A Game-Theoretic Approach for Threats Detection and Intervention in Surveillance

(Extended Abstract)

Wenjun Ma[1], Weiru Liu[1], Paul Miller[1], and Xudong Luo[2]
[1]School of Electronics, Electrical Engineering and Computer Science,
Queen's University Belfast, Belfast, BT7 1NN, UK
{w.ma, w.liu, p.miller}@qub.ac.uk
[2] Institute of Logic and Cognition, Sun Yat-sen University, Guangzhou, 510275, China
luoxd3@mail.sysu.edu.cn

ABSTRACT

Threat intervention with limited security resources is a challenging problem. An optimal strategy is to effectively predict attackers' targets (or goals) based on current available information, and use such predictions to disrupt their planned attacks. In this paper, we propose a game-theoretic framework to address this challenge which encompasses the following three elements. First, we design a method to analyze an attacker's types in order to determine the most plausible type of an attacker. Second, we propose an approach to predict possible targets of an attack and the course of actions that the attackers may take even when the attackers' types are ambiguous. Third, a game-theoretic based strategy is developed to determine the best intervention approaches taken by defenders (security resources).

Categories and Subject Descriptors

I.2.11 [**Artificial Intelligence**]: Distributed Artificial Intelligence—*Multi-agent Systems*

Keywords

Game theory, Surveillance System, Event Modeling

1. INTRODUCTION

Predicting attackers' intentions and allocating security resources are two research problems that are seldom addressed together in intelligence surveillance. In this paper, based on the Dempster-Shafer theory of evidence (D-S theory) [5] and the DS/AHP method [1], we first analyze information, using the multi-criteria event modeling framework [4], to determine the most plausible attacker type. Second, we propose *a principle of acceptable costs of minimax regret* to predict the most plausible strategy that each type of attacker may adopt, using a game matrix constructed from ambiguous information that is available. Following this, a method to allow a defender to select an optimal strategy is developed that will minimize the defender's maximum regret at an acceptable cost.

Our contributions are as follows: (i) a method to analyze the type of attacker with ambiguous information; (ii) a principle of acceptable costs of minimax regret to predict an attacker's target/goal based on information gathered and reasoned upon by an intelligent surveillance system; and (iii) a method to determine a defender's optimal strategy based on the principle of acceptable costs of minimax regret.

Appears in: *Alessio Lomuscio, Paul Scerri, Ana Bazzan, and Michael Huhns (eds.), Proceedings of the 13th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2014), May 5-9, 2014, Paris, France.*

2. ANALYZING ATTACKER TYPES

In order to find out the optimal threat intervention strategy, first we should determine the most plausible attacker type based on the information obtained by the multi-criteria event modeling framework [4] with a mass value m_i^c, criterion's weight ω_c and a utility function U^c for each criterion. Since the problem of analyzing an attacker's type can be considered as an ambiguous Multi-Criteria Decision Making problem as illustrated by Figure 1[1], we deploy the extended DS/AHP method to determine an attacker's type with the following steps.

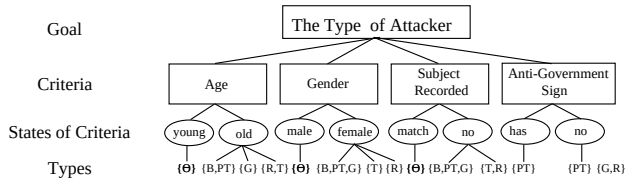

Figure 1: A hierarchical structure illustrating relationship between criteria and attacker's types

Step 1: Determine types in frame Θ*, select criteria w.r.t the observed events, and construct the preference matrix with* m_i^c*,* U^c *and weight.*

Step 2: Determine the mass function $m_{v_c^i,t}$ *about an attacker's type for each state of each criterion as follow:*

$$m_{v_c^i,t}(s_i)=\frac{a_i\omega_c}{\sum_{j=1}^{d} a_j\omega_c+\sqrt{d}},\ m_{v_c^i,t}(\Theta)=\frac{\sqrt{d}}{\sum_{j=1}^{d} a_j\omega_c+\sqrt{d}},$$

where s_i is one of d focal elements of a criterion's state that has the preference value a_i, ω is that criterion's weight.

Step 3: Obtain the mass functions set $\{m_{c,t}(x)\}$ *about an attacker's types for each criterion by Eq.(1). And then*

[1]where the types of attacker are Gunman (G), Bomber (B), Robber (R), Political Terrorist (PT), and Thief (T)

obtain the overall mass function of attacker's types by combining all elements in $\{m_{c,t}(x)\}$ with Dempster's combine rule [5].

$$m_{c,t}(x)=\sum_{v_c^i\in\Omega_c V}\sum_{\bigcap v_c^i\neq\emptyset}\frac{u(v_c^i)}{\sum\{u(v_c^j)\mid v_c^j\in V\}}m_c(V)m_{v_c^i,t}(x). \quad (1)$$

where m_c is a mass function over frame Ω_c, $m_{v_c^i,t}$ is a mass functions over Θ about an attacker's types for the states v_c^i of the criterion c, $u(v_c^i)$ is the utility value for the state v_c^i, and V is any subset of Ω_c.

3. THREAT INTERVENTION GAME

Since Different types of attackers might have different preferences over the choices of their next move, we can construct a game-theoretic model for the ambiguous threat intervention problem. Moreover, since such a game model is different from the static Bayesian game, or the Stackelberg's game, in which players' types are determined by a probability distribution, while in our problem, due to ambiguous information, players' types are determined by the mass function. We propose a new solution concept for the threat intervention game.

First, we will predict attacker's strategies by *the principle of acceptable costs of minimax regret*, which suggests that the decision maker will consider not only the maximin regret but also the minimum utility he can obtain in decision making. Formally, we have:

DEFINITION 1. *Let $S_2=\{s_2^1,\ldots,s_2^m\}$ be a set of an attacker's mixed strategies, $\sigma_{2,t}\in[0,1]$ be the threshold of acceptable costs that an attacker of type t can bear and $a_t\in A_1$ is the pure strategy of a defender. Then the optimal strategy for attacker type t, denoted as $s_{2,t}^*$ ($s_{2,t}^*\in S_2$), is given by:*

$$s_{2,t}^*=\arg\min\{\overline{r}(s_2^i)\,|\,\overline{r}(s_2^i)=\max_{a_h}\{\max_{j\neq i}u_{2,t}(a_h,s_2^j)-u_{2,t}(a_h,s_2^i)\}\}, \quad (2)$$

where

$$\min_{a_s}u_{2,t}(a_s,s_2^i)\geq\max_{s_2^k}\min_{a_r}u_{2,t}(a_r,s_2^k)-\sigma_{2,t}\varsigma_{a,t}, \quad (3)$$

$$\varsigma_{a,t}=\max_{s_2^k}\min_{a_r}u_{2,t}(a_r,s_2^k)-\min_{s_2^l}\min_{a_w}u_{2,t}(a_w,s_2^l). \quad (4)$$

In the above definition, the higher $\sigma_{2,t}$ is, the higher potential loss for the minimum utility a type of attacker can accept. Moreover, $\varsigma_{a,t}$ means the maximum costs that a type t attacker might pay in a threat intervention game. Thus, $\sigma_{2,t}\varsigma_{a,t}$ means the highest costs that a type t attacker is willing to pay given his type. Finally, in real-world applications, $\sigma_{2,t}$ for each type of attacker can be obtained by historical data and the judgement of criminology expects.

After that, the security team's optimal strategy for threat intervention can be obtained as follow:

DEFINITION 2. *Let $S_1=\{s_1^1,\ldots,s_1^n\}$ be a set of defender's mixed strategies, Θ be the set of types of an attacker, $\sigma_1\in[0,1]$ be the threshold of acceptable costs that a defender can bear, $EUI(s_1^i)=[\underline{E}(s_1^i),\overline{E}(s_1^i)]$ be an expected utility interval [6], $\delta(s_1^i)$ be the normalized nonspecificity degree [3], and $b_{l,t}^*\in A_{2,t}^*\subseteq A_2$ be a pure strategy for which the optimal mixed strategy $s_{2,t}^*$ assigns a positive probability. Then a defender's optimal strategy, denoted as s_1^*, is given by:*

$$s_1^*=\arg\min\{\overline{r}(s_1^i)\mid\overline{r}(s_1^i)=\max_{j\neq i}\varepsilon(s_1^j)-\underline{E}(s_1^i)\}, \quad (5)$$

where

$$\min_{s\in\Theta}\min_{b_{2,t}^*}u_{1,t}(s_1,b_{2,t}^*)\geq\max_{s_1^k}\min_{r\in\Theta}\min_{b_{2,r}^*}u_{1,t}(s_1^k,b_{2,r}^*)-\sigma_1\varsigma_d, \quad (6)$$

$$\varsigma_d=\max_{s_1^k}\min_{r\in\Theta}\min_{b_{2,r}^*}u_{1,t}(s_1^k,b_{2,r}^*)-\min_{s_1^l}\min_{u\in\Theta}\min_{b_{2,u}^*}u_{1,t}(s_1^l,b_{2,u}^*), \quad (7)$$

$$\varepsilon(s_1^j)=\underline{E}(s_1^j)+(1-\delta(s_1^j))(\overline{E}(s_1^j)-\underline{E}(s_1^j)). \quad (8)$$

Actually, Eq. (5) means that given the expected utility intervals of all defender's strategies, he will elicit the maximum regret of a given mixed strategy by a (counterfactual) comparison between the lower expected utility of a reality choice, and the maximum upper expected utility of a foregone rejected alternative that might have been. And because only one pure strategy of an attacker's optimal mixed strategy will actually be taken, Eqs. (6) and (7) mean a defender will consider the potential reduction of maximum minimum utility given such a pure strategy of an attacker. Hence, $\varepsilon(s_1^j)$ is an ambiguity aversion upper expected utility for a defender. From Eq. (8), nonspecificity degree $\delta(s_1^j)$ actually works as a discount factor: the higher the degree, the more the upper utility of a choice is discounted. In fact, Eq. (8) is based on the consideration of ambiguity aversion that describes an attitude of preference for known risks over unknown risks, when the decision maker faces an ambiguous decision problem [2].

In fact, by Definition 2, a security manager can tune the value of σ_1 to reflect different (real-time) situations at different security area. Thus, our method is more flexible in balancing returns and risks, where returns is interpreted as the expected payoff of successfully preventing an attack, while risks mean the possibility of unaffordable losses and the severity of loss that are caused by the failure of intervention.

4. CONCLUSION

This paper addresses the threat detection and intervention problem in intelligence surveillance. First, we introduced an attacker's type analysis method according to information obtained by a surveillance system. Then, we developed a game-theoretic model for the ambiguous threat intervention problem and proposed a principle of acceptable costs of minimax regret to predict the strategy of an attacker and accordingly determine the optimal strategy for a defender. Based on our method, we can address both the problem of predicting attackers' intentions and the problem of allocating security resources in intelligence surveillance.

5. ACKNOWLEDGEMENTS

This work has been supported by the EPSRC projects EP/G034303/1; EP/H049606/1; Bairen plan of Sun Yat-sen University; National Natural Science Foundation of China (No. 61173019).

6. REFERENCES

[1] M. Beynon. DS/AHP method: A mathematical analysis, including an understanding of uncertainty. *Eur. J. of Op. Research.*, 140(1):148–164, 2002.

[2] E. Daniel. Risk, ambiguous, and the savage axioms. *Quarterly Journal of Economics*, 75(4):643–669, 1961.

[3] D. Dubois and H. Prade. A note on measures of specificity for fuzzy sets. *Int. J. General Systems*, 10(4):279–283, 1985.

[4] J. Ma, W. Liu, and P. Miller. Event modelling and reasoning with uncertain information for distributed sensor networks. In *Procs. of SUM'10*, pages 236–249.

[5] G. Shafer. *A Mathematical Theory of Evidence.* Princeton University Press, 1976.

[6] T. M. Strat. Decision analysis using belief functions. *Int. J. of Approx. Rea.*, 4(5-6):391–417, 1990.