Optimal System Maneuver for Trust Management in Social Networks

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 hosts 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 host 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 in the cases of i) single independent host, and ii) multiple competing hosts as an optimal control problem and a non-cooperative differential game, respectively. Through rigorous analysis, optimal response strategies are identified for both cases, based on which the best system maneuver mechanism is presented. Extensive numerical studies further verify the analytical results.

I. INTRODUCTION

We have witnessed the prevailing usage 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, i.e., 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?

Our objective is to study the optimal response strategies of the malicious hosts in both single-host case and multiple-host 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-host case as an optimal control problem and the competition among multiple hosts as a non-cooperative differential game. Through rigorous analysis, we solve the optimal response strategies for each host in both cases, on the basis of which we identify the best system maneuver mechanism given any targeted level of overall malicious activities.

The contribution 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, both single-host and multiple-host cases are studied to gain the insight for the administrators of SNSs.

Through rigorous analysis, we identify the best response strategies for each host in both the single-host optimal control problem and the multiple-host 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 V, we present our model on the social trust and the threats from malicious hosts. We solve the single-host optimal control problem and the multiple-host differential game for both static and dynamic cases in Sec. V and Sec. VI respectively. Numerical studies under different system settings are presented in Sec. VII. Finally, Section VIII 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. Social trust is an assessment of the given host’s honest intention in a social network. In order to create and maintain trust communities, many interaction-based trust models have been proposed to evaluate the trustworthiness of a given host in online social networks [5] [6]. Adali et al. [5] propose a trust evaluation method based on communication behaviors, which can be further categorized into the popularity and the engagement trust. Tricfunovic et al. [6] separate social trust into two classes, where implicit trust is a measure of frequency and duration of the contact among a pair of members.

The application of trust frameworks and systems in social networks involves defending malicious activities, especially the spamming [7] [8] [9] [10]. As mentioned in [7], the behaviors of spammers are getting stealthy to evade from existing detection techniques. Yang et al. [8] 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, Sirivianos et al. [9] design an ad-hoc social trust based collaborative spam mitigation system. Hameed et al. [10] introduce a spam protection system depends on social circles and trust.

B. Optimal Control and Game Theory in Network Security

In security and networking field, optimal control has been used to figure out the best disposition for underlying resources [11]. Karnik et al. [12] employ optimal control to characterize the epidemic model in communication networks.

As a technique that naturally supports modeling decision-making for multiple agents, game theory has been extensively applied in security area. Alpcan and Basar [13] survey the application of game theory in network security. Hu et al. [14] propose a dynamic game model to analyze the interplay among attacker, defender and insiders. Omi et al. [15] combine the epidemic model with game theory in order to derive the optimal protection mechanism against infection, whereas Zhu et al. [16] 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 in 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 hosts, and discuss the problem models for optimal tradeoff between positive and negative interactions in two scenarios: i) optimal control for single malicious host; and ii) optimal response strategy for multiple malicious hosts in a non-cooperative competition. Important notations are summarized in Table 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 hosts get little attention and has 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. We normalize the total interactions in the SNSs into 1. Let $x_i(t) \in [0, 1]$ be

| Notation | Description |
|----------|-------------|
| $\alpha_i(t)$ | 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 time $t$ |
| $\dot{x}_i(t)$ | the evolving rate of $x_i$ at each time point |
| $x_0$ | initial value of the $x_i$ |
| $P_i(t)$ | long-term profit gain of $i$ from negative activities |
| $C_i(t)$ | long-term cost for positive activities of $i$ |
| $C_{neg}(t)$ | 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 $i$ |
the malicious host $i$’s share in the total interactions at time $t$, which is its social trust to the others in the social network at that time point. The initial value $x_i(0)$ is $x_{i0} \in [0, 1]$, which is a constant value given by the system setting. The value of $x_i(t)$ evolves over time. Its dynamics should be impacted by its actions, e.g., disseminating trustable/malicious information, (in the case with single malicious host) and the actions of other malicious hosts (in the case with multiple malicious hosts). Its details will be addressed in Sec. III-C and III-D for cases with single and multiple malicious hosts, respectively.

B. Payoff and Cost Functions for Malicious Hosts

A malicious host 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 host does not consistently provide pure baleful content to avoid diminishing its social trust and its information influence for later malicious actions. Instead, they move stealthily by mixing good content with malicious content. They can either mix both type of content into one post or by posting these two in separate claims [7] [8]. By doing so, they maintain an acceptable level of social trust, and make a balance between its instantaneous malicious gain of current action and its future profits.

We normalize the overall activity rate, i.e., having an interaction with other users, of each malicious host into 1. Let $\alpha_i(t) \in [0, 1]$ and $\beta_i(t) \in [0, 1]$ be the controls on the rate of posting trustable and malicious information at time $t$, respectively. We should have that $\beta_i(t) = 1 - \alpha_i(t)$.

The instantaneous malicious profit of host $i$ 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 host $i$ is defined as follows,

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

where $p_i$ is the unit malicious profit for host $i$, 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 host. As commonly applied in literature [16] [17] [18], 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,

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

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

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

To sum up, the net profit for malicious host $i$ is

$$P_i - C_{i1} - C_{i2}.$$

C. Single malicious host

In the case with a single malicious host, it judiciously takes control on the rate of positive and negative activities with users in the social network so as to maximize its long-term net profit as defined in Eqn. (4).

The dynamics of its social trust can be evaluated as follows,

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

where $\alpha_i(t)(1 - x_i(t))$ denotes the social trust gained by posting trustable information, i.e., $\alpha_i(t)$, that has positive response from $1 - x_i(t)$ (the share that is originally not positively interacting with host $i$); while $\beta_i(t)x_i(t)$ reflects the loss of social trust because of disseminating malicious content $\beta_i(t)$ to $x_i(t)$ (the share that is originally positively interacting with host $i$).

D. Multiple malicious hosts

Next, we practically consider the coexistence of multiple malicious hosts in the SNSs and the competition among them. Each of the malicious hosts acts independently and selfishly to maximize its individual net profit as defined in Eqn. (4).

As mentioned previously, social trust measures 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 [19]. 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 hosts ($n > 1$) in a online social network, we have that the sum of their social trust should be upper-bounded by the total interactions in the entire network as follows,

$$\sum_{i} x_i(t) \leq 1.$$  

Different from the previous single-host case, the dynamics of host $i$’s social trust should consider the joint actions of all
the malicious hosts and be formulated as follows,

\[
\dot{x}_i(t) = \alpha_i(t)(1 - x_i(t)) - \sum_{j \in -i} \alpha_j(t)x_j(t) - \beta_i(t)x_i(t),
\]

\[x_i(0) = x_{i0}\]

(7)

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 host; while \(\sum_{j \in -i} \alpha_j(t)x_j(t)\) denotes the accumulated loss rate of social trust, that is obtained by other malicious hosts, i.e., \(j \in -i\), who post trustworthy information and attract the share that is originally positively interacting with host \(i\). Above derivative equation captures the effect of "word-of-mouth" and diffusion progress, which are often used for advertising and marketing in economics field [17].

We can find from Eqn. (6) and Eqn. (7) that, each malicious host has to compete with each other for higher social trust, which leads to higher profit gain accordingly to Eqn. (7).

E. 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-host case; and ii) \(\sum_{j \in -i} \beta_j(t)\) for the multiple-host 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 host \(i\), 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 host’s setting while not controllable by the system administrator.

IV. THE SINGLE MALICIOUS HOST

In this section, we first analyze the behavior of a single malicious host in SNSs, which is an optimal control problem as defined in Sec. III-C. We propose optimal solutions to two cases: the static case and the dynamic case, with details in Sec. IV-A and Sec. IV-B respectively. Based on the optimal controls for the malicious host, we present the method to determine the system maneuver for controlling the malicious activity into any targeted level.

A. Static Case

In this subsection, we analyze the behavior of a single malicious host in the static case, where the rates of trustworthy and malicious actions are time-invariant and preconfigured before the deployment of the system. The malicious host aims to maximize its own net profit with optimal control. Based on the definition of its net profit as in Eqn. (4) and the dynamics of its social trust as in Eqn. (7), we can have the profit-maximization problem at each host \(i\) as follows (for simplicity, we denote \(\alpha_i(t)\) and \(\beta_i(t)\) as \(\alpha_i\) and \(\beta_i\) since they are time-invariant in this subsection),

\[
\max \lim_{T \to \infty} \frac{1}{T} \int_0^T \left[ p_i\beta_i x_i(t) - q_i\alpha_i^2 - r_i\beta_i^2 \right] dt
\]

s.t. \(\dot{x}_i = \alpha_i(1 - x_i(t)) - \beta_i x_i(t), \quad x_i(0) = x_{i0}, \quad \alpha_i, \beta_i \in [0,1], \quad \alpha_i + \beta_i = 1.\)

Since \(\beta_i\) can be directly computed through \(\alpha_i\), we use \(\alpha_i\) to denote the action of the malicious host. The optimal control problem can be converted into the following form,

\[
\max \lim_{T \to \infty} \frac{1}{T} \int_0^T \left[ p_i(1 - \alpha_i)x_i(t) - r_i(1 - \alpha_i)^2 - q_i\alpha_i^2 \right] dt
\]

s.t. \(\dot{x}_i = \alpha_i - x_i(t), \quad x_i(0) = x_{i0}, \quad \alpha_i \in [0,1]\).

(9)

By solving ordinary differential equation (10), we can acquire the fraction of online users who involve positive interaction with the given malicious host in the SNS at time \(t\) as

\[
x_i(t) = e^{-t}x_{i0} + \alpha_i(1 - e^{-t}).
\]

(11)

Substituting \(x_i(t)\) into Eqn. (9) and Eqn. (10), the optimal control problem for the malicious host can be further simplified into

\[
\max p_i(1 - \alpha_i)\alpha_i - r_i(1 - \alpha_i)^2 - q_i\alpha_i^2,
\]

s.t. \(\alpha_i \in [0,1]\).

We solve the above optimization problem with the following theorem.

Theorem 1: Consider the static case of single malicious host problem, an unique optimal control exists in \([0,1]\) and is given by

\[
\begin{align*}
\alpha^* &= \frac{p_i + 2r_i}{2(p_i + q_i + r_i)}, \\
\beta^* &= \frac{p_i + q_i + r_i}{2(p_i + q_i + r_i)}.
\end{align*}
\]

Proof: We define the objective function \(F(\alpha) = p(1 - \alpha_i)\alpha_i - r_i(1 - \alpha_i)^2 - q_i\alpha_i^2\). Taking the second order derivative of \(F(\alpha)\) we have

\[
F''(\alpha) = -2p_i - 2(q_i + r_i).
\]

As \(p\), \(q\) and \(r\) can only be positive values, the second order derivative \(F''(\alpha) < 0\) and it reflects the concavity of \(F(\alpha)\). Then checking the value of first order derivative at the end points,

\[
\begin{align*}
F'(0) &= 2r_i + p_i > 0, \\
F'(1) &= -2q_i - p_i < 0,
\end{align*}
\]

which reflects that the uniqueness of optimal solution \(\alpha_i^*\) and it exists in \((0,1)\). The value of \(\alpha_i^*\) can be directly computed from the first order derivative of \(F(\alpha)\),

\[
F'(\alpha_i) = p_i + 2r_i - 2(p_i + q_i + r_i)\alpha_i = 0.
\]
With Theorem [1], we can have the optimal system maneuver with the following theorem.

**Theorem 2:** Given a targeted level of malicious activity $\beta_i$, a unique optimal system maneuver exists for $r_i$ in the static single malicious host case and is given by

$$r_i^* = \frac{p_i + 2q_i}{2\beta_i} - p_i - q_i.$$

The proof to this theorem can be trivially derived from Theorem [1] and omitted for limited space.

**B. Dynamic Case**

In this subsection, we analyze the dynamic case for the optimal control problem of single malicious host, where the control rates, i.e., $\alpha_i(t)$ and $\beta_i(t)$, are dynamically decided and adapted to the system status, i.e., $x_i(t)$, at time $t$.

The optimal control problem is defined the same as follows,

$$\max \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i\beta_i x_i(t) - q_i\alpha_i^2(t) - r_i\beta_i^2(t) dt$$

s.t. \( \dot{x}_i(t) = \alpha_i(t)(1 - x_i(t)) - \beta_i(t)x_i(t), \quad x_i(0) = x_{i0}, \quad \alpha_i(t), \beta_i(t) \in [0,1], \quad \alpha_i(t) + \beta_i(t) = 1. \)

Let functions $f(\cdot)$ and $g(\cdot)$ denote the dynamics of social trust and the instantaneous net profit, respectively. We have that,

$$f(x_i(t), \alpha_i(t), \beta_i(t)) = \dot{x}_i(t) = \alpha_i(t)(1 - x_i(t)) - \beta_i(t)x_i(t), \quad g(x_i(t), \alpha_i(t), \beta_i(t)) = p_i\beta_i x_i(t) - q_i\alpha_i^2(t) - r_i\beta_i^2(t).$$

**Theorem 3:** Consider the dynamic single malicious host optimal control problem described in (15), there exists an unique optimal control solution at $t$.

**Proof:** To obtain the optimal control, we utilize the Pontryagin Maximum Principle [18]. We define the Hamiltonian function $H$ as

$$H(\alpha_i(t), \beta_i(t), x_i(t), \lambda_i(t)) = \lambda_i(t)f(x_i(t), \alpha_i(t), \beta_i(t)) + g(x_i(t), \alpha_i(t), \beta_i(t))$$

where $\lambda_i(t)$ is the costate (or adjoint) variable associated to state variable $x_i(t)$.

Since $\beta_i(t)$ can be substituted by $1 - \alpha_i(t)$, we simply the problem as finding a control $\alpha_i^*(t)$ that satisfies

$$\alpha_i^*(t) = \arg \max H(\alpha_i(t), x_i(t), \lambda_i(t)).$$

The state space $[0,1]$ is a compact set. Also, the Hamiltonian function $H$ is concave and continuously differentiable with respect to $\alpha_i(t)$. Thus, an unique optimal control exists.

We move forward to solve for $\alpha_i^*(t)$ for each time instance. 

**Lemma 1:** The optimal control solution for the dynamic single malicious host case is

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

**Proof:** From Eqn. (14), we can rewrite the $H$ as

$$H(\alpha_i(t), x_i(t), \lambda_i(t)) = \lambda_i(t)(\alpha_i(t) - x_i(t)) + p_i(1 - \alpha_i(t))x_i(t) - q_i\alpha_i^2(t) - r_i(1 - \alpha_i(t))^2(t).$$

Solving the following first order condition,

$$\frac{\partial H_i}{\partial \alpha_i} = -2\alpha_i(t)(q_i + r_i) + \lambda_i(t) - p_i x_i(t) + 2r_i = 0,$$

we acquire the optimal solution as Eqn. (15).

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

$$\frac{\partial H_i}{\partial \alpha_i} = 0,$$

$$\frac{\partial H_i}{\partial x_i} = \lambda_i.$$

When $x_i(t)$ reaches at its steady state, we can solve the costate variable analytically.

**Theorem 4:** As the state variable $x_i(t)$ of the dynamic single malicious host case converges to steady status, the optimal control is

$$\alpha_i^*(t) = \frac{p_i + 2q_i}{2(p_i + q_i + r_i)}, \quad \beta_i^*(t) = \frac{p_i + 2q_i}{2(p_i + q_i + r_i)},$$

which coincides with the solution for the static case of single malicious host optimal control problem.

**Proof:** Differentiating Eqn. (15) and recalling the adjoint equation Eqn. (19), we obtain

$$\dot{\alpha}_i(t) = \frac{1}{2(q_i + r_i)}(\dot{\lambda}_i(t) - p_i \dot{x}_i(t))$$

$$\dot{x}_i(t) = \frac{1}{2(q_i + r_i)}(\lambda_i(t) - p_i x_i(t) + 2r_i - p_i(\alpha_i(t) - x_i(t))).$$

Then substitute $\lambda_i(t) = 2(q_i + r_i)\alpha_i^*(t) + p_i x_i(t) - 2r_i$ derived from Eqn. (15), which yields

$$\dot{\alpha}_i(t) = \frac{1}{2(q_i + r_i)}(2\alpha_i(t)(q_i + r_i) + 2p_i x_i(t) - 2r_i - p_i)$$

$$= 0,$$

in

$$\alpha_i^*(t) = \frac{p_i + 2r_i - 2p_i x_i(t)}{2(q_i + r_i)}.$$  \hspace{1cm} (20)

Since $x_i(t)$ finally converge to stationary status in open-loop solution, we have

$$\dot{x}_i(t) = 0.$$

From (5), we then have

$$x(t) = \alpha_i(t),$$

which can be plugged back to (20) to solve for $\alpha^*(t)$.

With Theorem [4], we can have the optimal system maneuver with the following theorem.
Theorem 5: Given a targeted level of malicious activity $\beta_i$ when the system is in its stable status, a unique optimal system maneuver exists for $r_i$ in the dynamic single malicious host case and is given by

$$r^*_i = \frac{p_i + 2q_i}{2\beta_i} - p_i - q_i.$$  

The proof to this theorem can be trivially derived from Theorem 5 and omitted for limited space.

Remarks: Since $p_i$ (the unit malicious profit for host $i$) and $q_i$ (the unit cost of providing trustable information for host $i$) are uncontrollable parameters for the system administrator, the value of $r^*_i$ should be set inversely proportional to the targeted malicious activity level $\beta_i$. Intuitively, the system administrator should increase the unit penalty for each malicious host when it targets on a lower malicious activity. The contribution of Theorem 5 is deriving the analytical relation between the targeted malicious activity level and its corresponding system setting.

V. Social Trust Games

In this section, we study the competition among multiple malicious hosts and identify the best response strategy for each host. The competition is formulated into a non-cooperative differential game [18] that is continuously played among hosts.

In this subsection, our objective is to derive Nash equilibrium (NE) solutions for the differential game according to Eqn. (24) and Eqn. (25). 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 hosts other than $i$. $\phi_i(t) = \{\alpha_i(t), \beta_i(t)\}$ constitutes the strategy of $i$. We have the definition of the NE for the this differential game as follows.

Definition 2: Consider the game described by Eqn. (20). The strategy profile $\Phi^*(t) = \{\phi^*_i(t), ..., \phi_{n}^*(t)\}$ constitutes a Nash equilibrium solution if and only if, all following inequalities are satisfied

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

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

A. Static Case

Similar to the optimal control problem, we first analyze the static scenario of multiple competing malicious hosts. The activity variables of all malicious hosts, i.e., $\alpha_i(t)$ and $\beta_i(t)$ remain unchanged during the runtime of the game.

The goal of each malicious host 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. (3) and the dynamics of its social trust as in Eqn. (5), we can have the optimal control problem as follows (for simplicity, we denote $\alpha_i(t)$ and $\beta_i(t)$ as $\alpha_i$ and $\beta_i$ since they are time-invariant in this sub-section),

$$\max \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i(1 - r_i(t)) - q_i^2 - r_i^2 dt$$  

s.t. $x_i = \alpha_i(1 - x_i(t)) - \sum_{j \in S_i} \alpha_j x_j(t) - \beta_i x_i(t),$

$$x_i(0) = x_{i0},$$

$$\alpha_i, \beta_i \in [0, 1] \quad \alpha_i + \beta_i = 1.$$  

Since $\beta_i$ can be directly computed through $\alpha_i$ with $\alpha_i + \beta_i = 1$, we use $\alpha_i$ to denote the action set of $i$ and simplify the profit-maximization problem into

$$\max \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i(1 - \alpha_i) x_i(t) - r_i(1 - \alpha_i)^2 - q_i^2 dt$$  

s.t. $x_i = \alpha_i - x_i(t) - \sum_{j \in S_i} \alpha_j x_j(t), \quad x_i(0) = x_{i0},$

$$\alpha_i \in [0, 1]$$

We obtain the fraction of users who involves positive interaction with the malicious host $i$ in the SNS at time $t$ through solving the ODE Eqn. (25).

$$x_i(t) = e^{-\frac{1 + \sum_{j \in S_i} \alpha_j(t)}{t} + \frac{\alpha_i}{1 + \sum_{j \in S_i} \alpha_j(t)} - \frac{1 - \alpha_i}{t}}$$  

Substituting the above $x_i(t)$ into the profit-maximization problem as defined in Eqn. (24) and Eqn. (25), we can simplify the problem into,

$$\max \frac{p_i(1 - \alpha_i) x_i}{1 + \sum_{j \in S_i} \alpha_j} - r_i(1 - \alpha_i)^2 - q_i^2$$  

s.t. $\alpha_i \in [0, 1]$  

We can derive the best response of the malicious host $i$ by analyzing the structure of Eqn. (27).

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

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

Proof: The proof is similar to that of Theorem 1. The objective function of the malicious host $i$ is defined by $F_i(\alpha_i) = \frac{p_i(1 - \alpha_i) x_i}{1 + \sum_{j \in S_i} \alpha_j} - r_i(1 - \alpha_i)^2 - q_i^2$. The second order derivative of $F_i(\alpha_i)$ is

$$F''(\alpha_i) = -\frac{2p_i}{(1 + \sum_{j \in S_i} \alpha_j)} - 2(q_i + r_i) < 0.$$  

Thus, $F(\alpha_i)$ is a concave function w.r.t $\alpha_i$. By checking the value of first derivative at the end points, we know that the optimal action $\alpha_i^*$ exists in domain $(0, 1)$. Let the first derivative
of $F(\alpha_i)$ equals to 0,
\[
F'(\alpha_i) = -2\alpha_i\left(\frac{p_i}{\sum_{j \neq i} m_j \alpha_j + 1} + q_i + r_i\right) \\
+ \frac{p_i}{\sum_{j \neq i} m_j \alpha_j + 1} + 2r_i = 0.
\]

Hence, we obtain the value of $\alpha_i^*$ as defined in Eqn. (28), which constitutes the best response of $i$.

**Theorem 6:** There exists a Nash equilibrium for the static social trust game.

**Proof:** Let $B_i(a_{-i}) = a_i^* : [0, 1] \to [0, 1]$ 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 [20]. 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 hosts. The general term of dynamics has been described in Eqn. (7). For each malicious host $i \in \{1, 2, ..., n\}$, it solves a profit-maximization problem in the game as follows,

\[
\max \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i \beta_i(t) x_i(t) - q_i \alpha_i(t) - r_i \alpha_i^2(t) dt
\]

s.t. 
\[
\dot{x}_i(t) = \alpha_i(t)(1 - x_i(t)) - \sum_{j \neq i} \alpha_j(t)x_j(t) - \beta_i(t)x_i(t),
\]

\[
x_i(0) = x_{i0},
\]

\[
\alpha_i(t), \beta_i(t) \in [0, 1], \quad x_i(t) + \beta_i(t) = 1,
\]

which can be further simplified by substituting each $\beta_i(t)$ with $1 - \alpha_i(t)$ into the following,

\[
\max J_i(\alpha_i, \alpha_{-i}) = \lim_{T \to \infty} \frac{1}{T} \int_0^T p_i(1 - \alpha_i(t))x_i(t) - q_i \alpha_i^2(t) - r_i(1 - \alpha_i(t))^2 dt
\]

s.t. 
\[
\dot{x}_i(t) = \alpha_i(t) - x_i(t) - \sum_{j \neq i} \alpha_j(t)x_j(t),
\]

\[
x_i(0) = x_{i0},
\]

\[
\alpha_i(t) \in [0, 1].
\]

It is unrealistic to say that malicious hosts would reveal their states to the competitors as the game evolves, we say that the information structure of this dynamic game is open-loop. Open-loop information structure means that the players do not acquire further information except the common knowledge of the state vector at initial time $t = 0$ [13]. We follow the procedure in [21] to look for the open-loop Nash equilibrium.

**Lemma 2:** For the dynamic case of multiple competing malicious hosts, the best response of the malicious host $i = 1, ..., n$ is given by

\[
\alpha_i^*(t) = \begin{cases} \\
\frac{\lambda_i(t) - px_i(t) + 2r_i}{2(q_i + r_i)} & \lambda_i(t) > px_i(t) - 2r_i, \\
0 & \text{otherwise.} \end{cases}
\]

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

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

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 $i$,

\[
\alpha_i^*(t) = \arg\max \{H_i(\alpha_i(t), \alpha_{-i}(t), x_i(t), \lambda_i(t))\}.
\]

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. (31) by solving $\frac{\partial H_i}{\partial \alpha_i} = 0$.

**B. Dynamic Case**

**Theorem 7:** The best response of malicious host $i$ at the open-loop equilibrium is given by

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

**Proof:**

Differentiating (31) and let it be 0, we obtain

\[
\alpha_i^*(t) = \frac{p_i(1 - 2 \sum_{j \neq i} \alpha_j(t)x_j(t) - 2x_i(t))}{2(q_i + r_i)(1 + \sum_{j \neq i} \alpha_j(t))} + 2r_i(1 + \sum_{j \neq i} \alpha_j(t))
\]

Recall the stationary condition mentioned in (21), yielding

\[
x_i = \frac{\alpha_i}{(1 + \sum_{j \neq i} \alpha_j)},
\]

We then derive Eqn. (32) by plugging Eqn. (34) back to Eqn. (33).

**Remarks:** For $n$ non-cooperative malicious hosts, 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 $r$ respectively.

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

\[
r^*(t) = (p + q)(3 - 2\beta(t))^2 - \frac{1}{4}(3p + q),
\]
where $\beta(t)$ is the control of malicious behavior from two symmetric malicious hosts in steady status.

Proof: From Theorem 1, we know that

$$\beta^*(t) = \beta_1^*(t) = \beta_2^*(t) = \frac{1}{2}(3 - \sqrt{(p+q)(3p+q+4r)})$$

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

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

VI. NUMERICAL STUDY

In this section, we illustrate the results with numerical examples. We build our simulation on Matlab platform. The evolution of dynamic cases can be implemented with bvp4c toolbox.

A. Optimal Control for Single Malicious Host

We analyze three configurations for the single malicious host scenario in order to capture the influences of $r_i$ on the evolution of the state and the control. Each case contains both static and dynamic graphs to depict the evolution progress of the state $x_i(t)$ and the control $\alpha_i$. The left subgraph stands for the static case and the right one for the dynamic case. The system start with the initial state as $x_i(0) = 0$.

- Configuration 1: $q_i = r_i$, e.g., $p_i = 0.4$ and $q_i = r_i = 0.2$. With Fig. 1 we have that both the control and the state converge to 0.5.

- Configuration 2: $q_i > r_i$, e.g., $p_i = 0.4$, $q_i = 0.2$ and $r_i = 0.1$. The cost factor of positive activities is higher than the value of negative factor in this case. With Fig. 2 we have that both the control and the state converge to 0.42.

- Configuration 3: $q_i < r_i$, e.g., $p_i = 0.4$, $q_i = 0.2$ and $r_i = 0.3$; Contrast to the previous case, the cost factor of negative activities is higher than the one for the positive one in this case. With Fig. 3 both the control and the state converge to 0.56.

As we can see, the evolving trend exactly matches the analytical results in Sec. IV. By comparing these three cases, we can have that $\alpha_i(t)$ has the highest value when $r_i = 0.3$. In both the static case and the dynamic case, the converged values for $\alpha_i(t)$ and $x_i(t)$ coincide. Also, $\alpha_i(t)$ will finally converge to same value in both static case and dynamic case. That is also the same for $x_i(t)$. This finding matches the analysis as introduced in Theorem 1 and Theorem 4.

As we can see, in all cases the rate of behaving normally at the beginning is higher than the one at the convergence point. This implies the malicious host choose to suppress its aggressivity when it first enters the social network.

B. Multiple Competing Malicious Hosts

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

From Fig. 4 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 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. 5 $\alpha^*(t)$ begins at 0.5 and converges to 0.35 when $n$ increases, whereas $\beta^*$ starts from 0.5 and converges to 0.65. This observation means that the competition does not motivate good behaviors by hosts.

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. 6 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.
Fig. 3. The trajectory of the states and the controls for single malicious host when $p_i = 0.4, q_i = 0.2, r_i = 0.3$.

Fig. 4. The trajectory of the states for two identical players.

Fig. 5. The variation of $\alpha^*$ and $\beta^*$ when $n$ increases.

Fig. 6. The trajectory of the states and the controls for two players with different system maneuvers.
VII. CONCLUSION

This paper investigates the social-trust-based information dissemination by malicious hosts in social network sites. An interaction-based social trust model is presented. For studying the best response strategies of malicious hosts with a long-term objective, we formulate the maximization of malicious gains in a long time-span for cases of i) single independent malicious host, and ii) multiple competing malicious hosts, as an optimal control problem and a non-cooperative differential game, respectively. Through rigorous analysis, optimal response strategies for each malicious host is identified in both cases, based on which the best system maneuver mechanisms are presented so as to achieve the targeted level of overall malicious activities in the system. Extensive numerical studies further verifies the analytical results.

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 Proceedings of the 27th Annual Computer Security Applications Conference, pp. 93–102, ACM, 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 IEEE Symposium on Security and Privacy (SP), pp. 447–462, 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] S. Adali, R. Escriva, M. K. Goldberg, M. Hayyanovich, M. Magdon-Ismail, B. K. Szymanski, W. A. Wallace, and G. Williams, “Measuring behavioral trust in social networks,” in IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 150–152, 2010.

[6] S. Trifunovic, F. Legendre, and C. Anastasiades, “Social trust in opportunistic networks,” in INFOCOM IEEE Conference on Computer Communications Workshops, 2010, pp. 1–6, IEEE, 2010.

[7] G. Stringhini, C. Kruegel, and G. Vigna, “Detecting spammers on social networks,” in Proceedings of the 26th Annual Computer Security Applications Conference, pp. 1–9, ACM, 2010.

[8] 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.

[9] M. Sirivianos, K. Kim, and X. Yang, “Socialfilter: introducing social trust to collaborative spam mitigation,” in Proc. of IEEE INFOCOM, pp. 2300–2308, 2011.

[10] S. Hameed, X. Fu, P. Hui, and N. Sastry, “Lens: Leveraging social networking and trust to prevent spam transmission,” in Proc. of IEEE ICNP, pp. 13–18, 2011.

[11] M. Khouzani, S. Sarkar, and E. Altman, “Optimal control of epidemic evolution,” in Proc. of IEEE INFOCOM, pp. 1683–1691, 2011.

[12] A. Karnik and P. Dayama, “Optimal control of information epidemics,” in Proc. of IEEE COMSNETS, pp. 1–7, 2012.

[13] T. Alpcan and T. Basar, Network Security: A Decision and Game-Theoretic Approach. Cambridge University Press, 2010.

[14] P. Hu, H. Li, H. Fu, D. Cansever, and P. Mohapatra, “Dynamic defense strategy against advanced persistent threat with insiders,” in Computer Communications (INFOCOM), 2015 IEEE Conference on, pp. 747–755, IEEE, 2015.

[15] J. Omic, A. Orda, and P. Van Mieghem, “Protecting against network infections: A game theoretic perspective,” in Proc. of IEEE INFOCOM, pp. 1485–1493, 2009.

[16] Q. Zhu, L. Bushnell, and T. Basar, “Game-theoretic analysis of node capture and cloning attack with multiple attackers in wireless sensor networks,” in Proc. of IEEE CDC, pp. 3404–3411, Dec 2012.

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

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

[19] B. Jiang, N. Hegde, L. Massoulié, and D. Towsley, “How to optimally allocate your budget of attention in social networks,” in Proc. of IEEE INFOCOM, pp. 2373–2381, 2013.

[20] K. C. Border, “Fixed point theorems with applications to economics and game theory,” Cambridge Books, 1990.

[21] 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.