Analysis of Strategic Security Through Game Theory for Mobile Social Networks

Anna V. Guglielmi and Leonardo Badia
Dept. of Information Engineering, University of Padova, via Gradenigo 6B, 35131 Padova, Italy
email: {guglielm, badia}@dei.unipd.it

***Abstract*—in mobile social networks, legitimate transmitting nodes can be contrasted by malicious attackers acting on the purpose of disrupting communication. Our aim is to use game theory to identify malicious nodes. With respect to previous similar formulations, we consider a wider array of action options for the players, notably we include a choice about whether to engage or not in packet exchanging, and also malicious activity and its prevention. This leads to a structured analysis of the resulting game, resulting in different equilibria. We use a Bayesian Game where the critical parameter is the likelihood that the unknown agent is malicious. Investigating on the Nash equilibria in pure/mixed strategies found, we can see how in some cases the malicious behavior of the nodes can be tolerated since a trade-off between their presence and the effect of damages that they caused can be reached.**

***Index Terms*—Mobile social networks; Game theory; Network security; Bayesian games; Repeated games.**

I. Introduction

SOCIAL mobile networking is rapidly expanding, as most of the users connected to social networks also enjoy wireless connectivity and the seamless integration brought by next generation networking [1]. However, this poses unprecedented challenges in terms of network security. Connecting to unprotected wireless hot-spots and/or interacting with other mobile nodes in a machine-to-machine communication may imply the risk of having sensible information stolen and/or services disrupted by malicious adversaries [2], [3].

Mobile nodes are usually unaware of the intentions of the terminals they are interacting with. From a theoretical perspective, the network layer of a social mobile network operates assuming full cooperation and commonality of intents. However, not all the involved nodes may be benign: some can be adversaries pretending to collaborate, while instead pursuing some different purposes than the network wellbeing. In general, there exist multiple kinds of malicious clients with different purposes and abilities. These clients could act as hacker, cracker, cybercriminals, cyber terrorist, and malicious insiders [4]. As a consequence, we deal with several security problems in which the identity of a client is unknown to a legitimate node searching for collaboration. Thereafter, we refer to this scenario as a server-client interaction, even though this is done only to follow the common choices in the literature and by no means limited to this kind of interaction. Hence, in what follows "server" denotes a legitimate node of the network that interacts with a "client" of unknown intentions [5].

If this server trusts the clients to be benign and does not implement a surveillance procedure, attackers would be able to disrupt network operation by simply pretending to be legitimate clients. On the other hand, it the server considers each client as a potential malicious agent, it will implement some costly surveillance procedure that decreases the quality of service provided, not only for the server itself, but also for all the legitimate benign clients. Therefore, the most compelling network security problem is to correctly define a proper operation where both types of clients are considered and efficient defense strategies are designed with the purpose of preventing malicious activities and providing good quality of services to benign nodes [6].

Game theory has been widely used to tackle security problems in communication networks [7]. This approach enables low-cost distributed management of the network, as well as defining some theoretical performance bounds. However, most security issues are generally modeled as the interaction between a defender and a node that is certain to be a malicious attacker. In reality, the correct identification of a malicious client is challenging, especially if the attacker adopts a strategic behavior that tries to keep its true intent hidden from the network [9]. For this reason, it is not always possible to infer the identity of a malicious agent known to the other nodes of the network, but rather, only an *estimate* about the character of the client can be made; thus, the specific instrument to use in this scenario is that of Bayesian games [8].

Our purpose is to analyze a server/client game where the client can be either benign or malicious with a known prior probability. The nodes have an available set of possible actions that they can perform in the network. We remark that the existing literature usually considers just two actions available to each player, which enables a simple approach to the problem but on the other hand makes it easy to identify the client type (depending on whether he plays its collaborative action or not). We aim at expanding this scenario by including additional moves that enable a malicious client to get undetected, which in our opinion is a more realistic option.

A sample scenario could be a packet forwarding situation that is often used in the context of security. In this case, sensor nodes would have the option to transmit a packet, and supporting clients acting as relays can choose among forwarding it without damaging it, ignoring it, or adding some intentional errors to its transmission. This wider set of available options makes the analysis more realistic. Still, we are able to analyze the game in closed-form, and show some examples of analysis with given numerical values. Specifically, we consider a Bayesian game framework to capture the

uncertainty about the client's nature, and identify the Bayesian Nash equilibria (BNEs) of this game, and discuss how they are affected by the numerical parameters. We also suggest some implications on the security strategies of the network.

The rest of this paper is organized as follows. In Section II we discuss some related works. In Section III, we describe the system model and the game theory application; we compute the BNEs in Section IV. Section V presents some numerical results. Finally, Section VI draws the conclusions.

II. Related works

In [5] and [8], authors survey and classify existing game theoretic approaches to network security issues. They state that game theory can address network security and identify some promising research directions. In [9], a new Bayesian hybrid detection approach is suggested for a network defender. A lightweight monitoring system is used to estimate the opponent's actions, and a heavyweight monitoring system acts as a last resort of defense. The results show that the dynamic game produces energy-efficient monitoring strategies for the defender, while improving the overall hybrid detection power.

Similar studies are [10] and [11]. In particular, a new exact method, called *DOBSS*, for finding the optimal strategy for the leader in a Bayesian Stackelberg game is presented in [10]. In these kinds of games, one agent, called the leader, must commit to a possibly mixed strategy that can be observed by other agents, called the followers, before they choose their own strategies. The leader is uncertain about the types of adversary it may face, therefore such games are extremely valuable in modeling domains involving security, including patrolling, setting up checkpoints, network routing, transportation systems and others. Solution techniques such as *DOBSS* are also relevant for efficiently solving such games. In [11], using a game-theoretic approach, authors propose a selective and dynamic mechanism for counter-fingerprinting. They first model and analyze the interaction between a fingerprinter and a target as a signaling game. Then, they derive the Nash equilibrium strategy profiles based on the information gain analysis and they design a mechanism to prevent or to significantly slow down fingerprinting attacks. Their game-theoretic approach appropriately distinguishes a fingerprinter from a benign client and mystifies packets to confuse the fingerprinter, while minimizing the effects on legitimate clients. This mechanism can reduce the probability of success of the fingerprinter significantly, without deteriorating the overall performance of other clients.

Paper [12] presents an application of the Bayesian game theory to model node behaviors in trajectory privacy preservation activities in mobile wireless sensor networks. The characteristics of autonomous nodes, including selfish, malicious, and cooperative, are formulated in a game, and the trustworthiness of the unknown type node has been evaluated. Then, authors derive the equilibrium strategies of the game in both theoretical and simulation results.

In [13], the authors study two-player security games which can be viewed as sequences of non-zero-sum matrix games played by an attacker and a defender. It is assumed that, at each stage of the game iterations, the players make imperfect observations of each other's previous actions. The underlying decision process can be seen as a fictitious play (FP) game, but in their analysis the communication channels that carry action information from one player to the other, or the sensor systems, are error prone. Two possible scenarios are addressed: if the error probabilities associated with the sensor systems are known to the players, then the analysis provides guidelines for each player to reach a Nash equilibrium related to the underlying static game; if the error probabilities are not known to the players, then they evaluate the effect of observation errors on the convergence to the Nash equilibrium and the final outcome of the game. Moreover, both the classical FP and the stochastic FP are discussed, where for the latter the payoff function of each player includes an entropy term randomizing its own strategy, which can be interpreted as a way of concealing its true strategy.

Finally, in many formalizations, service providers are assumed to be unaware of the type of their clients that can either be malicious (and attack at any time during their connections) or legitimate agents. In [7], a general framework is provided for modeling security problems subject to different types of clients connected to service providers. The authors develop an incomplete-information two-player game to capture the interaction between the service provider and an unknown client. They consider two types of clients, i.e. attacker and benign clients, and they analyze the game using perfect BNEs with different conditions. As a result, they design an algorithm using the computed BNE strategy profiles to find the best defense strategy that the server should use.

III. System model

We consider a Bayesian Game, in which players can be of different *types*, and each type implies a different utility function. This is a way to capture different behaviors that the players may have. Each player is aware of its own type only. About the other players' types, each player only knows the prior probability distribution over types, which is common knowledge among the players. Possible assumptions about an opponent types are integrated in the beliefs [14]. A *strategy* of a node is identified as a complete plan of actions that covers every contingency of the game, also including types.

For the sake of presentation, we consider the interaction between two nodes. This can be further extended to a larger number of nodes by repeating the pairwise interaction. In a mobile social network, each node can be either malicious or benign, depending on how they interact with the rest of the network. Malicious nodes damage communications in the network so as to disrupt its operation; we only focus on this kind of malicious attacks, that are aimed at denying service [6]. Also, we consider the typical interaction of multi-hop networks, where nodes exchange packets and require mutual collaboration in the form of relaying data to the end destination. In this setting, a server can decide to transmit a packet to another node, monitor the network in order to detect malicious nodes without transmitting any packet, or do nothing and remain inactive. Furthermore, considering that some nodes

		Player 2		
		Forward	**Ignore**	**Damage**
Player 1	**Nothing**	$0, -1$	$0, 0$	$-1, -\infty$
	Packet	$1, 1$	$-1, 0$	$-2, -\infty$
	Surveillance	$-3, -2$	$-3, 0$	$2, -\infty$

TABLE I
NORMAL-FORM (MATRIX) OF THE GAME WITH A BENIGN PLAYER 2

		Player 2		
		Forward	**Ignore**	**Damage**
Player 1	**Nothing**	$0, -1$	$0, 0$	$-1, 1$
	Packet	$1, -1$	$-1, 0$	$-2, 3$
	Surveillance	$-3, -2$	$-3, 0$	$2, -3$

TABLE II
NORMAL-FORM (MATRIX) OF THE GAME WITH MALICIOUS PLAYER 2

may have malicious purposes, when a node is asked to forward a packet, it can really do so, or ignore the packet, or even forward a corrupted version of it, to cause harm.

We assume that one of the nodes, denoted as node 1, is a benign node (i both nodes are malicious, it is pointless to consider the defense from an adversary). The other node instead, denoted as player 2, may be either a legitimate node or a malicious attacker. We define p as the prior probability of being malicious for node 2. Note that the type of player 2 is private information, which means that 2 is aware of being either malicious or not, while 1 does not know (but it can estimate it via the probability p). Node 1 acts as a *server* that can transmit packets to a client or protect the network against malicious agents through surveillance. Node 2 acts as a *client* and it can forward a received packet without any damage, ignore a received packet, or forward a received packet corrupting it depending on its behavior. We can model this scenario using a two-player static Bayesian game in which the server and the client are the *players*. The set of $actions$ that can be taken is $\{Nothing,\ Packet,\ Surveillance\}$ for player 1 and $\{Forward,\ Ignore,\ Damage\}$ for player 2. These actions reflect the three types of possible interactions. Every pair of actions yields an arbitrary quantification of the goodness coming from its resulting outcome. Because of the Bayesian context, it is assumed that each player has a type θ_i, with values in $\{0, 1\}$, that describes its being malicious. If a generic player i is malicious, its type is $\theta_i = 1$, otherwise it is $\theta_i = 0$. In this paper, we assume that only player 2 has a type. As a consequence, $\theta_1 = 0$ and $\theta_2 \in \{0, 1\}$; finally, we formally state that $p = \text{Prob}[\theta_2 = 1]$.

Before analyzing the Bayesian game, we separately describe separately the cases in which player 2 is benign or malicious, as two different static games without any Bayesian element. We set arbitrary numbers for the utility of each outcome, that satisfy an ordinal criterion (most preferred situations have a higher utility). The actual numbers are chosen just for the sake of easy computations; the same analysis is still valid for other choices of values.

First, consider the scenario in which player 2 is not malicious. Table I shows the normal-form matrix describing this game. Player 2 will reach a payoff equal to $-\infty$ if it chooses to damage the transmission. Therefore, it will never choose that action that can be considered as a strictly *dominated* strategy. For this reason, we can neglect this strategy in our analysis. Once we neglect the $Damage$ action of player 2, a similar observation iteratively holds for the $Surveillance$ strategy of player 1. In particular, this strategy is now strictly dominated by $Nothing$ that allows to the player to earn a higher payoff regardless of what opponent may do. This is quite intuitive because, since player 2 is not malicious, it does not have any incentive to damage network, therefore, it is not necessary for player 1 to monitor and control the network against malicious agents. After these considerations, the 3×3 matrix can be simplified in a 2×2 matrix keeping the rows denoted by $Nothing$ and $Packet$, and the columns $Forward$ and $Ignore$ from the original normal-form matrix. At this point, we can easily observe that two Nash Equilibria (NEs) in pure strategies can be found. In more details, these NEs are: ($Nothing$, $Ignore$) and ($Packet$, $Forward$). In addition, we investigate on the existence of NEs in mixed strategies. Defining α as the probability that player 1 chooses $Nothing$ and β the probability that player 2 plays $Forward$, this game has a single mixed NE $(\frac{1}{2}, \frac{1}{2})$. This means that player 1 will plays $Nothing$ with probability $\frac{1}{2}$ and $Packet$ with probability $\frac{1}{2}$. In more details, we find that player 1 will play $Nothing$ if $\beta < \frac{1}{2}$, it will play $Packet$ if $\beta > \frac{1}{2}$, finally it will be indifferent between playing $Nothing$ of $Packet$ if $\beta = \frac{1}{2}$. Similar observations hold for player 2; it will choose $Forward$ if $\alpha > \frac{1}{2}$, it will choose $Ignore$ if $\alpha < \frac{1}{2}$, and it will be indifferent if $\alpha = \frac{1}{2}$. Eqs. (1) and (2) describe the best response strategy of player 1 and player 2, respectively.

$$BR_1 = \begin{cases} Nothing, & \text{if } \beta \leq \frac{1}{2} \\ Packet, & \text{if } \beta \geq \frac{1}{2} \end{cases}, \tag{1}$$

$$BR_2 = \begin{cases} Forward, & \text{if } \alpha \geq \frac{1}{2} \\ Ignore, & \text{if } \alpha \leq \frac{1}{2} \end{cases}. \tag{2}$$

Table II shows the normal-form matrix describing the simple game in which player 2 acts as a malicious player. With respect to Table I, few meaningful changes have been made. In particular, acting to $Damage$ the network is now possible. Also, the mutual benefit of cooperation that was previously described by the outcome for ($Packet$, $Forward$) has been altered for the malicious player. Focusing on player 2's payoffs, $Forward$ is strictly dominated by $Ignore$. Then, for player 1 it holds that $Packet$ is strictly dominated by $Nothing$. This is intuitive because, since player 2 is malicious, it does has incentive to damage nodes communication in order to gain a higher payoff; therefore, it is better for player 1 to monitor and control the network against malicious agents or to do nothing instead of transmitting a packet which will be corrupted. After these considerations, also in this case the 3×3 matrix can be simplified in a 2×2 matrix keeping the rows denoted by $Nothing$ and $Surveillance$, and the columns $Ignore$ and $Damage$ from the original normal-form matrix. At this point, we can observe that there are no BNE in pure strategies. Concerning the existence of BNEs in mixed strategies and defining α as the probability that player 1 chooses $Nothing$

		Player 2			
		IF	**II**	**DF**	**DI**
Player 1	**N**	0, p-1	0, 0	-p, 2p-1	-p,p
	P	1-2p, 1-p	-1, 0	1-3p, 2p+1	-p-1, 3p
	S	-3, -2+2p	-3, 0	5p-3, -p-2	5p-3, -3p

TABLE III
NORMAL-FORM (MATRIX) OF THE BAYESIAN GAME

and β the probability that player 2 plays $Ignore$, this game has a single mixed BNE denoted as $(\frac{3}{4}, \frac{1}{2})$, i.e., player 1 will plays $Nothing$ with probability $\frac{3}{4}$ and $Surveillance$ with probability $\frac{1}{4}$ and player 2 will choose $Ignore$ or $Damage$ with equal probability. The best response of players 1 and 2 is shown in (3) and (4), respectively.

$$BR_1 = \begin{cases} Nothing, & \text{if } \beta \geq \frac{1}{2} \\ Surveillance, & \text{if } \beta \leq \frac{1}{2} \end{cases}, \tag{3}$$

$$BR_2 = \begin{cases} Ignore, & \text{if } \alpha \leq \frac{3}{4} \\ Damage, & \text{if } \alpha \geq \frac{3}{4} \end{cases}. \tag{4}$$

IV. BNE ANALYSIS

Now we consider an incomplete information game in which player 1 is not aware of the type of its opponent, i.e., benign or malicious. Either of the two games previously described is actually played with probability $1 - p$ or p respectively, with the parameter p (but not the actual type) being known in advance by player 1. That is, p is the prior probability of player 2 being of malicious type. Table III shows the payoff of each player according to the different strategies that it can take. For the sake of exposition, we shorten the action names by writing N, P, S, F, I, D, instead of "$Nothing$," "$Packet$," "$Surveillance$," "$Forward$," "$Ignore$," and "$Damage$," respectively. As previously stated for the two simple games, when player 2 is not malicious it actually plays only $Forward$ or $Ignore$, instead when it is malicious it can play $Ignore$ or $Damage$. Furthermore, under a Bayesian framework we represent strategies by defining an action for each player's type. Thus, while player 1's possible strategies are the same as the previous two games (i.e., they coincide with actions N, P, S), the strategies of player 2 are instead *pairs* of actions, since we need to specify what player 2 does depending on its type. This means that player 2 has four possible strategies in the Bayesian game, namely its strategy set is $\{IF,\ II,\ DF,\ DI\}$. We denote each strategy as a pair of actions to be played when the client is either malicious or benign, respectively. For example, strategy IF means that player 2 will play $Ignore$ if malicious, $Forward$ otherwise. Note that, for the sake of brevity, we already discarded player 2's strictly dominated strategies, depending on its type. To compute the payoffs earned by the players, we take expectations over beliefs for types. For example, for actions (N,IF): player 1 earns 0 if player 2 is not malicious, that is, with probability p, and 0 if player 2 is malicious, therefore the payoff is still 0; player 2 earns -1 if it is not malicious, that is, with probability p, and 0 otherwise, therefore its payoff is $0 \cdot p - 1 \cdot (1 - p) = p - 1$. To compute the BNEs, we consider different cases depending on whether $p < \frac{1}{2}$ or $p > \frac{1}{2}$.

		Player 2	
		DF	**DI**
Player 1	**N**	-p, 2p-1	-p, p
	P	1-3p, 2p+1	-p-1, 3p

TABLE IV
SIMPLIFIED NORMAL-FORM (MATRIX) OF THE BAYESIAN GAME

A. Likely benign client ($p < \frac{1}{2}$)

From Table III, we notice that S is a strategy strictly dominated by N for player 1. As a consequence, we can neglect S as a possible action of player 1. Moreover, we can also state that II and IF are strictly dominated strategies for player 2; the former is strictly dominated by DI, the latter is strictly dominated by DF. Table IV summarized the game considering these observations. There exist two BNEs in pure strategies: $(N,\ DI)$ and $(P,\ DF)$. Defining α as the probability that player 1 plays N and β as the probability that player 2 chooses DF, we also find a mixed BNE: $\left(\frac{1}{2}, \frac{1}{2(1-p)}\right)$. In more detail, the best strategy for player 1 is to choose N if $\beta < \frac{1}{2(1-p)}$, P if $\beta > \frac{1}{2(1-p)}$, and to be indifferent if $\beta = \frac{1}{2(1-p)}$. At the same time, the best strategy for player 2 is to play DF if $\alpha < \frac{1}{2}$, DI if $\alpha > \frac{1}{2}$, and to be indifferent if $\alpha = \frac{1}{2}$. The best response strategies of the two players are:

$$BR_1(\theta_2) = \begin{cases} N, & \text{if } \beta \leq \frac{1}{2(1-p)} \\ P, & \text{if } \beta \geq \frac{1}{2(1-p)} \end{cases}, \tag{5}$$

$$BR_2(\theta_1) = \begin{cases} DF, & \text{if } \alpha \leq \frac{1}{2} \\ DI, & \text{if } \alpha \geq \frac{1}{2} \end{cases}. \tag{6}$$

B. Likely malicious client ($p > \frac{1}{2}$)

In Table III, P is now a strategy dominated by N for player 1. As a consequence, we can neglect P as a possible action of player 1. We can also state that, focusing on player 2 payoff, IF is dominated by II. Moreover, it can be observed that there exists a linear combination of the payoffs that player 2 can reach playing DI and II that gives a higher payoff with respect to playing DF. In other terms, there exists μ such as $\mu \cdot u_2(DI) + (1 - \mu) \cdot u_2(II) > u_2(DF)$, where u_i is the payoff of player i. Eq. (7) below shows that, since $p > \frac{1}{2}$, $\frac{2p-1}{p} < \frac{p+2}{3p}$, there exists an interval for μ in which the condition on the payoffs is satisfied:

$$\frac{2p-1}{p} < \mu < \frac{p+2}{3p}. \tag{7}$$

Thus, we can neglect DF as a dominated strategy for player 2. Table V summarized the game considering these observations. In this case, the game has no BNEs in pure strategies. However, defining α as the probability that player 1 plays N and β as the probability that player 2 chooses II, there exists a mixed BNE: $\left(\frac{3}{4}, \frac{2p-1}{2p}\right)$. In more details, the best strategy for player 1 is to choose N if $\beta > \frac{2p-1}{2p}$, S if $\beta < \frac{2p-1}{2p}$, and to be indifferent if $\beta = \frac{2p-1}{2p}$. At the same time, the best

		Player 2	
		IF	**DI**
Player 1	**N**	0, 0	-p, p
	S	-3, 0	5p-3, -3p

TABLE V
SIMPLIFIED NORMAL-FORM (MATRIX) OF THE BAYESIAN GAME

strategy for player 2 is to play II if $\alpha < \frac{3}{4}$, DI if $\alpha > \frac{3}{4}$, and to be indifferent if $\alpha = \frac{3}{4}$. The best responses are:

$$BR_1(\theta_2) = \begin{cases} N, & \text{if } \beta \geq \frac{2p-1}{2p} \\ S, & \text{if } \beta \leq \frac{2p-1}{2p} \end{cases}, \tag{8}$$

$$BR_2(\theta_1) = \begin{cases} II, & \text{if } \alpha \leq \frac{3}{4} \\ DI, & \text{if } \alpha \geq \frac{3}{4} \end{cases}. \tag{9}$$

C. Uniform prior probability ($p = \frac{1}{2}$)

The final case is when the server considers both types of the client to be equally likely, which may also mean that the server cannot make any assumption about the client's character. If we consider this situation of maximum uncertainty about player 2, we obtain the payoffs shown in Table VI. As it can be noted from the table, in this case there are no dominated strategies. Therefore, the analysis for the computation of the equilibria is much harder with respect to previous cases. However, a mixed equilibrium is bound to exist, and also continuity of the expected utility in the Bayesian case can be exploited (so that we can find the Nash equilibrium via the left and right limits that fall within the two previous cases).

V. NUMERICAL RESULTS

In this section, we discuss some numerical results, obtained by both analysis and simulation. We consider that each player plays a mixed equilibrium strategy. We limit our results to the repetition of a static Bayesian game, meaning that at each stage we do not consider the update of the belief that player 1 has on the type of its opponent. However, we can model this scenario through the repetition of a dynamic Bayesian game in which the game evolution is taken into account and the defender can dynamically update its beliefs based on new observations of actions chosen by its opponent and the game history in order to adjust its monitoring strategy accordingly. We will consider this aspect as a future work.

Figs. 1 and 2 show the payoff reached by player 1 and player 2, respectively, versus the value of p, i.e., the prior probability that player 2 is malicious. In these figures, we compare the payoff obtained through the theoretical analysis

		Player 2			
		IF	**II**	**DF**	**DI**
Player 1	**N**	0, $-1/2$	0, 0	$-1/2$, 0	$-1/2$, $1/2$
	P	0, $1/2$	-1, 0	-1/2, 2	-3/2, 3/2
	S	-3, -1	-3, 0	-1/2, -5/2	-1/2, -3/2

TABLE VI
NORMAL-FORM (MATRIX) OF THE BAYESIAN GAME FOR $p = 1/2$

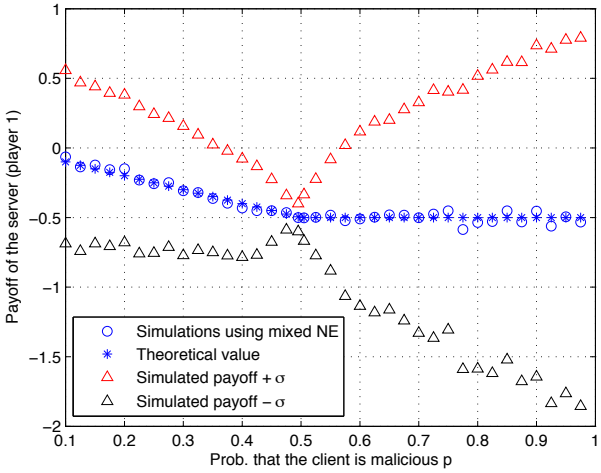

Fig. 1. Comparison of player 1 payoff obtained via the theoretical analysis and simulations considering mixed BNE.

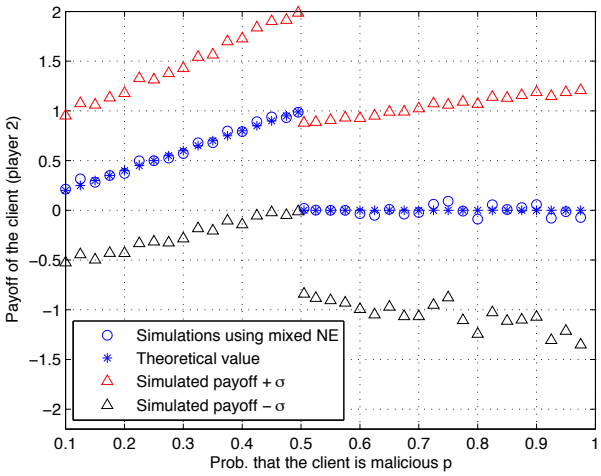

Fig. 2. Comparison of player 2 payoff obtained via the theoretical analysis and simulations considering mixed BNE.

discussed in previous sections represented by the blue-dotted line with respect to the payoff obtained via simulations described by the blue-star line. As it can be observed by the figures, the simulations results follow very well the theoretical development. In particular, looking at Fig. 1 we can see that, for $p < \frac{1}{2}$, the lower the value of p, the higher the value of player 1 payoff. This is reasonable because the lower p, the less likely that player 2 is malicious, then it is not necessary to apply surveillance against a possible attacker in the network. As a consequence, player 1 can reach a higher payoff by transmitting packets. On the other hand, for $p > \frac{1}{2}$ the value of player 1's payoff is constantly equal to -0.5 (also confirmed by simulations). This is because players play the mixed BNE, then for player 1 is more likely to play N and, as a consequence, it is more likely that player 2 will play DI. Focusing on Fig. 2, we can notice that, for $p < 0.5$, the lower p, the lower player 2's payoff. Indeed, the higher p, the lower the value of $(2(1-p))^{-1}$; therefore, player 2 will reach a higher payoff damaging the network. For $p > \frac{1}{2}$, the

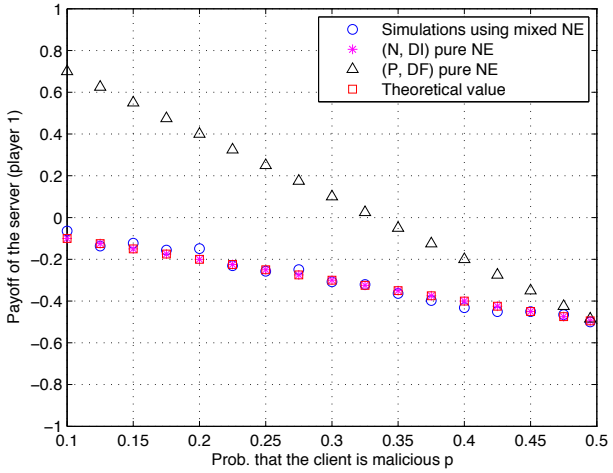

Fig. 3. Comparison of player 1 payoff obtained via the theoretical analysis, simulations considering mixed BNE, and considering pure BNEs for $p < \frac{1}{2}$.

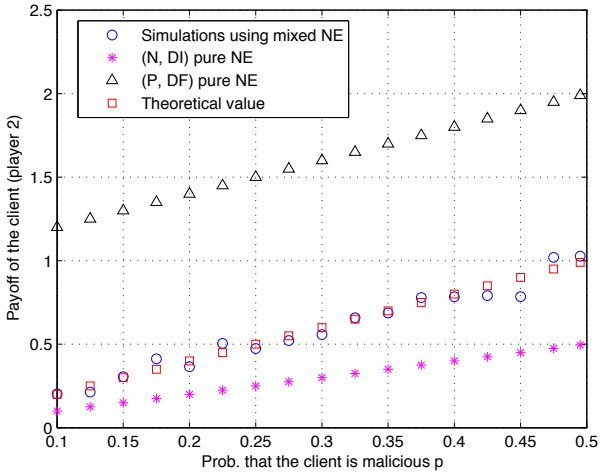

Fig. 4. Comparison of player 2 payoff obtained via the theoretical analysis, simulations considering mixed BNE, and considering pure BNEs for $p < \frac{1}{2}$.

higher is p, the lower is the probability that player 2 will play strategy DI. Since player 1 is more likely to play N at the equilibrium, player 2's payoff will be 0.

In Figs. 1 and 2 we also show the simulated payoff value plus or minus its standard deviation (obtained by simulation) with the red/black-triangle lines. Looking at Fig. 1 we can observe that for too low and high values of p there is a greater dispersion of the individual observations around the average value. This aspect is less pronounced in Fig. 2.

Finally, Fig. 3 and Fig. 4 show the development of the players payoff considering theoretical analysis results with the red-square line, simulations results described by the blue-dotted line, and the results obtained considering that players play the two pure BNEs shown by magenta-star line and black-triangle line. In these figures, we consider p varying in the interval $[0.1\ 0.5]$; indeed, only in this interval there exists the pure BNEs. As we can notice, one of the two pure BNEs, that is $(P,\ DF)$, outperforms the other cases. In particular, this is more evident for player 2's payoff shown in Fig. 4.

VI. Conclusions

We employed game theory to characterize interactions in social online networking, and we considered a Bayesian game where a server is operating with incomplete information, i.e., only a prior estimate, on the client actual type (malicious/benign). The analysis confirms practical aspects of network surveillance. In particular, malicious clients are hard to defeat without a proper surveillance mechanism that may be very costly. The best strategy for the server would not be to always identify malicious clients, but rather to force them to strategically play some less harmful strategies. This would lead to a more effective implicit surveillance.

Our numerical findings confirm that it may be more advantageous even for a malicious client not to harm the network in fear of retaliation. As a development of our analysis, we can consider a strategic client that does not apply just a myopic optimization of its own payoff, but rather tries to avoid being identified. The best way to do so would actually be to occasionally cooperate with the network, which would lead to a transparent surveillance in which the client itself has the right incentive to behave correctly.

References

[1] N. Vastardis and K. Yang, "Mobile social networks: Architectures, social properties, and key research challenges," *IEEE Commun. Surv. & Tut.*, vol. 15, no. 3, pp. 1355-1371, 2013.

[2] G. Quer, F. Librino, L. Canzian, L. Badia, and M. Zorzi, "Inter-network cooperation exploiting game theory and Bayesian networks," *IEEE Trans. Commun.*, vol. 61, no. 10, pp. 4310-4321, 2013.

[3] V. Vadori, M. Scalabrin, A. V. Guglielmi, and L. Badia, "Jamming in underwater sensor networks as a Bayesian zero-sum game with position uncertainty," *Proc. IEEE Globecom*, 2015.

[4] R. Sabillon, J. Cano, V. Cavaller, and J. Serra, "Cybercrime and cybercriminals: A comprehensive study," *Int. J. Comp. Netw. Commun. Sec.*, vol. 4, no. 6, 165-176, 2016.

[5] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Basar, J. P. Hubaux, "Game theory meets network security and privacy," *ACM Comput. Surv.*, vol. 45, n. 3, pp. 1-39, 2013.

[6] M. Jain, B. An, M. Tambe, "Security games applied to real-world: Research contributions and challenges," *Moving Target Defense II*, pp. 15-39, Springer, 2013.

[7] S. Farhang, M. H. Manshaei, M. N. Esfahani, Q. Zhu, "A dynamic bayesian security game framework for strategic defense mechanism design," *R. Poovendran and W. Saad (eds.) Proc. GameSec, LNCS*, vol. 8840, pp. 317-326, Springer, 2014.

[8] X. Liang, Y. Xiao, "Game theory for network security," *IEEE Comm. Surv.& Tut.*, vol. 15, n. 1, pp. 472-486, 2013.

[9] Y. Liu, C. Comaniciu, H. Man, "A Bayesian game approach for intrusion detection in wireless ad hoc networks," *Proc. Workshop on Game Theory for Communications and Networks*, ACM, 2006.

[10] P. Paruchuri, J.P. Pearce, J. Marecki, M. Tambe, F. Ordonez, and S. Kraus, "Playing games for security: an efficient exact algorithm for solving Bayesian Stackelberg games," *Proceedings of AAMAS*, pp. 895-902, 2008.

[11] M. A. Rahman, M. H. Manshaei, and E. Al-Shaer, "A game-theoretic approach for deceiving remote operating system fingerprinting," *IEEE CNS*, pp. 73-81, 2013.

[12] X. Jin, N. Pissinou, S. Pumpichet, C. A. Kamhoua, and K. Kwiat, "Modeling cooperative, selfish and malicious behaviors for trajectory privacy preservation using Bayesian game theory," *Local Computer Networks (LCN)*, pp. 835-842, IEEE, 2013.

[13] K.C. Nguyen, T. Alpcan, T. Basar, "Security games with incomplete information," *Proc. IEEE ICC*, 2009.

[14] M. J. Osborne and A. Rubinstein, *A course in game theory*. Cambridge, USA: The MIT Press, 1994.

[15] S. Tadelis, *Game theory: an introduction*. Princeton Univ. Press, 2013.