Trust Exploitation and Attention Competition: A Game Theoretical Model

Hao Fu, Hongxing Li, Zizhan Zheng, Pengfei Hu, Prasant Mohapatra

Department of Computer Science, University of California, Davis, CA, USA.

{haofu, honli, cszheng, pfhu, pmohapatra}@ucdavis.edu

Abstract—The proliferation of Social Network Sites (SNSs) has greatly reformed the way of information dissemination, but also provided a new venue for hosts with impure motivations to disseminate malicious information. *Social trust* is the basis for information dissemination in SNSs. Malicious nodes judiciously and dynamically make the balance between maintaining its social trust and selfishly maximizing its malicious gain over a long time-span. Studying the optimal response strategies for each malicious node could assist to design the best system maneuver so as to achieve the targeted level of overall malicious activities. In this paper, we propose an interaction-based social trust model, and formulate the maximization of long-term malicious gains of multiple competing nodes as a non-cooperative differential game. Through rigorous analysis, optimal response strategies are identified and the best system maneuver mechanism is presented. Extensive numerical studies further verify the analytical results.

I. Introduction

We have witnessed the prevailing useage of Social Network Sites (SNSs), including Facebook, Twitter and Google+, for the information sharing among users on their personal pages and the interaction with friends or followers [1]. While SNSs provide excellent platforms for information dissemination among millions of users [2], they also attract hosts with impure motivations to exploit their massive influence for malicious activities, such as spam, click fraud, identity theft and phishing [3].

Unique feature of SNSs is that the information dissemination is primarily dependent on the *social trust* among users, [4], *e.g.*, a user's post is more likely to be reposted by his/her followers instead of others with no social tie. As a result, the one-time gain from a malicious action is positively related with the social trust of the malicious user, *i.e.*, the higher the social trust is, the more users will be influenced by the malicious action.

The social trust of a user reflects the confidence that this user will behave in an expected way, and can be evaluated by his/her frequency of non-malicious interactions with other users [4]. A positive interaction, *e.g.*, posing a trustworthy news, will improve the social trust of the user leading to larger influence for information dissemination, while a negative interaction, *e.g.*, maliciously spreading a rumor, will result in a degradation in the trust and hurting his/her potential of information dissemination in the future. Hence, for a malicious user aiming to maximize his/her overall personal benefits over a long time span, a tradeoff should be made between dynamically conducting positive and negative interactions with others,

e.g., obtaining malicious gain through negative interactions while accumulating better trust by positive interactions for larger malicious gain later.

It is desirable to understand the malicious host's best action strategy towards this tradeoff, and to accordingly propose optimal system maneuver mechanism for the social trust management so as to confine the malicious activities in the system. However, it is non-trivial to find the optimal balance between positive and negative interactions so as to maximize the long term malicious gain: how can we quantify the impact of an action on the future malicious gains and judiciously conduct positive/negative actions dynamically?

The difficulty further escalates when we practically extend the problem of optimizing the malicious gain at one individual user to the picture of interplays among multiple malicious users, who compete for the social trust, *i.e.*, interaction densities with other normal users, in order to selfishly maximize their own influence in information dissemination and thus malicious gains. Each action taken by an individual user will have an impact on the potential gain of other malicious users and vice versa. The following questions should be answered: how to evaluate the impact of an action on one's own and others' malicious gains in the future; what is the best strategy for each malicious user to dynamically adjust his/her positive/negative interactions in this competition?

Each user can be viewed as a node in the online social network. Our objective is to study the optimal response strategies of the malicious nodes in both single-node case and multiple-node case, respectively, such that we could find a better system maneuver accordingly in order to manage the trust evaluation and control the malicious activities. We propose an interaction-based *social trust* evaluation model, and formulate the single-node case as an optimal control problem and the competition among multiple nodes as a non-cooperative differential game. Through rigorous analysis, we solve the optimal response strategies for each node in both cases, on the basis of which we identify the best system maneuver mechanism given any targeted level of overall malicious activities.

The main contributions of this paper can be summarized as follows.

- We investigate *social trust* and its impact on the malicious information dissemination in SNSs.
- We propose a general framework to model the *social trust* using the frequency of interactions in the SNSs. Based on this model, we gain the insight for the administrators of SNSs to control the overall malicious activity.

- Through rigorous analysis, we identify the best response strategies for each node in both the single-node optimal control problem and the multiple-node differential game. Best system maneuver strategies are presented for each case, so as to maintain the overall malicious activities at any given level.
- Extensive numerical studies further verify our analytical results in differential system settings.

The rest of the paper is organized as follows. We highlight the related work in Section II. In Section IV, we present our model on the social trust and the threats from malicious nodes. We solve the multiple-node differential game for both static and dynamic cases in Section IV. Numerical studies under different system settings are presented in Section V. Finally, Section VI concludes the paper.

II. RELATED WORK

A. Social Trust and Trust Management

As an emerging research topic, social trust in social networks has been extensively discussed in [4] and references therein. The application of trust frameworks and systems in social networks involves defending malicious activities, especially the spamming [5] [6] [7]. As mentioned in [5], the behaviors of spammers are getting stealthy to evade from existing detection techniques. Yang *et al.* [6] state that malicious hosts can dilute their vicious posts and raise the opportunities to survive through mixing normal content with malicious content. To effectively eliminate the threat from spammers, Wang *et al.* [7] design an trust based collaborative spam mitigation system.

B. Game Theory in Cybersecurity

As a technique that naturally supports modeling decision-making for multiple agents, game theory has been extensively applied in security area. Hu *et al.* [8] and Feng *et al.* [9], [10] propose dynamic game models to analyze the interplay among attacker, defender and insider. Omic *et al.* [11] combine the epidemic model with game theory in order to derive the optimal protection mechanism against infection, whereas Zhu *et al.* [12] utilize differential games to analyze the infection process.

To the best of our knowledge, our work is the first in literature as the application of game theory for online social trust, and presents provably optimal system maneuver mechanism for SNSs.

III. SOCIAL TRUST AND THREAT MODEL

In this section, we first define the *social trust* in social networks and its dynamics based on the mutual *positive interactions*. Next, we formulate the threats from malicious nodes, and discuss the problem models for optimal tradeoff between positive and negative interactions. Important notations are summarized in Table I.

TABLE I: Important notations.

$\alpha_i(t)$	rate of posting trustable information from i
$\beta_i(t)$	rate of posting malicious information from i
$x_i(t)$	fraction of online users who are interacting with i at t
$\dot{x_i}(t)$	the evolving rate of x_i at each time point
x_{i0}	initial value of the x_i
$P_i(\cdot)$	long-term profit gain of i from negative activities
$C_{i1}(\cdot)$	long-term cost for positive activities of i
$C_{i2}(\cdot)$	long-term cost for negative activities of i
$\alpha_{-i}(t)$	action profile of positive activities for all players except i
$\beta_{-i}(t)$	action profile of negative activities for all players except i

A. Social trust

As a measurement of the confidence that an entity will behave in an expected way, trust moves to the center of data dissemination in SNSs. To build a trust community where users provide healthy information and feel free to share with each other, an effective and convenient trust system is required.

The interactions between a pair of users provides a natural way to assess one's *social trust* [4]. Users with high social trust draw more attention from others and involve high frequency of positive interaction with their neighbors, whereas un-trusted nodes get little attention and have limited influence of data dissemination over the SNSs. Current SNSs offer features to reflect one's social trust level based on users' reactions on the posted information, *e.g.*, Facebook users normally click "like" or "share" if they are in a comfortable interaction and they could choose to report a spam if they feel offended by the content.

In this paper, we use a general model to characterize one's social trust to the rest of the social network. Let N denote the total number of users in the social network. Let $X_i(t)$ denote the number of users that trust node i at time t, which is a random variable in general. We model the social trust of node i at as $x_i(t) = \mathbb{E}(X_i(t)/N)$, which evolves over time. Its dynamics is determined by its initial value $x_{i0} \in [0,1]$, which is a constant, and its actions on disseminating trustable/malicous information as discussed below. Alternatively, we can consider $x_i(t)$ as the fraction of nodes that interact with node i (assuming a node only interacts with the set of nodes that it trusts).

B. Dynamics of Social Trust

A malicious node delivers malicious content to as many users as possible for a profit. However, it does not target at a one-time profit from disseminating the malicious information. Instead, it tries to persistently make profits over a long time-span by continuously spreading malicious information. As discussed previously, social trust determines the influence of the information on the SNSs. Hence, a malicious node does not consistently provide pure baleful content to avoid diminishing its social trust and its information influence for later malicious actions. Instead, it moves stealthily by mixing good content with malicious content. It can either mix both type of content into one post or by posting these two in separate claims [5] [6]. By doing so, it maintains an acceptable level of social trust,

and makes a balance between its instantaneous malicious gain of current action and its future profits.

Single malicious node: Let us first consider the case of a single malicious node. Consider a malicious node i that posts some content $c(t)$ at time t. We model the impact of the content on the dynamics of social trust of node i by a pair of transition probabilities. Let $p_1(c(t), \delta)$ denote the probability that a node distrusting i at time t becomes trusting i at time $t + \delta$ after the content is posted for a small time period δ, which depends on both the content posted and δ. Similarly, let $p_2(c(t), \delta)$ denote the probability that a node trusting i at time t becomes distrusting i at time $t + \delta$. Intuitively, p_1 models the negative influence of malicious content through direct interaction with node i, and p_2 models the positive influence of benign content that propagates indirectly, e.g., through the "word-of-mouth" effect. In both cases, the influence is assumed to be independent across nodes. It follows that

$$\mathbb{E}(X_i(t + \delta) - X_i(t) | X_i(t)) = p_1(c(t), \delta)(N - X_i(t)) \\ - p_2(c(t), \delta)X_i(t), \quad (1)$$

Taking the expectation (with respect to $X_i(t)$) of both sides, we have

$$x_i(t + \delta) - x_i(t) = p_1(c(t), \delta)(1 - x_i(t)) - p_2(c(t), \delta)x_i(t). \quad (2)$$

Dividing both sides by δ and letting $\delta \to 0$, we obtain the following dynamics of social trust of node i,

$$\dot{x}_i(t) = \frac{dx_i(t)}{dt} = \alpha_i(t)(1 - x_i(t)) - \beta_i(t)x_i(t), \quad (3)$$
$$x_i(0) = x_{i0},$$

where $\alpha_i(t) = \lim_{\delta \to 0} \frac{p_1(c(t), \delta)}{\delta}$ and $\beta_i(t) = \lim_{\delta \to 0} \frac{p_2(c(t), \delta)}{\delta}$, which are assumed to exist. Instead of modeling the details of p_1 and p_2, we consider $(\alpha_i(t), \beta_i(t))$ as the strategy of node i in this work. We note that the differential equation is intuitive by itself. In particular, $\alpha_i(t)(1 - x_i(t))$ can be viewed as the social trust gained by posting trustable information that has positive response from $1 - x_i(t)$ (the share that is originally not positively interacting with node i); while $\beta_i(t)x_i(t)$ reflects the loss of social trust because of disseminating malicious content to $x_i(t)$ (the share that is originally positively interacting with node i). To simplify the description, we normalize $\alpha_i(t)$ and $\beta_i(t)$ so that $\alpha_i(t) + \beta_i(t) = 1$.

Multiple malicious nodes: Next, we consider the coexistence of multiple malicious nodes in the SNSs and the competition among them. As mentioned previously, social trust can be viewed as the frequency of interactions among users. In a continuous-time environment as in real-life applications, a content viewer in SNSs only involves in an effective interaction with one content provider at one time point. For instance, a user cannot click "like" for two separate posts concurrently at exactly the same time. Moreover, each online user has a limit *budget of attention* as suggested in [13]. The notion budget of attention quantifies the constraint on one's frequency of pulling content from the neighbors. Since attention is the foundation and the necessary condition for interaction, we extend the concept so as to characterize the upper bound exists on user's interaction rate.

Definition 1 (Budget of interaction): Budget of interaction is a constrained rate of a user that quantifies all kinds of its positive actions, which exclusively happen in continuous time at a social network site.

That is to say, malicious nodes have to compete with each other to gain social trust from their potential victims in order to maximize their individual profits.

Given n competing malicious nodes ($n > 1$) in a online social network, we assume that the sum of their social trust should be upper-bounded by the total interactions in the entire network, which is normalized to 1. That is,

$$\sum_i x_i(t) \leq 1. \quad (4)$$

Different from the previous single-node case, the dynamics of node i's social trust should consider the joint actions of all the malicious nodes and be formulated as follows,

$$\dot{x}_i(t) = \alpha_i(t)(1 - x_i(t)) - \sum_{j \in -i} \alpha_j(t)x_i(t) - \beta_i(t)x_i(t),$$
$$x_i(0) = x_{i0} \quad (5)$$

where $\alpha_i(t)(1 - x_i(t))$ and $\beta_i(t)x_i(t)$ have the same meaning as that in its counterpart with single malicious node; while $\sum_{j \in -i} \alpha_j(t)x_i(t)$ denotes the accumulated loss rate of social trust, that is obtained by other malicious nodes, *i.e.*, $j \in -i$, who post trustable information and attract the share that is originally positively interacting with node i. Above derivative equation captures the effect of "word-of-mouth" and diffusion progress, which are often used for advertising and marketing in economics field [14].

We can find from Eqn. (4) and Eqn. (5) that, each malicious node has to compete with each other for higher social trust, which leads to higher profit gain accordingly to Eqn. (6).

C. Payoff and Cost Functions for Malicious Nodes

The instantaneous malicious profit of node i at time t should be proportional to its malicious activity rate $\beta_i(t)$ and its social trust $x_i(t)$, *i.e.*, the amount of interactions could be influenced. Hence, the long-term profit gain P_i for node i is defined as follows,

$$P_i = \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i \beta_i(t) x_i(t) dt, \quad (6)$$

where p_i is the unit malicious profit for node i, with a positive constant value.

However, every activity comes with an operational cost. Both positive activity $\alpha_i(t)$ and negative activity $\beta_i(t)$ consume money in manpower at the malicious node. As commonly applied in literature [12] [14] [15] , we utilize the quadratic cost function to capture the instantaneous operational costs. The long-term costs for positive activities, C_{i1} and negative activities, C_{i2}, are evaluated as follows,

$$C_{i1} = \lim_{T \to \infty} \frac{1}{T} \int_0^T q_i \alpha_i^2(t) dt, \quad (7)$$

$$C_{i2} = \lim_{T \to \infty} \frac{1}{T} \int_0^T r_i \beta_i^2(t) dt, \quad (8)$$

where q_i and r_i are the unit cost of providing trustable content and the unit penalty for each malicious activity, respectively. p_i, q_i and r_i are all positive.

To sum up, the net profit for malicious node i is

$$P_i - C_{i1} - C_{i2}. \tag{9}$$

In the case of multiple malicious nodes, each of the malicious nodes acts independently and selfishly to maximize its individual net profit as defined in Eqn. (9).

D. System Maneuver

The objective of this paper is to find the optimal system maneuver mechanism, *i.e.*, configuration of the system parameters, in order to control the overall malicious activity within the targeted level.

The overall malicious activity is defined as i) $\beta_i(t)$ for the single-node case; and ii) $\sum_{i \in [1,n]} \beta_i(t)$ for the multiple-node case, when $\beta_i(t)$ has converged to its optimal strategy.

As for the system administrator, it can adjust the value of r_i, which could be the unit penalty for malicious activities of node i, at the start of the system so as to achieve its targeted level of overall malicious activity. Note that p_i and q_i are constants that are only related with the malicious node's setting while not controllable by the system administrator.

IV. SOCIAL TRUST GAMES

In this section, we study the competition among multiple malicious nodes and identify the best response strategy for each node. The competition is formulated into a non-cooperative differential game [15] that is continuously played among nodes. Note that the optimal control problem for the single malicious node setting can be easily derived from the game result and its result is given in our online technical report [16], since it can be viewed as a degenerate case of the differential game.

For each malicious node $i \in \{1, 2, ..., n\}$, it solves a profit-maximization problem in the game as follows,

$$\max \quad J_i(\alpha_i(t), \beta_i(t), \alpha_{-i}(t), \beta_{-i}(t))$$
$$= \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i \beta_i(t) x_i(t) - q_i \alpha_i^2(t) - r_i \beta_i^2(t) dt \tag{10}$$
$$\text{s.t.} \quad \dot{x}_i(t) = \alpha_i(t)(1 - x_i(t)) - \sum_{j \in -i} \alpha_j(t) x_i(t) - \beta_i(t) x_i(t),$$
$$x_i(0) = x_{i0}, \ \alpha_i(t), \beta_i(t) \in [0, 1], \ \alpha_i(t) + \beta_i(t) = 1,$$

We denote $\Phi(t) = \{\alpha_i(t), \alpha_{-i}(t); \beta_i(t), \beta_{-i}(t)\}$ as the strategy profile, where $\{\alpha_{-i}(t), \beta_{-i}(t)\}$ is the action set of malicious nodes other than i. $\phi_i(t) = \{\alpha_i(t), \beta_i(t)\}$ constitutes the strategy of i. Our objective is to derive the *open-loop* Nash Equilibrium (NE) defined as follows.

Definition 2: Consider the game described by Eqn. (16). The strategy profile $\Phi^*(t) = \{\phi_1^*(t), ..., \phi_n^*(t)\}$ constitutes a Nash

equilibrium solution if and only if, all following inequalities are satisfied

$$J_1(\phi_1^*(t), ..., \phi_n^*(t)) \geq J_1(\phi_1(t), ..., \phi_n^*(t)),$$
$$\vdots$$
$$J_n(\phi_i^*(t), ..., \phi_n^*(t)) \geq J_n(\phi_1^*(t), ..., \phi_n^*(t)).$$

Note that it is unrealistic for a malicious node to reveal its state to the competitors as the game evolves. Therefore, we consider the open-loop information structure in the game, which means that the players do not acquire further information except the common knowledge of the state vector at initial time $t = 0$ [15].

A. Static Case

We first analyze the static scenario of multiple competing malicious nodes, where the activity variables of all malicious nodes, *i.e.*, $\alpha_i(t)$ and $\beta_i(t)$ remain unchanged during the runtime of the game. The goal of each malicious node is to maximize the individual net profit through choosing its optimal action before the game starts. Based on the definition of its net profit as in Eqn. (9) and the dynamics of its social trust as in Eqn. (3), we can have the optimal control problem as follows (for simplicity, we denote $\alpha_i(t)$ and $\beta_i(t)$ as α_i and β_i since they are time-invariant in this subsection),

$$\max \quad J_i(\alpha_i, \alpha_{-i}) = \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i(1 - \alpha_i)x_i(t) - r_i(1 - \alpha_i)^2$$
$$- q_i \alpha_i^2 dt \tag{11}$$
$$\text{s.t.} \quad \dot{x}_i = \alpha_i - x_i(t) - \sum_{j \in -i} \alpha_j x_i(t), \quad x_i(0) = x_{i0}, \tag{12}$$
$$\alpha_i \in [0, 1]$$

where we have used the fact that $\alpha_i + \beta_i = 1$.

We obtain the fraction of users who involves positive interaction with the malicious node i in the SNS at time t through solving the ODE Eqn. (12).

$$x_i(t) = e^{-(1 + \sum_{j \in -i} \alpha_j)t} x_{i0} \tag{13}$$
$$+ \frac{\alpha_i}{1 + \sum_{j \in -i} \alpha_j}(1 - e^{-(1 + \sum_{j \in -i} \alpha_j)t})$$

Substituting the above $x_i(t)$ into the profit-maximization problem as defined in Eqn. (11) and Eqn. (12), we can simplify the problem into,

$$\max \quad \frac{p_i(1 - \alpha_i)\alpha_i}{1 + \sum_{j \in -i} \alpha_j} - r_i(1 - \alpha_i)^2 - q_i \alpha_i^2 \tag{14}$$
$$s.t. \quad \alpha_i \in [0, 1]$$

We can derive the best response of the malicious node i by analyzing the structure of Eqn. (14) and the proof is given in our online technical report [16].

Proposition 1: For the static case of multiple competing malicious nodes, the best response for the malicious node i, where $i = 1, ..., n$ is given by

$$\alpha_i^* = \frac{p_i + 2r_i(1 + \sum_{j \in -i} \alpha_j)}{2[p_i + q_i + r_i + (q_i + r_i) \sum_{j \in -i} \alpha_j]} \tag{15}$$

Theorem 1: There exists a Nash equilibrium for the static social trust game.

Proof: Let $B_i(a_{-i}) = a_i^* : [0,1] \rightarrow [0,1]$ be the best response function of i. The action set [0, 1] is compact and convex. Also, the best response function B_i is continuous over [0, 1]. Thus, there exists a fixed point that satisfies the equation $\alpha_i^* = B_i(\alpha_i^*)$ based on Brouwer's fixed point theorem [17]. Since a NE satisfies the fixed point equation, we prove the existence of NE. ∎

B. Dynamic Case

In this subsection, we analyze the dynamic case for the game among multiple malicious nodes. The general term of dynamics has been described in Eqn. (5). For each malicious node $i \in \{1, 2, ..., n\}$, it solves a profit-maximization problem in the game as follows (using the fact that $\alpha_i(t) + \beta_i(t) = 1$),

$$\max \quad J_i(\alpha_i, \alpha_{-i}) = \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i(1 - \alpha_i(t))x_i(t) - q_i\alpha_i^2(t)$$
$$- r_i(1 - \alpha_i(t))^2(t)dt \quad (16)$$
$$\text{s.t.} \quad \dot{x}_i(t) = \alpha_i(t) - x_i(t) - \sum_{j \in -i} \alpha_j(t)x_i(t),$$
$$x_i(0) = x_{i0}, \quad \alpha_i(t) \in [0,1].$$

We follow the procedure in [18] to look for the open-loop Nash equilibrium.

Lemma 1: For the dynamic case of multiple competing malicious nodes, the best response of the malicious node $i = 1, ..., n$) is given by

$$\alpha_i^*(t) = \begin{cases} \frac{\lambda_i(t) - p_i x_i(t) + 2r_i}{2(q_i + r_i)} & \lambda_i(t) > p_i x(t) - 2r_i, \\ 0 & \text{otherwise.} \end{cases} \quad (17)$$

Proof: To obtain the best response of each malicious nodes, we solve an optimal control problem for i. The Hamiltonian function of i is denoted by H_i as:

$$H_i(\alpha_i, \alpha_{-i}, t) = \lambda_i(t)(\alpha_i(t) - x_i(t) - \sum_{j \in -i} \alpha_j(t)x_i(t))$$
$$+ p_i(1 - \alpha_i(t))x_i(t) - q_i\alpha_i^2(t) - r_i(1 - \alpha_i)^2(t)$$

where $\lambda_i(t)$ is the co-state variable attached to $x_i(t)$.

We apply the Pontryagin maximum principle to derive the best response of i,

$$\alpha_i^*(t) = \text{argmax} \ \{H_i(\alpha_i(t), \alpha_{-i}(t), x_i(t), \lambda_i(t))\}.$$

The state space is compact and convex. Also, H_i is concavity and differentiable with respect to control $\alpha_i(t)$. We then obtain the best response Eqn. 17 by solving $\frac{\partial H_i}{\partial \alpha_i} = 0$. ∎

At each time instance, we are able to solve for explicit value of $\alpha_i^*(t)$ through numerical approach from following Pontryagin necessary conditions

$$\frac{\partial H_i}{\partial \alpha_i} = 0, \quad -\frac{\partial H_i}{\partial x_i} = \dot{\lambda}_i.$$

We can now acquire the open-loop NE in steady status from Pontryagin necessary conditions (see our online technical report [16] for the proof).

Theorem 2: The best response of malicious node i at the open-loop equilibrium is given by

$$\alpha_i^*(t) = \frac{p_i + 2r_i(1 + \sum_{j \in -i} \alpha_j(t))}{2[p_i + q_i + r_i + (q_i + r_i)\sum_{j \in -i} \alpha_j(t)]}. \quad (18)$$

Remark 1: In the steady status, the optimal dynamic control coincides with the static solution for the single malicious node setting ($n = 1$).

Remark 2: For n non-cooperative malicious nodes, we are able to get n best responses in steady status separately. Thus, we have n simultaneous equations with n unknown variables. It is trivial to obtain explicit solutions in some cases using analytical or numerical techniques.

To bound our analysis in a controllable scope, we take the situation of two symmetric players as a simple illustration. The symmetric means that the payoff and the cost factors are same for two players. We simply denote them as p, q and r respectably.

Corollary 1: Consider two symmetric competing malicious nodes exist in the SNS, the optimal system maneuver is given by

$$r^*(t) = (p + q)(3 - 2\beta(t))^2 - \frac{1}{4}(3p + q), \quad (19)$$

where $\beta(t)$ is the control of malicious behavior from two symmetric malicious nodes in steady status.

Proof: From Theorem (2), we know that

$$\beta^*(t) = \beta_i^*(t) = \beta_j^*(t) = \frac{1}{2}(3 - \frac{\sqrt{(p+q)(3p+q+4r)}}{2(p+q)}).$$

which can be used to derive the system maneuver $r^*(t)$. ∎

We can further obtain the following corollary for general situations.

Corollary 2: Consider the game among two non-cooperative players i and j. At the open-loop equilibrium, the best response function of i is negatively sloped for all p_i, q_i and $r_i \in (0, \infty]$. In absolute value, the slope is everywhere decreasing in r_i.

Proof: The slope of best response function at the open-loop equilibrium is given by

$$\frac{\partial \alpha_i}{\partial \alpha_j} = -\frac{4r_i(q_i + r_i) + p_i(q_i + 3r_i)}{2(p_i + (q_i + r_i)(1 + \alpha_j)^2)} < 0.$$

∎

V. NUMERICAL STUDY

In this section, we illustrate the results with numerical examples. We build our simulation on Matlab platform with bvp4c toolbox.

Suppose there is an existing malicious node that has already reached its steady state. Now we introduce another homogenous malicious node with identical configurations with the existing node. Let $p = 0.4, q = 0.2, r = 0.2$ for both nodes.

From Fig. 1(a), we can observe that the player I deviates from its previously steady state 0.5 and its $x_1(t)$ begins decreasing, meanwhile, $x_2(t)$ of player II starts from 0 and increases until finally converging to the steady position, which

(a) state trajectory for two identical players   (b) variation of α^* and β^* when n increases   (c) trajectory of the states and the controls for two players with different system maneuvers
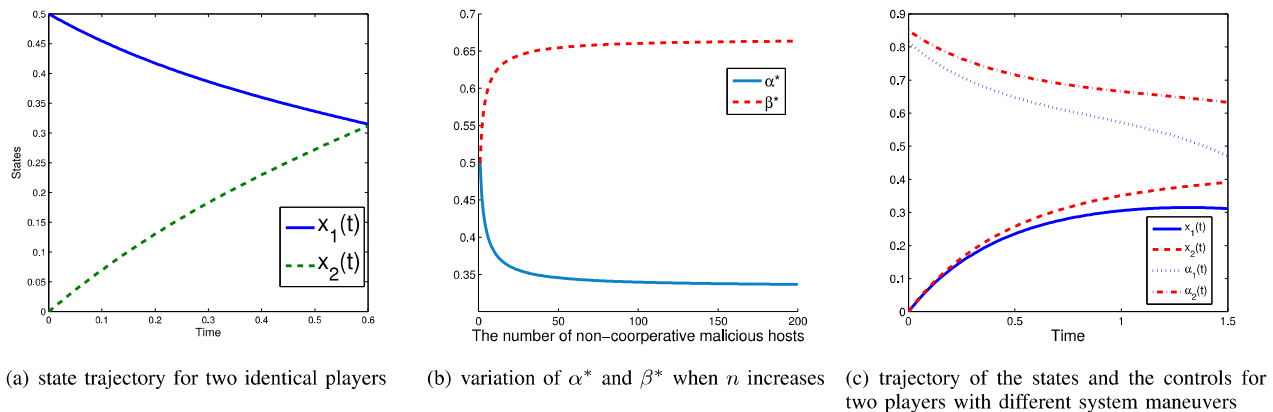
Fig. 1: Multiple Competing Malicious Nodes

matches the analytical result for steady position. Since the factors of two players are symmetric, it is not surprise that two players finally converge to the same state.

Next, we examine how the control in the steady state evolves when the amount of players increases. The parameters are set as same, *i.e.*, $p = 0.4$ and $q = r = 0.2$. As shown in Fig. 1(b), $\alpha^*(t)$ begins at 0.5 and converges to 0.35 when n increases, whereas β^* starts from 0.5 and converges to 0.65. This observation means that the competition does not motivate good behaviors by nodes.

We then study the influences of the system maneuver r_i on the controls and the states of a two-player game scenario. Let $p_1 = p_2 = 0.5$ and $q_1 = q_2 = 0.1$, Fig. 1(c) depicts the evolution progress of the states and the controls of two players with $r_1 = 0.2$ and $r_2 = 0.3$ separately. We can see that the higher system maneuver comes with the lower negative activity rate in social trust games.

VI. CONCLUSION

This paper investigates the social-trust-based information dissemination by malicious nodes in social network sites. An interaction-based social trust model is presented. For studying the best response strategies of malicious nodes with a long-term objective, we formulate the maximization of malicious gains in a long time-span of multiple competing malicious nodes as a non-cooperative differential game. Through rigorous analysis, optimal response strategies for each malicious node are identified and the best system maneuver mechanisms are presented in order to achieve the targeted level of overall malicious activities in the system. The numerical studies further verify the analytical results.

ACKNOWLEDGEMENTS

REFERENCES

[1] N. B. Ellison *et al.*, "Social network sites: Definition, history, and scholarship," *Journal of Computer-Mediated Communication*, vol. 13, no. 1, pp. 210–230, 2007.

[2] Y. Boshmaf, I. Muslukhov, K. Beznosov, and M. Ripeanu, "The socialbot network: when bots socialize for fame and money," in *ACSAC*, 2011.

[3] K. Thomas, C. Grier, J. Ma, V. Paxson, and D. Song, "Design and evaluation of a real-time url spam filtering service," in *SP*, 2011.

[4] W. Sherchan, S. Nepal, and C. Paris, "A survey of trust in social networks," *ACM Computing Surveys (CSUR)*, vol. 45, no. 4, 2013.

[5] G. Stringhini, C. Kruegel, and G. Vigna, "Detecting spammers on social networks," in *ACSAC*, 2010.

[6] C. Yang, R. C. Harkreader, and G. Gu, "Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers," in *Recent Advances in Intrusion Detection*, pp. 318–337, Springer, 2011.

[7] X. Wang, H. Fu, C. Xu, and P. Mohapatra, "Provenance logic: Enabling multi-event based trust in mobile sensing," in *IPCCC*, 2014.

[8] P. Hu, H. Li, H. Fu, D. Cansever, and P. Mohapatra, "Dynamic defense strategy against advanced persistent threat with insiders," in *INFOCOM*, 2015.

[9] X. Feng, Z. Zheng, D. Cansever, A. Swami, and P. Mohapatra, "Stealthy attacks with insider information: A game theoretic model with asymmetric feedback," in *MILCOM*, 2016.

[10] X. Feng, Z. Zheng, P. Hu, D. Cansever, and P. Mohapatra, "Stealthy attacks meets insider threats: A three-player game model," in *MILCOM*, 2015.

[11] J. Omic, A. Orda, and P. Van Mieghem, "Protecting against network infections: A game theoretic perspective," in *INFOCOM*, 2009.

[12] Q. Zhu, L. Bushnell, and T. Basar, "Game-theoretic analysis of node capture and cloning attack with multiple attackers in wireless sensor networks," in *CDC*, 2012.

[13] B. Jiang, N. Hegde, L. Massoulié, and D. Towsley, "How to optimally allocate your budget of attention in social networks," in *INFOCOM*, 2013.

[14] S. Jørgensen, "A survey of some differential games in advertising," *Journal of Economic Dynamics and Control*, vol. 4, pp. 341–369, 1982.

[15] Z. Han, *Game theory in wireless and communication networks: theory, models, and applications*. Cambridge University Press, 2012.

[16] H. Fu, H. Li, Z. Zheng, P. Hu, and P. Mohapatra, "Trust exploitation and attention competition: A game theoretic model." Technical Report, available online at http://spirit.cs.ucdavis.edu/pubs/tr/fu-trust-report.pdf.

[17] K. C. Border, "Fixed point theorems with applications to economics and game theory," *Cambridge Books*, 1990.

[18] R. Cellini and L. Lambertini, "A differential oligopoly game with differentiated goods and sticky prices," *European Journal of Operational Research*, vol. 176, no. 2, pp. 1131–1144, 2007.