Cooperation among Malicious Agents: A General Quantitative Congestion Game Framework (Extended Abstract)

Zaojie Rui
School of Computer Science and Engineering, Southeast University, Nanjing 211189, China;
Research Center for Learning Science, Southeast University, Nanjing 210096, China
culturejie@gmail.com

Tuanjie Fu, Darong Lai
School of Computer Science and Engineering, Southeast University, Nanjing 211189, China
tuanjiefu@gmail.com
daronglai@seu.edu.cn

Yichuan Jiang*
School of Computer Science and Engineering, Southeast University, Nanjing 211189, China;
State Key Laboratory for Novel Software Technology, Nanjing University, Nanjing 210093, China
yjiang@seu.edu.cn

ABSTRACT

Malicious behaviors and cooperation have been well studied separately. However, rare systematic study has been conducted on the combination of them: malicious cooperation. In this paper, a general quantitative utility function of malicious cooperation is firstly formulated in a congestion game framework. Both objective and subjective factors are incorporated (e.g., malicious social networks and moral degrees). Then, Nash equilibrium and the condition of malicious cooperation are given theoretically. Meanwhile, we show empirically that malicious cooperation may even improve system performance (i.e., *catfish effect*).

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence−*Multiagent Systems*

General Terms

Theory, Performance, Experimentation

Keywords

Malicious Cooperation, Malicious Cooperative Agents, Catfish Effect, Congestion Game

1. MOTIVATION

Malicious agents are traditionally regarded as non cooperative. However, they may cooperate with each other to conduct malicious behaviors. For instance, coordinated attacks (e.g. DDoS) in cyberspace are launched via cooperation among hackers. We call this type of cooperation malicious cooperation and these agents malicious cooperative agents (MCA). Although malicious behaviors and cooperation have been well studied in agent society, rare attention is paid to malicious cooperation. Thus, we look into the MCA and the malicious cooperation. We mainly focus on two problems: 1) the attraction of malicious cooperation (utility of the MCA from malicious cooperation) and 2) the effect of malicious cooperation on the system.

Appears in: *Proceedings of the 11th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2012)*, Conitzer, Winikoff, Padgham, and van der Hoek (eds.), 4–8 June 2012, Valencia, Spain.

* Corresponding author.

2. MODEL DESCRIPTION

For the sake of generality, we present our model in a modified congestion game framework [1], where malicious cooperation is explicitly incorporated. There are two types of agents in the game, *MCA*, and regular agents (*RA*). We assume that all agents are informed of the structure of road network and cost functions. Also Agents are self-interested and are aimed at minimizing own time delay of the journey. The behaviors of agents are based on that of gossip agents in *Gossip Networks* [2]. Agents will communicate with other agents to exchange load information of roads in order to make a more reliable route. Furthermore, we made a modification to MCA to reflect malicious cooperation. Besides tampering the road congestion information in their own routes to be high congested to cheat others out of their path as in [2], MCA in our model will also cheat others for his cooperators.

3. UTILITY FUNCTION OF MCA

Here, a general utility function of the MCA is proposed in a congestion game framework. In order to imply the social factors, which are often ignored in agent research, the function is based on that of in social study [3]. Also, we add the formulation of malicious cooperation to overcome its absence in [3].

***Network and Malicious Effort*:** The adjacency matrix G of network **g** denotes the direct connections of the social network among the MCA. Meanwhile, malicious cooperation is among their social connections. We denote cooperation network as P. Then malicious effort (e_i) of a MCA i is defined as the percentage of malicious peers in its social networks, i.e., $e_i = \sum_{j=1}^{n} p_{ij} / \sum_{j=1}^{n} g_{ij}$.

***Benefit*:** The benefits of agent i come from both his own behaviors and his cooperators' malicious behaviors for him. Thus $B_i = a + e_i \sum_{j=1}^{n} g_{ji} b_j$, where $a>0$ is a constant, b_j denotes the ability of agent j to make benefits from its malicious behaviors. Here b_j in a congestion game roots in cheating RA, so $b_j = b\, \gamma\, |RA_j|$, where b denotes the benefit per successful cheating, γ denotes the success possibility of cheating, and $|RA_j|$ denotes the number of encountered RA in the agent's journey.

***Cost*:** MCA's cost includes *i*) the cost of being detected (c_i^{de}) (including the cost of his malicious behaviors for both himself and his cooperators), *ii*) ignoring high congested information from RA in his or his cooperators' routes (c_i^{ig}), *iii*) *moral cost* for malicious behaviors and *conformity cost* of failing to conform to friends as indicated in [3]. So the sum cost of agent i is $C_i = \sum_{j=1}^{n} p_{ij} f |RA_i| + ce_i^2 + d(e_i - \overline{e_i})^2$, where α, β denotes the possibility of being detected and the false information turning to be true, respectively; c, d denotes *moral degree* and *conformity degree*, respectively; $\overline{e_i}$ denotes average malicious efforts of friends, and $f = c^{de}\alpha + c^{ig}\beta$.

Utility Function: Finally, utility function of agent i is:

$$U_i = a + e_i \sum_{j=1}^{n} g_{ij}(b_j - f|RA_i|) - ce_i^2 - d(e_i - \overline{e_i})^2 . \quad (1)$$

4. PROPERTIES OF MCA

Proposition 1 (Nash Equilibrium): Consider the general case when 1) all MCA have different ability to make benefits, 2) social network among MCA are heterogeneous, 3) agents' *conformity degrees* are different. Assume $b_j > f\,|RA_i|$. Then a unique Nash equilibrium in pure strategies of the game is:

$$e_i^* = d\overline{e_i} / c + d + \sum_{j=1}^{n} g_{ij}(b_j - f\,|\,RA_i\,|) / 2(c+d). \quad (2)$$

Proof 1. The utility function is nearly the same as the one in [3], with (b_i-pf) replaced by $\sum_{j=1}^{n} g_{ij}(b_j - f\,|\,RA_i\,|)$. Then we can apply $\alpha = \sum_{j=1}^{n} g_{ij}(b_j - f\,|\,RA_i\,|)$ into the proof of *Proposition 2* in [3]. The assumption of $b_i - pf > 0$ in our case is $\sum_{j=1}^{n} g_{ij}(b_j - f\,|\,RA_i\,|) > 0$. It is always satisfied since $b_j > f\,|RA_i|$. □

Proposition 2 (Condition of Malicious Cooperation): Assume $|RA_i| \sim N(u,\sigma^2)$, agents interact in equal possibility and c=d=0, then the condition that MCA trend to contribute more malicious efforts is: $b / f > u$.

Proof 2. Under the conditions, the utility function turns to be $U_i = a + e_i \sum_{j=1}^{n} g_{ij}(b\gamma\,|\,RA_j\,| - f|RA_i|)$. Thus $E(U_i) = a + e_i u \sum_{j=1}^{n} g_{ij}(b\gamma - f)$. Then we get $\partial E(U_i) / \partial e_i = u \sum_{j=1}^{n} g_{ij}(b\gamma - f)$. If $b\gamma - f > 0$, then MCA will trend to make more benefits if they contribute more malicious efforts. As agents interact in equal possibility, $\gamma = 1/u$. Thus $b\,\gamma - f > 0$ turns to be $b\,/f > u$. □

Note that ***Proposition 2*** is consistent with the conclusion in a recent *Nature* letter [4], which demonstrates an extraordinary simple rule that cooperators are advantageous over defectors if the benefit of the cooperative act (b), divided by the cost (c), is larger than the average number of neighbors (k), i.e., $b/c > k$. Their c and k are f and u in our model, respectively.

5. EXPERIMENTAL RESULTS

Based on [2], the network simulates a city center which consists of 162 edges and 100 junctions. Each simulation consists of ten iterations. Two hundred agents are randomly generated and they will finish at least 20 journeys from their sources to targets in the iteration. And each journey is tagged with different source-target pair. MCA are initialed with a social network. And each MCA is initialized with a malicious effort ([0, 1]). Similar to [2], the attraction of malicious cooperation in simulations is evaluated when the detection is absent. And the utility of agents in simulations are denoted by time delay per kilometer (*TDK*). A smaller *TDK* indicates a higher utility. Noting that moral cost and peers effect have been validated in [3], we just focus on second term of Equation 1. In below, n denotes the number of MCA.

Validation of Utility Function: Two set of simulations are done to test the effect of characteristics of social networks and malicious efforts.

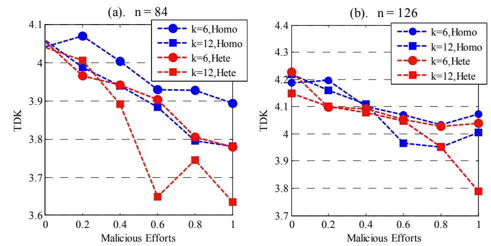

Figure 1. Effect of Social Networks and Malicious Efforts

Figure 1 illustrates that utility increases as MCA contributes more efforts (e_i), which can be explained by the increase of second term of Equation 1. Also, compared to homogeneous social networks, heterogeneous networks can improve the utility of MCA.

Catfish Effect: Here, we will show the effect of malicious cooperation on RA and the system. We find a counterintuitive phenomenon, denominated as "*catfish effect*": malicious cooperation may benefit the system and even RA, comparing to the situation where malicious cooperation is absent (i.e., $e_i = 0$).

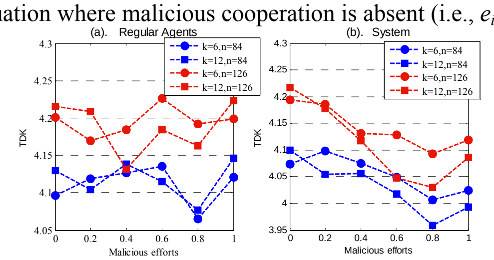

Figure 2 Effect of Malicious Cooperation on RA and System

Figure 2 illustrates that there exists an appropriate malicious effort, i.e., *catfish effort*, where *TDK* is minimal for RA or system. It implies that a catfish effect exists if e_i is appropriate. A potential reason for this is that false information helps RA avoid encountering MCA, which may avoid high congested roads to some extent.

6. CONCLUSION AND FUTURE WORK

We have presented a general quantitative congestion game framework for analyzing cooperation among malicious agents. We show theoretically the Nash equilibrium of malicious efforts and condition of malicious cooperation. Also, we show the catfish effect via simulations. Our future work may study the malicious cooperation in other evaluation scenarios. Furthermore, it is interesting to find a way to utilize catfish effect to benefit system.

Acknowledgements

This work is supported by projects NSFC (61170164, 60803060), NCET (NCET-09-0289) and SRFDP (20090092110048).

7. REFERENCES

[1] I. Ashlagi, D. Monderer, and M. Tennenholtz. Two-terminal routing games with unknown active players. *Artificial Intelligence*, 173(15): 1441-1455, 2009.

[2] R. Lin, S. Kraus, and Y. Shavitt. On the benefits of cheating by self-interested agents in vehicular networks. *In AAMAS*, 2007.

[3] E. Patacchini, and Y. Zenou. Juvenile delinquency and conformism. *Journal of Law, Economics, and Organization*, 2009.

[4] H. Ohtsuki, et al. A simple rule for the evolution of cooperation on graphs and social networks. *Nature,* 441(7092): 502-505, 2006.