MILCOM 2016

SECURE COMMUNICATIONS AT THE SPEED OF CYBER

Trust Exploitation and Attention Competition: A Game Theoretical Model

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

University of California, Davis

BALTIMORE, MD • NOVEMBER 1–3, 2016

Background

- Prevailing usage of Social Network Sites (SNSs)
  - Facebook, Twitter, Google+…
- But also attract malicious activities …
  - spam, click fraud, identity theft and phishing

Social Trust

- The confidence that a node will behave in an expected way in a social network
- Determined by its frequency of non-malicious interactions with other nodes
  - a positive interaction, e.g., posing a trustworthy news, will improve the social trust of the node leading to larger influence for information dissemination
  - a negative interaction, maliciously spreading a rumor, will result in a degradation in the trust and hurting its potential of information dissemination in the future

Social Trust

- The confidence that a node will behave in an expected way in a social network
- Determined by its frequency of non-malicious interactions with other nodes
  - a positive interaction, e.g., posing a trustworthy news, will improve the social trust of the node leading to larger influence for information dissemination
  - a negative interaction, maliciously spreading a rumor, will result in a degradation in the trust and hurting his/her potential of information dissemination in the future

Social Trust

- Malicious node aims to maximize its overall personal benefits over a long time span
- A tradeoff 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
- To understand the malicious host's best action strategy towards this tradeoff, and to accordingly propose optimal system maneuver mechanism to confine the malicious activities in the social network

Social Trust

- $X_i(t)$ is a random variable denotes the number of nodes that trust node *i* at time *t*. *N* denotes the total number of users in the social network.
- Social trust

$$x_i(t) = E\left(\frac{X_i(t)}{N}\right)$$

It evolves overtime, and its dynamics is determined by its initial value $x_{i0} \in [0, 1]$, and its actions on disseminating trustable/malicious information

Dynamics of Social Trust: Single Malicious Node

- Malicious nodes mix good content with bad to persistently make profits over a long time-span
- Malicious node *i* posts some content *c(t)* at time *t*. The impact of the content on the dynamics of social trust of *i* is 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.
  - Similarly, let $p_2(c(t), \delta)$ denote the probability that a node trusting *i* at time *t* becomes distrusting *i* at time $t + \delta$.

Dynamics of Social Trust: Single Malicious Node

Assume the impact is independent across nodes, we have,

$$E\left(X_i(t+\delta) - X_i\ \ (t)\middle|X_i\ \ (t)\right) = \ p_1(c(t),\delta)\left(N - X_i\ \ (t)\right) - p_2(c(t),\delta)X_i(t))$$

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

$$x_i(t+\delta) - x_i\ \ (t) = \ p_1(c(t),\delta)\left(N - x_i\ \ (t)\right) - p_2(c(t),\delta)x_i(t))$$

Dividing both sides by δ and let$\delta \to 0$, we obtain the following dynamics of social trust of *I*, $\dot{x}_i(t) = \frac{dx_i(t)}{dt} = \alpha_i(t)(1 - x_i(t)) - \beta_i(t)x_i(t)$

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

$$\text{where } \alpha_i(t) = \lim_{\delta\to 0}\frac{p_1(c(t),\delta)}{\delta} \text{ and } \beta_i(t) = \lim_{\delta\to 0}\frac{p_2(c(t),\delta)}{\delta}$$

α and β are considered as the strategy of node *i.* α can be viewed as the social trust gained by posting trustable information that has positive response from $1\ -\ x_i\ \ (t)$; β reflects the loss of social trust because of disseminating malicious content to $x_i\ \ (t)$. We normalize them so that $\alpha +\ \beta = 1$.

Multiple Malicious Nodes

- **Budget of attention**: 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.
  - a user cannot click “like” for two separate posts concurrently at exactly the same time.
  - malicious nodes have to compete with each other to gain social trust from their potential victims to maximize their individual profits
- The dynamics of node *i*'s social trust should consider the joint actions of all the malicious nodes,

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

$\Sigma_{j \in -i} \alpha_j(t) x_i$(t) denotes the accumulated loss rate of social trust, that is obtained by other malicious nodes

Payoff and Cost Functions for Malicious Nodes

- The instantaneous malicious profit of node *i* at time *t* is proportional to its malicious activity rate $\beta_i(t)$ and its social trust $x_i\ (t)$. The long-term profit gain P_i is

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

- We utilize the commonly applied quadratic cost function to capture the instantaneous operational costs. The long-term costs for positive activities, $C_{i1}\ (t)$ and negative activities, $C_{i2}\ (t)$ 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 *p* is the unit gain, *q* and *r* are the unit cost.

- To sum up, the net profit for malicious node *i* is $P_i -\ C_{i1}$ - C_{i2}

System Maneuver

- Our objective is finding 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 in β
  - As for the system administrator, it can adjust the value of *r*, which is the unit penalty for malicious activities , at the start of the system so as to achieve its targeted level of overall malicious activity.

Social Trust Games

- 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 that is continuously played among nodes.
- For each malicious node *i* ∈ {1, 2, ..., *n*}, it solves a profit-maximization problem in the game as follows

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

- Our objective is to derive the open-loop Nash Equilibrium (NE)

Open-loop Nash Equilibrium

- The strategy profile $\Phi(t) = \{\alpha_i(t), \alpha_{-i}(t); \beta_i(t), \beta_{-i}(t)\}$ forms a open-loop Nash Equilibrium iff all following inequalities are satisfied

$$\begin{gathered} J_1(\phi_1^*(t), ..., \phi_n^*(t)) \geq J_1(\phi_1(t), ..., \phi_n^*(t)), \\ \vdots \\ J_n(\phi_i^*(t), ..., \phi_n^*(t)) \geq J_n(\phi_1^*(t), ..., \phi_n^*(t)). \end{gathered}$$

Steady

"a stable state of a system involving the interaction of different participants, in which no participant can gain by a unilateral change of strategy if the strategies of the others remain unchanged."

Static Case

- The activity variables of all malicious nodes, i.e., α and β 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.
- We prove that there exists a Nash equilibrium for the static social trust game.
- The best response for the malicious node *i* is given by

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

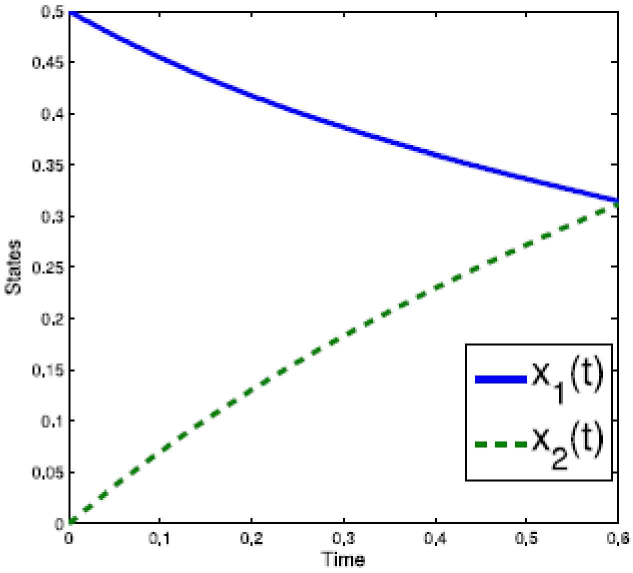
Dynamic Case

- We derive the open-loop NE and show that the optimal dynamic control coincides with the static solution for the single malicious node setting.
- For the situation of two symmetric players, the optimal system maneuver is given by

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

Numerical Study

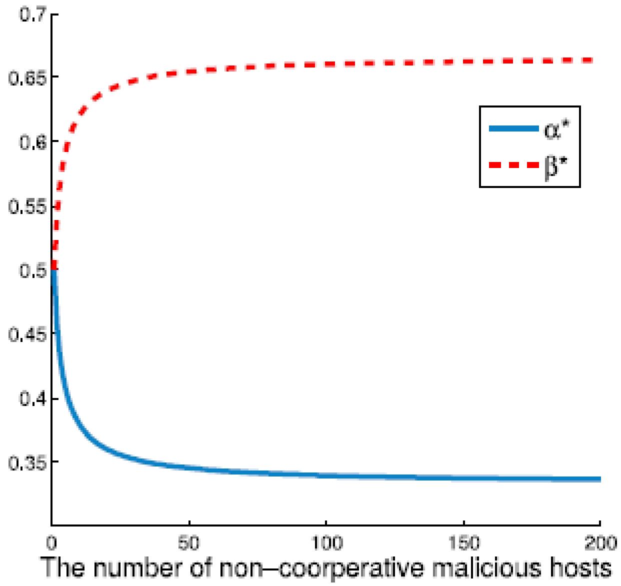
- Suppose there is an existing malicious node that has already reached its steady state.
- Now we introduce another homogenous malicious node with identical configurations with the existing node.
- The player I deviates from its previously steady state and its *x* begins decreasing, meanwhile, x_2 of player II starts from 0 and increases until finally converging to the steady position, which matches the analytical result for steady position.



(a) state trajectory for two identical players

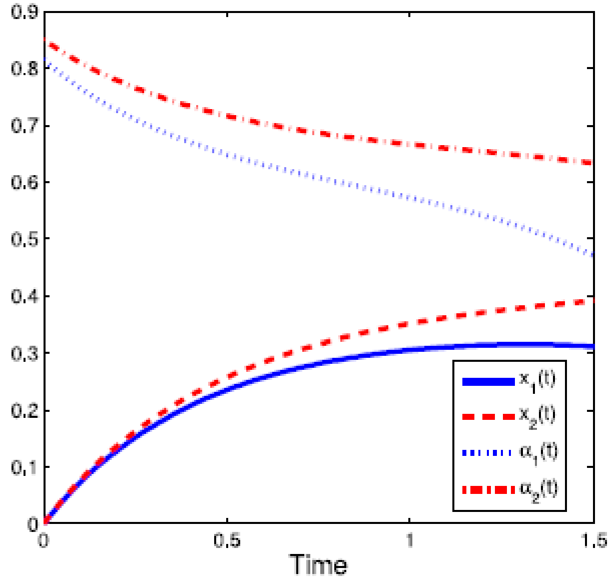
Numerical Study

- We examine how the control in the steady state evolves when the amount of players increases.
- As shown in the figure, α 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.



Numerical Study

- The influences of the system maneuver *r* on the controls and the states of a two-player game scenario.
- The figure depicts the evolution progress of the states and the controls of two players with *r* = 0.2 and *r* = 0.3.
- We can see that the higher system maneuver comes with the lower negative activity rate in social trust games.



(c) trajectory of the states and the controls for two players with different system maneuvers

Summary

- 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.
- Extensive numerical studies further verify our analytical results.

Thank you!