile nodes can get high incentive to change pseudonyms within the non-monitored area, i.e. where the local adversary has no coverage. This would be less costly for nodes since they would not need to stay quiet during the operation. Based on work done by Buttyán *et al.* [5], we could model uncertainty and mixing efficiency of such passive mix zones.

Bibliography

- [1] http://research.nokia.com/research/rich_context_modeling. [cited at p. 1]
- [2] <http://en.wikipedia.org/wiki/Bluedating>. [cited at p. 1]
- [3] Alastair R. Beresford and Frank Stajano. Location privacy in pervasive computing. In *Pervasive Computing, IEEE*, pages 2(1):46–55, 2003. [cited at p. 2, 3, 9]
- [4] Alastair R. Beresford and Frank Stajano. Mix zones: User privacy in location-aware services. *PerSec*, March 2004. [cited at p. 2, 6]
- [5] Levente Buttyán, Tamás Holczer, and István Vajda. On the effectiveness of changing pseudonyms to provide location privacy in VANETs. In *ESAS*, 2007. [cited at p. 44]
- [6] Levente Buttyán and Jean-Pierre Hubaux. *Security and Cooperation in Wireless Networks*. Cambridge University Press, 2008. [cited at p. 1]
- [7] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. In *Communications of the ACM*, 1981. [cited at p. 6]
- [8] David Chaum and Eugène van Heyst. Group signatures. In *EUROCRYPT*, 1991. [cited at p. 5]
- [9] Thomas M. Cover and Joy A. Thomas. *Elements of information theory*. John Wiley and Sons, Inc., 1991. [cited at p. 14]
- [10] Yves Détraigne and Anne-Marie Escoffier. La vie privée à l'heure des mémoires numériques. Pour une confiance renforcée entre citoyens et société de l'information. <http://www.senat.fr/noticerap/2008/r08-441-notice.html>. [cited at p. 1]

- [11] Yves Eudes. Alex is watching you. http://www.lemonde.fr/technologies/article/2008/11/28/alex-is-watching-you_1124462_651865.html. [cited at p. 1]
- [12] Julien Freudiger, Hossein Manshaei, Jean-Pierre Hubaux, and David Parkes. On non-cooperative location privacy in mobile ad hoc networks. Submitted to CCS, 2009. [cited at p. 17, 20]
- [13] Julien Freudiger, Maxim Raya, and Jean-Pierre Hubaux. Self-organized location privacy in mobile networks. Technical report, 2008. [cited at p. 5]
- [14] Julien Freudiger, Reza Shokri, and Jean-Pierre Hubaux. On the optimal placement of mix zones. In *PETS*, 2009. [cited at p. 3, 6, 9, 10, 12, 13]
- [15] Robert G. Gallager. *Information Theory and Reliable Communication*. Wiley, 1968. [cited at p. 23]
- [16] Denos C. Gazis. *Traffic Theory*. Springer, 2002. [cited at p. 13]
- [17] Robert Gibbons. *A Primer in Game Theory*. FT Prentice Hall, 1992. [cited at p. 18]
- [18] Marco Gruteser and Dirk Grunwald. Anonymous usage of location-based services through spatial and temporal cloaking. In *MobiSys*, pages 31–42, 2003. [cited at p. 5]
- [19] Baik Hoh and Marco Gruteser. Protecting location privacy through path confusion. In *First International Conference on Security and Privacy for Emerging Areas in Communication Networks*, 2005. [cited at p. 7]
- [20] Leping Huang, Kanta Matsuura, Hiroshi Yamane, and Kaoru Sezaki. Enhancing wireless location privacy using silent period. In *Wireless Communications and Networking Conference, 2005 IEEE*, volume 2, pages 1187–1192 Vol. 2, March 2005. [cited at p. 6]
- [21] Leping Huang, Hiroshi Yamane, Kanta Matsuura, and Kaoru Sezaki. Towards modeling wireless location privacy. In *Proceedings of PET*, pages 59–77, 2005. [cited at p. 2]
- [22] Leping Huang, Hiroshi Yamane, Kanta Matsuura, and Kaoru Sezaki. Silent cascade: enhancing location privacy without communication QoS degradation. In *SPC*, 2006. [cited at p. 7]
- [23] Annie Kahn. Avec JCDecaux, l'INRIA imagine la publicité sur mesure pour le chaland. *Le Monde*, 31 mars 2006. [cited at p. 1]

- [24] Leping Huang Mingyan Li, Krishna Sampigethaya and Radha Poovendran. Swing & swap: User-centric approaches towards maximizing location privacy. *WPES*, 2006. [cited at p. 6]
- [25] Roger B. Myerson. *Game Theory*. Harvard University Press, 1991. [cited at p. 17]
- [26] Andreas Pfitzmann and Marit Köhntopp. Anonymity, unobservability, and pseudonymity - a proposal for terminology. In *Designing Privacy Enhancing Technologies: Proceeding of the International Workshop on the Design Issues in Anonymity and Observability, LNCS 2009*, 2000. [cited at p. 6]
- [27] Ronald Rivest, Adi Shamir, and Yael Tauman. How to leak a secret. In *ASIACRYPT*, 2001. [cited at p. 5]
- [28] Elmar Schoch, Frank Kargl, Tim Leimüller, Stefan Schlott, and Panos Papadimitratos. Impact of pseudonym changes on geographic routing in VANETs. In *ESAS*, 2006. [cited at p. 20]
- [29] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In *PET, LNCS 2482*, 2002. [cited at p. 6, 12]
- [30] Claude Shannon. The mathematical theory of communication. In *Bell Systems Technical Journal*, pages 30:50–64, 1948. [cited at p. 12]
- [31] Reza Shokri, Julien Freudiger, Murtuza Jadliwala, and Jean-Pierre Hubaux. A distortion-based metric for location privacy. Submitted to *WPES*, 2009. [cited at p. 6]
- [32] Daniel J. Solove. “I’ve got nothing to hide” and other misunderstandings of privacy, 2007. [cited at p. 1]
- [33] Latanya Sweeney. k-anonymity: a model for protecting privacy. In *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, pages 10:557–570, 2002. [cited at p. 5]

Appendices

Appendix A

Game results

A.1 Equality game

Derivation of equations (5.2) and (5.3):

$$\begin{aligned} u_n &= \sum_{i=1}^4 m_i \left[a_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda \cdot 3\lambda \cdot \log 3 \cdot \sigma^2}{4} + (1 - a_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda(3\lambda\sigma^2 \log 3)}{4} \right] \\ &\quad - \gamma 4\lambda \sum_{i=1}^4 m_i - \frac{4\lambda\mu \sum_{i=1}^4 m_i}{4 \cdot 16\lambda\mu - 4\lambda\mu \sum_{i=1}^4 m_i} \\ &= \sum_{i=1}^4 m_i [a_i(3\lambda^2\sigma^2 \log 3) + (1 + a_i)3\lambda^2\sigma^2 \log 3] \\ &\quad - \gamma 4\lambda \sum_{i=1}^4 m_i - \frac{4\lambda\mu \sum_{i=1}^4 m_i}{64\lambda\mu - 4\lambda\mu \sum_{i=1}^4 m_i} \\ &= \sum_{i=1}^4 m_i \left[3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda - \frac{4\mu\lambda}{64\lambda\mu - 4\mu\lambda \sum_{j=1}^4 m_j} \right] \\ &= \sum_{i=1}^4 m_i \left[3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda - \frac{1}{16 - \sum_{j=1}^4 m_j} \right] \end{aligned} \tag{A.1}$$

$$\begin{aligned}
u_a &= \sum_{i=1}^4 a_i \left[m_i \sum_{j=4(i-1)}^{4i-1} \frac{\lambda(3\lambda\sigma^2 \log 3 - 3\lambda\sigma^2 \log 3)}{4} + (1 - m_i) \sum_{j=4(i-1)}^{4i-1} \frac{\lambda(3\lambda\sigma^2 \log 3)}{4} \right] \\
&\quad - \frac{\beta \sum_{i=1}^4 a_i}{4} \\
&= \sum_{i=1}^4 a_i \left[3\lambda^2\sigma^2 \log 3(1 - m_i) - \frac{\beta}{4} \right]
\end{aligned} \tag{A.2}$$

A.1.1 Nodes' best response

Theorem A.1. *For fixed values of γ and σ^2 , the nodes' best response is a function of λ :*

$$N_n^*(\lambda) = \begin{cases} 0 & \text{if } \lambda < \lambda_1 \\ \left[16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} \right] & \text{if } \lambda_1 \leq \lambda \leq \lambda_2 \\ 4 & \text{if } \lambda > \lambda_2 \end{cases} \tag{A.3}$$

$$\text{where } \lambda_1 = \frac{2\gamma + \sqrt{4\gamma^2 + 3\sigma^2 \log 3 \left(\frac{8}{31}\right)^2}}{3\sigma^2 \log 3} \text{ and } \lambda_2 = \frac{2\gamma + \sqrt{4\gamma^2 + 3\sigma^2 \log 3 \left(\frac{8}{25}\right)^2}}{3\sigma^2 \log 3}$$

Proof. From theorem 5.1, we know that maximal payoff is reached for

$$x = 16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} \tag{A.4}$$

Then, in order to get $N_n^* = 4$, we must have $x > 3.5$, i.e.,

$$16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} > 3.5 \tag{A.5}$$

After some calculation steps, we reach the following quadratic inequality:

$$3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda - \left(\frac{8}{25}\right)^2 > 0 \tag{A.6}$$

Solving (A.6) at equality, we get two solutions:

$$\lambda_{1,2} = \frac{2\gamma \pm \sqrt{4\gamma^2 + 3\sigma^2 \log 3 \left(\frac{8}{25}\right)^2}}{3\sigma^2 \log 3} \tag{A.7}$$

Since λ is positive, the only acceptable solution is $\lambda_1 = \frac{2\gamma + \sqrt{4\gamma^2 + 3\sigma^2 \log 3 \left(\frac{8}{25}\right)^2}}{3\sigma^2 \log 3}$. As the quadratic function in (A.6) is convex, x is greater than 3.5 for

$$\lambda > \frac{2\gamma + \sqrt{4\gamma^2 + 3\sigma^2 \log 3 \left(\frac{8}{25}\right)^2}}{3\sigma^2 \log 3} \tag{A.8}$$

In the same way, we must have verify that $x < 0.5$ to get $N_n^* = 0$. This is reached for

$$\lambda < \frac{2\gamma + \sqrt{4\gamma^2 + 3\sigma^2 \log 3 \left(\frac{8}{31}\right)^2}}{3\sigma^2 \log 3} \quad (\text{A.9})$$

□

Theorem A.2. For fixed values of λ and σ^2 , the nodes' best response is a function of γ :

$$N_n^*(\gamma) = \begin{cases} 0 & \text{if } \gamma > \gamma_1 \\ \left[16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} \right] & \text{if } \gamma_1 \leq \gamma \leq \gamma_2 \\ 4 & \text{if } \gamma < \gamma_2 \end{cases} \quad (\text{A.10})$$

where $\gamma_1 = \frac{3\lambda^2\sigma^2 \log 3 - \left(\frac{8}{31}\right)^2}{4\lambda}$ and $\gamma_2 = \frac{3\lambda^2\sigma^2 \log 3 - \left(\frac{8}{25}\right)^2}{4\lambda}$

Proof. In the same way as proof of theorem A.1, we get the inequality (A.6). As it is linear with respect to γ , we just need to isolate γ . Hence, in order to have $N_n^* = 4$, γ must verify

$$\gamma < \frac{3\lambda^2\sigma^2 \log 3 - \left(\frac{8}{25}\right)^2}{4\lambda} \quad (\text{A.11})$$

In the same way, in order to have N_n^* equal to zero, γ must satisfy

$$\gamma > \frac{3\lambda^2\sigma^2 \log 3 - \left(\frac{8}{31}\right)^2}{4\lambda} \quad (\text{A.12})$$

□

Theorem A.3. For fixed values of λ and γ , the nodes' best response is a function of σ^2 :

$$N_n^*(\sigma^2) = \begin{cases} 0 & \text{if } \sigma^2 < \sigma_1^2 \\ \left[16 - \frac{4}{\sqrt{3\lambda^2\sigma^2 \log 3 - 4\gamma\lambda}} \right] & \text{if } \sigma_1^2 \leq \sigma^2 \leq \sigma_2^2 \\ 4 & \text{if } \sigma^2 > \sigma_2^2 \end{cases} \quad (\text{A.13})$$

where $\sigma_1^2 = \frac{\left(\frac{8}{31}\right)^2 + 4\gamma\lambda}{3\lambda^2 \log 3}$ and $\sigma_2^2 = \frac{\left(\frac{8}{25}\right)^2 + 4\gamma\lambda}{3\lambda^2 \log 3}$

Proof. The demonstration is very similar to the proof of theorem A.2. We can also reuse inequality (A.6) of theorem A.1 and get a linear inequality with respect to σ^2 . Therefore, $N_n^* = 4$ if the following condition is verified

$$\sigma^2 > \frac{\left(\frac{8}{25}\right)^2 + 4\gamma\lambda}{3\lambda^2 \log 3} \quad (\text{A.14})$$

In the same way, $N_n^* = 0$ whenever

$$\sigma^2 < \frac{\left(\frac{8}{31}\right)^2 + 4\gamma\lambda}{3\lambda^2 \log 3} \quad (\text{A.15})$$

□