Defense Against Advanced Persistent Threats: A Colonel Blotto Game Approach

Minghui Min*, Liang Xiao*, Caixia Xie*, Mohammad Hajimirsadeghi†, Narayan B. Mandayam†
*Dept. of Communication Engineering, Xiamen University, Xiamen, China. Email: lxiao@xmu.edu.cn
†WINLAB, Dept. of ECE, Rutgers University, Piscataway, NJ. Email: {mohammad, narayan}@winlab.rutgers.edu

***Abstract*—An Advanced Persistent Threat (APT) attacker applies multiple sophisticated methods to continuously and stealthily attack targeted cyber systems. In this paper, the interactions between an APT attacker and a cloud system defender in their allocation of the Central Processing Units (CPUs) over multiple devices are formulated as a Colonel Blotto game (CBG), which models the competition of two players under given resource constraints over multiple battlefields. The Nash equilibria (NEs) of the CBG-based APT defense game are derived for the case with symmetric players and the case with asymmetric players each with different total number of CPUs. The expected data protection level and the utility of the defender are provided for each game at the NE. An APT defense strategy based on the policy hill-climbing (PHC) algorithm is proposed for the defender to achieve the optimal CPU allocation distribution over the devices in the dynamic defense game without being aware of the APT attack model. Simulation results have verified the efficacy of our proposed algorithm, showing that both the data protection level and the utility of the defender are improved compared with the benchmark greedy allocation algorithm.**

***Index Terms*—Colonel Blotto game, advanced persistent threats, cloud system, policy hill-climbing, data protection level**

I. Introduction

Cloud systems are vulnerable to Advanced Persistent Threats (APTs), in which an attacker applies multiple sophisticated methods such as the injection of multiple malwares to continuously and stealthily steal data from the target cyber system [1]. Being difficult to be detected and completely blocked, APT attacks cause serious data privacy leakage and millions of dollars' loss [2], [3]. The seminal work in [4] formulates a stealthy takeover game between an APT attacker and a defender, who compete to control the target cyber system. The cloud APT game among an APT attacker, a cloud defender and a mobile device was formulated in [5] which combines the APT defense game in [4] with the signaling game between the cloud and the mobile device. The APT defense game proposed in [6] extends the Flipit game in [4] to multi-node systems with limited resources, and derives the Nash Equilibrium (NE) for the case with deterministic attack time and fully observable defense. The joint attack of the APTs and the insider attacks was addressed in [7], which analyzes both the APT game and the information-trading game. APT attackers with a subjective view to decide the attack interval were analyzed in [8], [9] based on Prospect theory [10].

We investigate the defense strategies against the APT attacker with limited computation resources, such as Central Processing Units (CPUs) over multiple devices in the target cloud system. Both defender and attacker must deal with the challenging task of strategically allocating their constrained resources. The competitive nature of this process is closely related to the Colonel Blotto game (CBG) [11], [12], which is a two player zero-sum game to model the competition of two colonels with resource constraints over multiple battlefields. The player who allocates most resources to a battlefield wins the game, and the overall payoff of a player is proportional to the number of the won battlefields. The Nash equilibria of the continuous Colonel Blotto game derived in [11] describe the force distribution under different resource constraints across symmetric battlefields with the same value, or under the same resource constraint across asymmetric battlefields in [12]. The CBG with discrete resources studied in [13] evaluates the case with symmetric resources and symmetric battlefields.

Colonel Blotto game with mixed strategy has been used in [14] to study the spectrum allocation of network service providers, and a fictitious play based allocation approach has been developed to compute the equilibrium of the game with discrete spectrum resources. The CBG-based jamming resistance game for Internet of Things (IoT) formulated in [15] shows that neither the defender nor the attacker dominates if the computational resources are limited. The CBG-based jamming defense game in [16] has studied jamming attacks against IoT with continuous and asymmetric radio power resources.

In this paper, a Colonel Blotto game is used to formulate the defense of a cloud defender against an APT attacker, by choosing the allocation of a limited number of CPUs to defend/attack each device. As the number of CPUs is an integer by definition, the APT defense game is a Colonel Blotto game with discrete resources. On the other hand, a defender can apply time sharing (division) to scan multiple cloud devices with a single CPU, which approximately yields a continuous CBG. The pure strategy CBG game in which a player allocates a predetermined amount of resources to each storage device achieves an NE only under rare circumstance. This turns our attention to the CBG with mixed strategies, in which the attacker and the defender choose the distribution of each CPU allocation strategy over devices and introduce randomness in their actions to fool the opponent.

The NEs of the static defense game with mixed-strategy are derived to disclose the impacts of the data storage size,

This research was supported in part by National Natural Science Foundation of China under Grant 61671396 and CCF-Venustech Hongyan Research Initiative (2016-010), and in part by the U.S. National Science Foundation (NSF) under Grant ACI-1541069.

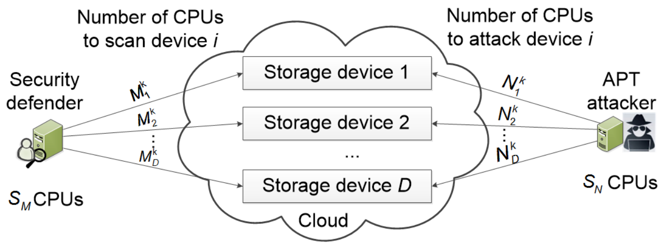


Fig. 1. APT defense game, in which a defender with S_M CPUs totally protects the D cloud devices against an APT attacker with S_N CPUs.

the total number of CPUs of the cloud defender and that of the attacker on the security performance. As a model-free reinforcement learning technique, Q-learning can derive the optimal strategy in a Markov decision process of the dynamic game with pure strategy. The policy hill-climbing (PHC) algorithm, as an extension of Q-learning for the mixed-strategy game can achieve the optimal mixed strategy by incrementally changing the policy [17]. Therefore, we propose a PHC-based APT defense algorithm for the defender to choose the probability distribution of each CPU allocation scheme to scan the cloud devices without being aware of the APT attack model in the dynamic APT defense game. Simulations are performed to evaluate the resulting data protection level defined as the normalized size of the "safe" data which are successfully protected by the defender and the utility of the defender and compare them with the benchmark schemes.

The rest of this paper is organized as follows: We present the system model in section II and study the CBG-based APT defense game in Section III. We develop a PHC-based APT defense algorithm for the dynamic game in Section IV. We present simulation results in Section V and draw conclusions in Section VI.

II. System Model

As illustrated in Fig. 1, we consider a cloud storage system consisting of D storage devices, where device i stores data of size $B_i \geq 0$, with $1 \leq i \leq D$. Let $\mathbf{B} = [B_i]_{1 \leq i \leq D}$, and $\widehat{B} = \sum_{i=1}^{D} B_i$ denote the total amount of data in a system. An APT attacker aims to jeopardize the D devices while remaining undetected by applying N_i^k CPUs to attack device i at time k to steal the stored data, following the constraint on the total number of CPUs, i.e., $\sum_{i=1}^{D} N_i^k \leq S_N$, where S_N is the total number of the CPUs that the attacker has at its disposal. For simplicity, we set $\mathbf{N}^k = \left[N_i^k\right]_{1 \leq i \leq D}$. The time index k in the superscript can be omitted if no confusion occurs.

The defender with S_M CPUs allocates M_i^k CPUs to scan device i to detect APTs following the CPU budget, i.e., $\sum_{i=1}^{D} M_i^k \leq S_M$, with the defense CPU allocation $\mathbf{M}^k = \left[M_i^k\right]_{1 \leq i \leq D}$. The data stored on device i are assumed to be safe if the defender applies more CPUs to scan the device than the attacker, with $M_i^k > N_i^k$, and vice versa. If $M_i^k = N_i^k$, both players have an equal opportunity to control the device. Therefore, the data protection level denoted by R is defined as the normalized size of the "safe" data protected by the defender and is given by

$$R = \frac{1}{\widehat{B}} \sum_{i=1}^{D} B_i sgn\left(M_i^k - N_i^k\right), \tag{1}$$

where

$$sgn(x) = \begin{cases} 1, & x > 0 \\ 0, & x = 0 \\ -1, & x < 0. \end{cases}$$

For ease of reference, our commonly used notation is summarized in Table I.

TABLE I
SUMMARY OF SYMBOLS AND NOTATIONS

D	Number of storage devices
$S_{M/N}$	Total number of CPUs of the defender/attacker
M_i^k/N_i^k	Number of CPUs allocated to scan/attack device i at time k
$\mathbf{M}^k/\mathbf{N}^k$	The defense/attack CPU allocation vector at time k
$\triangle_{D/A}$	Action set of the defender/attacker
$U_{D/A}$	Utility of the defender/attacker
B_i	Data size on device i
$\widehat{B}$	The total amount of data in a cloud system
$\mathbf{B}$	The data size vector of devices in a cloud system
R	Data protection level
$Q(\mathbf{s}, \mathbf{M})$	Q function of the defender with action $\mathbf{M}$ at state $\mathbf{s}$
$V(\mathbf{s})$	V function of the defender
$\pi(\mathbf{s}, \mathbf{M})$	The probability of the defender chooses action $\mathbf{M}$ at state $\mathbf{s}$

III. CBG-Based APT Defense Game

Colonel Blotto game (CBG) is a powerful method to study the strategic resource allocation in a competitive environment. The interactions between the APT attacker and the defender in the cloud storage system can be formulated as a Colonel Blotto game with D cloud devices, in which the defender chooses the defense CPU allocation vector $\mathbf{M}^k$ to scan the D devices, and attacker chooses the attack CPU allocation $\mathbf{N}^k$ at time k. In this pure-strategy game, the action set of the defender is

$$\triangle_D = \left\{ d_{1 \leq i \leq D} \middle| \, 0 \leq d_i \leq S_M; \, \sum_{i=1}^{D} d_i = S_M \right\},$$

and the action set of the attacker is

$$\triangle_A = \left\{ d_{1 \leq i \leq D} \middle| \, 0 \leq d_i \leq S_N; \, \sum_{i=1}^{D} d_i = S_N \right\}.$$

In this game, the utility of the defender and the attacker are denoted by u_D^k and u_A^k, respectively. Since the Colonel Blotto game is a zero-sum game, we have $u_A^k = -u_D^k$. The utility of the defender and the attacker in the CBG-based APT defense game can be defined as

$$\begin{aligned} & u_D^k\left(\mathbf{M}^k, \mathbf{N}^k\right) = -u_A^k\left(\mathbf{M}^k, \mathbf{N}^k\right) \\ & = R \cdot \widehat{B} = \sum_{i=1}^{D} B_i sgn\left(M_i^k - N_i^k\right). \end{aligned} \tag{2}$$

The CBG-based APT defense game with pure strategies has no NE, because the attacker can always adjust its CPU allocation over the devices according to the defense CPU allocation to improve its utility. Therefore, we study the CBG-based APT defense game denoted by $\mathbb{G}$ with mixed strategies, in which both the APT attacker and the defender randomize the CPU allocation strategy to fool the opponent and choose the probability distributions of the CPU allocations. More specifically, the defender in $\mathbb{G}$ chooses the mixed strategy of the CPU allocation at time k, $\mathbf{x}^k = \left[x_{i,j}^k\right]_{1\le i\le D,\ 0\le j\le S_M}$, where $x_{i,j}^k$ is the probability that the defender allocates j CPUs to device i at time slot k, i.e., $x_{i,j}^k = \Pr\left(M_i^k = j\right)$. The action set of mixed-strategy of the defender is

$$\mathrm{P}_{S_M} = \left\{ \left[p_j\right]_{0\le j\le S_M} \mid p_j \ge 0,\ \forall j;\ \sum_{j=0}^{S_M} p_j = 1 \right\}.$$

Similarly, the attacker chooses the probability of the CPU distribution over the D devices denoted by $\mathbf{y}^k = \left[y_{i,j}^k\right]_{1\le i\le D,\ 0\le j\le S_N}$, where $y_{i,j}^k = \Pr\left(N_i^k = j\right)$, with

$$\mathbf{y}^k \in \mathrm{P}_{S_N} = \left\{ \left[q_j\right]_{0\le j\le S_N} \mid q_j \ge 0,\ \forall j;\ \sum_{j=0}^{S_N} q_j = 1 \right\}.$$

The expected utility of the defender (or attacker) averaged over all the realizations of the mixed strategies is denoted by U_D^k (or U_A^k), and is given by

$$\begin{aligned} U_D^k\left(\mathbf{x}^k, \mathbf{y}^k\right) &= -U_A^k\left(\mathbf{x}^k, \mathbf{y}^k\right) \\ &= E_{(\mathbf{x},\mathbf{y})}\left[\sum_{i=1}^{D} B_i sgn\left(M_i^k - N_i^k\right)\right]. \end{aligned} \tag{3}$$

The NE of the CBG based APT defense game $\mathbb{G}$ denoted by $(\mathbf{x}^*, \mathbf{y}^*)$ provides the best-response policy in which no player can increase his or her utility by unilaterally changing the strategy, where $D \times (S_M + 1)$ matrixes $\mathbf{x}^*$ and $\mathbf{y}^*$ are given by definition as

$$U_D\left(\mathbf{x}^*, \mathbf{y}^*\right) \ge U_D\left(\mathbf{x}, \mathbf{y}^*\right),\quad \forall\ \mathbf{x} \in \mathrm{P}_{S_M} \tag{4}$$

$$U_A\left(\mathbf{x}^*, \mathbf{y}^*\right) \ge U_A\left(\mathbf{x}^*, \mathbf{y}\right),\quad \forall\ \mathbf{y} \in \mathrm{P}_{S_N}. \tag{5}$$

First, we consider a static CBG-based APT defense game denoted by $\mathbb{G}_1$ for the case of symmetric resources where the attacker and the defender have the same amount of computational resources, i.e., $S_M = S_N$. Let $\mathcal{U}(i,j)$ be the uniform distribution in the interval $[i,j]$, $\mathbf{1}_{m\times n}$ ($\mathbf{0}_{m\times n}$) be the all-1 (or 0) $m \times n$ matrix, and $\lfloor\ \rfloor$ be the lower floor function.

Theorem 1. *If $S_M = S_N$ and $B_i < \sum_{1\le h\ne i\le D} B_h$, the static APT defense game $\mathbb{G}_1$ has an NE denoted by $(\boldsymbol{x}^*, \boldsymbol{y}^*)$, where*

$$\boldsymbol{x}^* = \boldsymbol{y}^* = \begin{bmatrix} \frac{1}{\lfloor\beta B_1\rfloor+1}\mathbf{1}_{1\times(\lfloor\beta B_1\rfloor+1)} & \mathbf{0}_{1\times(S_M-\lfloor\beta B_1\rfloor)} \\ \frac{1}{\lfloor\beta B_2\rfloor+1}\mathbf{1}_{1\times(\lfloor\beta B_2\rfloor+1)} & \mathbf{0}_{1\times(S_M-\lfloor\beta B_2\rfloor)} \\ \vdots & \vdots \\ \frac{1}{\lfloor\beta B_D\rfloor+1}\mathbf{1}_{1\times(\lfloor\beta B_D\rfloor+1)} & \mathbf{0}_{1\times(S_M-\lfloor\beta B_D\rfloor)} \end{bmatrix}, \tag{6}$$

and $\beta = 2S_M/\widehat{B}$.

Proof: By Proposition 1 in [12], the NEs of the continuous CBG with symmetric players are $(\mathbf{x}^*, \mathbf{x}^*)$, with the i-th element given by

$$x_i^* \sim \mathcal{U}\left(0, \frac{2S_M B_i}{\widehat{B}}\right). \tag{7}$$

For simplicity, according to [13], we use (7) to approximate the case with discrete variables. More specifically, as the number of CPUs M_i is a natural number, we have $M_i \sim \mathcal{U}\left(\{0, 1, 2, ..., \lfloor\beta B_i\rfloor\}\right)$, and thus

$$x_{i,j}^* = \Pr\left(M_i = j\right) = \frac{1}{\lfloor\beta B_i\rfloor + 1},\ \forall\ 0 \le j \le \lfloor\beta B_i\rfloor, \tag{8}$$

which results in (6). ■

Corollary 1. *The symmetric APT defense game $\mathbb{G}_1$ has zero expected data protection level and $U_D = U_A = 0$ at the NE.*

Proof: By (1) and (6), we have

$$E_{(\mathbf{x}^*,\mathbf{x}^*)}\left[R\right] = E_{(\mathbf{x}^*,\mathbf{x}^*)}\left(\frac{1}{\widehat{B}}\sum_{i=1}^{D} B_i sgn\left(M_i - N_i\right)\right) \tag{9}$$

$$= \frac{1}{\widehat{B}}\sum_{i=1}^{D} B_i E_{(\mathbf{x}^*,\mathbf{x}^*)}\left(sgn\left(M_i - N_i\right)\right) \tag{10}$$

$$= \frac{1}{\widehat{B}}\sum_{i=1}^{D} B_i\left(\Pr\left(N_i < M_i\right) - \Pr\left(N_i > M_i\right)\right) = 0. \tag{11}$$

By (2) and (3) , we have $U_D = U_A = 0$.

■

Remark: If the APT attacker and the defender have equal total number of CPUs, and there is no dominant device, i.e., $B_i < \sum_{1\le h\ne i\le D} B_h$, it is shown in (8) that both the attacker and the defender use equal probability $1/\left(\lfloor\beta B_i\rfloor + 1\right)$ to choose a number from $\{0, 1, ..., \lfloor\beta B_i\rfloor\}$ to scan a device. Since the game makes a tie, the expected data protection level is 0, so does the utility of the defender.

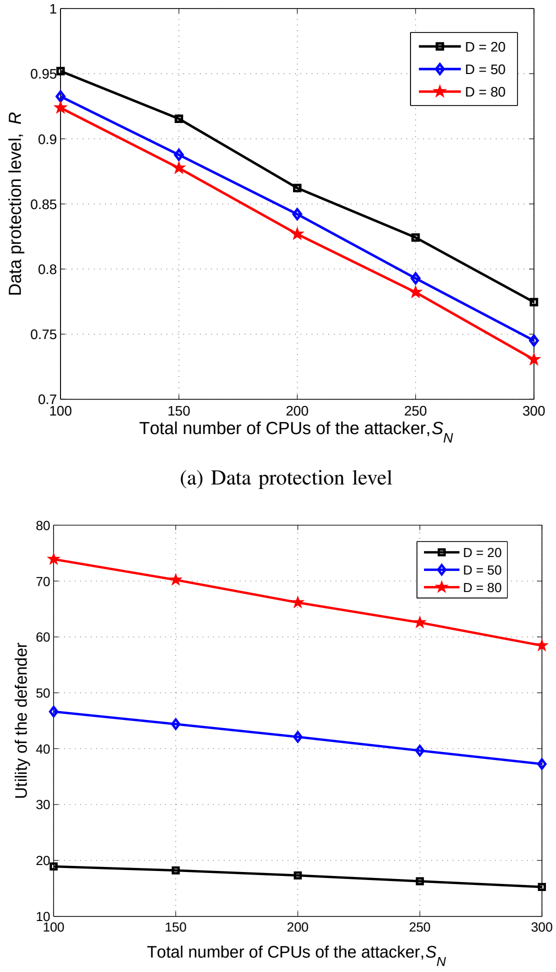
A CBG-based APT defense game with asymmetric players is denoted by $\mathbb{G}_2$, in which the attacker and the defender have different amount of resources and are competing over storage devices with equal data size, i.e., $B_i = B,\ \forall\ 1 \le i \le D$. Theorem 2 characterizes the Nash equilibrium for this game.

Theorem 2. *If $\frac{2}{D} \le \frac{S_N}{S_M} \le 1$ and $B_i = B,\ \forall\ 1 \le i \le D$, the NE of the APT defense game $\mathbb{G}_2$ is given by*

$$\boldsymbol{x}^* = \begin{bmatrix} \mathbf{0}_{D\times 1} & \frac{1}{\lfloor\frac{2S_M}{D}\rfloor}\mathbf{1}_{D\times\lfloor\frac{2S_M}{D}\rfloor} & \mathbf{0}_{D\times\left(S_M-\lfloor\frac{2S_M}{D}\rfloor\right)} \end{bmatrix} \tag{12}$$

$$\boldsymbol{y}^* = \Bigg[\left(1 - \frac{S_N}{S_M}\right)\mathbf{1}_{D\times 1} \quad \left(\frac{S_N}{S_M\lfloor\frac{2S_M}{D}\rfloor}\right)\mathbf{1}_{D\times\lfloor\frac{2S_M}{D}\rfloor} \quad \mathbf{0}_{D\times\left(S_M-\lfloor\frac{2S_M}{D}\rfloor\right)} \Bigg]. \tag{13}$$

Proof: According to Theory 2 in [11], the NEs of the



(a) Data protection level

(b) Utility of the defender

Fig. 2. Performance of the CBG-based defense game $\mathbb{G}_2$ at the NE, with $S_M = 1000$ and $B = 1$.

continuous CBG are given by

$$x_i^* \sim \mathcal{U}\left(0, \frac{2S_M}{D}\right) \tag{14}$$

$$y_i^* \sim \left(1 - \frac{S_N}{S_M}\right)\delta(N_i) + \frac{S_N}{S_M}\mathcal{U}\left(0, \frac{2S_M}{D}\right). \tag{15}$$

According to [13], we use (14) and (15) to approximate the CBG with discrete resources, and thus

$$x_{i,j}^* = \Pr\left(M_i = j\right) = \frac{1}{\lfloor \frac{2S_M}{D} \rfloor}, \ \forall\, 1 \leq j \leq \lfloor \frac{2S_M}{D} \rfloor \tag{16}$$

$$y_{i,j}^* = \Pr\left(N_i = j\right) = \begin{cases} 1 - \frac{S_N}{S_M}, & j = 0 \\ \frac{S_N}{S_M \lfloor \frac{2S_M}{D} \rfloor}, & \forall\, 1 \leq j \leq \lfloor \frac{2S_M}{D} \rfloor. \end{cases} \tag{17}$$

Thus, we have (12) and (13). ■

Corollary 2. *At the NE of the APT defense game $\mathbb{G}_2$, the average data protection level is $1 - S_N/S_M$ and*

$$U_D = -U_A = \left(1 - \frac{S_N}{S_M}\right)\widehat{B}. \tag{18}$$

Proof: According to (1), (12), (13) and (9)-(11), as $B_i = B$, we have

$$\begin{aligned} E_{(\mathbf{x}^*,\mathbf{y}^*)}\left[R\right] &= \frac{1}{\widehat{B}}\sum_{i=1}^{D} B_i \left(\Pr\left(N_i < M_i\right) - \Pr\left(N_i > M_i\right)\right) \\ &= \frac{1}{\widehat{B}}\sum_{i=1}^{D} B_i \Big(\left(\Pr\left(N_i < M_i\right) - \Pr\left(N_i > M_i\right)\right)|_{N_i=0} \\ &\quad + \left(\Pr\left(N_i < M_i\right) - \Pr\left(N_i > M_i\right)\right)|_{N_i \neq 0}\Big) \\ &= \frac{1}{\widehat{B}}\sum_{i=1}^{D} B_i \left(1 - \frac{S_N}{S_M}\right) = 1 - \frac{S_N}{S_M}. \end{aligned} \tag{19}$$

Similarly, by (2) and (3), we have (18). ■

Remark: If the defender is more powerful than the attacker, the defender scans each device with at least one CPU with probability one, and a subset of the devices is not attacked by the APT attacker since the attacker has to match the defender on the remaining devices. Therefore, the defender wins the game, and the utility increases with the total data amount. The expected data protection level increases with the resources advantage of the defender over the attacker, i.e., S_N/S_M.

The performance of the defense game $\mathbb{G}_2$ with $S_M = 1000$, $D = 20$ and $B = 1$ is presented in Fig. 2, showing that the data protection level decreases from 0.95 to 0.77, as S_N increases from 100 to 300. In addition, the data protection level decreases while the system size increases from 20 to 80 if $S_N = 100$, e.g., R changes from 0.95 to 0.925, because the defender does not have enough CPUs to scan all the devices to beat the attacker. Therefore, the utility of the defender decreases as S_N increasing given a fixed system.

IV. Dynamic APT Defense Game

The repeated interactions of the APT attacker and the defender can be formulated as a dynamic APT defense game, in which the defender can apply a policy hill-climbing based defense strategy to derive the optimal CPU allocation distribution over the D devices via trial-and-error without knowing the APT attack model. As a simple extension of Q-learning for the game with mixed strategies, the PHC algorithm maintains a quality function of Q-function and the current mixed strategy (strategy table π). The defender chooses the CPU allocation distribution $\mathbf{x}$ according to the observed system state, and allocates the CPUs following the probability distribution table π.

At time k, the defender uses the CPU allocations of the attacker at time $k-1$ as the system state, i.e., $\mathbf{s}^k = \mathbf{N}^{k-1}$, which is the basis to choose the mixed strategy $\mathbf{x}$. Then the defender observes the CPU allocation of the APT attacker over D devices $\mathbf{N}^k$, which is used as the new system state at time $k+1$. The defender also evaluates its utility at time k, and then updates the Q-function as follows:

$$\begin{aligned} Q\left(\mathbf{s}^k, \mathbf{M}^k\right) \leftarrow &\left(1-\alpha\right) Q\left(\mathbf{s}^k, \mathbf{M}^k\right) \\ &+ \alpha\left(u_D + \gamma V\left(\mathbf{s}^{k+1}\right)\right) \end{aligned} \tag{20}$$

$$V\left(\mathbf{s}^k\right) = \max_{\hat{\mathbf{M}} \in \triangle_D} Q\left(\mathbf{s}^k, \hat{\mathbf{M}}\right), \tag{21}$$

Algorithm 1 Resource allocation against APT using PHC algorithm.

```
Initialize $\alpha = 0.5$, $\gamma = 0.9$, $\delta = 0.02$, $\pi(\mathbf{s}, \mathbf{M}) = \frac{1}{|\triangle_D|}$, $Q(\mathbf{s}, \mathbf{M}) = 0$,
        $V(\mathbf{s}) = 0$, $\forall \mathbf{s}$, $\mathbf{M}$
Randomly choose $\mathbf{N}^0 \in \triangle_A$
For $k = 1, 2, 3, ...$
   Update the state $\mathbf{s}^k$ = $\mathbf{N}^{k-1}$
   Choose $\mathbf{M}^k$ with probability $\pi(\mathbf{s}^k, \mathbf{M}^k)$
   Observe $u_D$ and the CPU allocation of the attacker, $\mathbf{N}^k$
   Update $Q(\mathbf{s}^k, \mathbf{M}^k)$ via Eq. (20)
   Update $V(\mathbf{s}^k)$ via Eq. (21)
   Update $\pi(\mathbf{s}^k, \mathbf{M}^k)$ via Eq. (22)
End for
```

where $\alpha \in (0, 1]$ is the learning rate of the defender, $\gamma \in [0, 1]$ is the discount factor that indicates the weight of a future payoff over the current one, and $V\left(\mathbf{s}^k\right)$ is the value function representing the highest value of Q-function at state $\mathbf{s}^k$.

The defender updates the mixed strategies table π by increasing the probability that it selects the highest valued action by δ, with $0 < \delta \leq 1$, and decreasing the other probability by $-\delta / \left(|\triangle_D| - 1\right)$, i.e.,

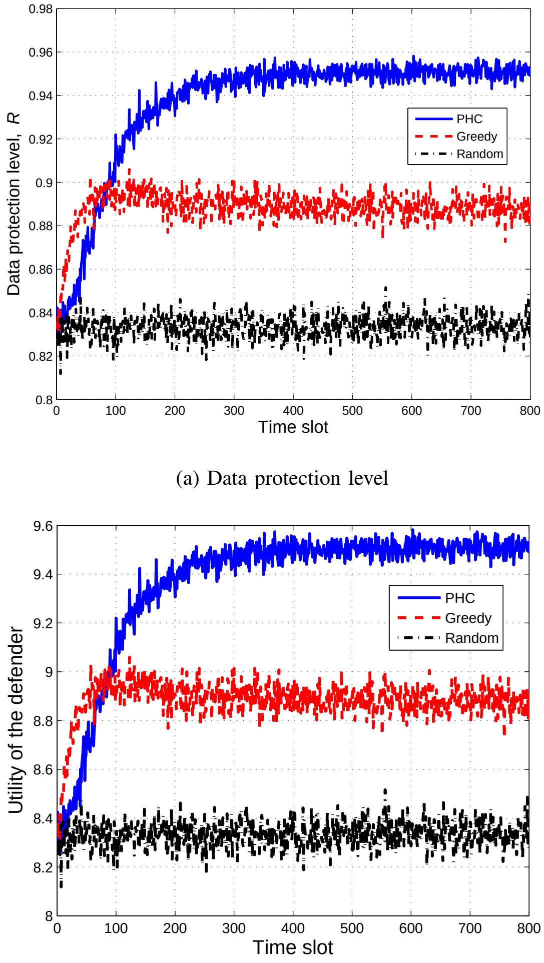
$$\pi(\mathbf{s}^k, \mathbf{M}^k) \leftarrow \pi(\mathbf{s}^k, \mathbf{M}^k) + \begin{cases} \delta & \text{if } \mathbf{M}^k = \underset{\hat{\mathbf{M}} \in \triangle_D}{\arg\max}\, Q\left(\mathbf{s}^k, \hat{\mathbf{M}}\right) \\ \frac{-\delta}{|\triangle_D|-1} & \text{otherwise} \end{cases}. \tag{22}$$

The PHC based APT defense algorithm is summarized in Algorithm 1.

V. Simulation Results

Simulations were performed to evaluate the performance of the proposed PHC based APT defense strategy, with the PHC parameters set at: $\alpha = 0.5$, $\gamma = 0.9$ and $\delta = 0.02$. As benchmarks, a random strategy and a greedy strategy were considered, and in a greedy defense strategy the defender chose the resource allocation to maximize its estimated immediate utility based on the previous attack strategy.

We first evaluate the PHC based dynamic APT defense game in which the defender and the attacker have total budget of 1000 and 100 CPUs, respectively. The data sizes at 10 devices are equal, i.e., $B = 1$. As shown in Fig. 3 (a), the data protection level increases over time in our proposed PHC based defense scheme, and converges to 0.91 after 210 time slots, which is about 7.34% and 14.46% higher than the greedy strategy and the random strategy, respectively. Consequently, as shown in Fig. 3 (b), the utility of the defender increases quickly after the start of the learning process, and converges to a certain value that is much higher than the two benchmark strategies. For example, our proposed scheme exceeds the random strategy by 14.46% at time 350 because the PHC scheme adjusts the defender's policy over time. We evaluate the average performance of the dynamic APT defense game in Fig. 4, showing that the data protection level decreases with the number of CPUs of the attacker, because more data can be stolen, and our proposed approach protects up to 99% and 94% of data against APT with 50 and 100 CPUs, respectively.



(a) Data protection level

(b) Utility of the defender

Fig. 3. Performance of the dynamic APT defense game, with $S_M = 1000$, $S_N = 100$, $D = 10$ and $B = 1$.

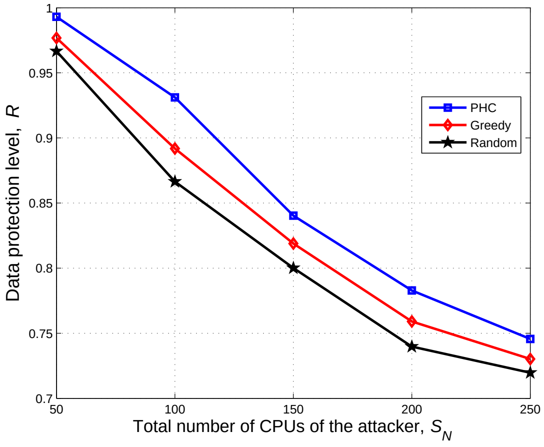


Fig. 4. Average performance of the dynamic APT defense game, with $S_M = 1000$, $D = 10$ and $B = 1$.

VI. Conclusion

In this paper, we modeled the computational resource (or CPU) allocation problem under APT attacks as a two-player zero-sum game, where the defender aims at maximizing the data protection level of the cloud system by randomizing the amounts of CPUs allocated to each cloud device, which is converted to a Colonel Blotto Game. We derived the NE of the static APT defense game, and investigated the impacts of the data size, the number of devices and the CPU resource constraints on the data protection level and the defender's utility. We also proposed a policy hill-climbing based APT defense strategy for the defender to scan the devices while the attack model is unknown in the dynamic defense game. Simulation results show that our proposed strategy can significantly improve the data protection level and the utility of the storage defender. For instance, the data protection level increases by 14.46% compared with the random strategy. In our future work, we plan to use deep Q-learning techniques to further accelerate the learning speed and improve the data privacy of the cloud system for the case with a large number of CPUs.

References

[1] C. Tankard, "Advanced persistent threats and how to monitor and deter them," *Network Security*, vol. 2011, no. 8, pp. 16–19, Aug. 2011.

[2] J. Vukalović and D. Delija, "Advanced persistent threats-detection and defense," in *Proc. IEEE Inf. Commun. Technol. Electronics and Microelectronics*, pp. 1324–1330, Opatija, May 2015.

[3] N. Cao, C. Wang, M. Li, K. Ren, and W. Lou, "Privacy-preserving multi-keyword ranked search over encrypted cloud data," *IEEE Trans. Parallel and Distributed Systems*, vol. 25, no. 1, pp. 222–233, Jan. 2014.

[4] M. Van Dijk, A. Juels, A. Oprea, and R. L. Rivest, "Flipit: The game of "stealthy takeover"," *J. of Cryptology*, vol. 26, no. 4, pp. 655–713, Oct. 2013.

[5] J. Pawlick, S. Farhang, and Q. Zhu, "Flip the cloud: Cyber-physical signaling games in the presence of advanced persistent threats," in *Proc. Int. Conf. Decision and Game Theory for Security*, pp. 289–308, London, UK, Nov. 2015.

[6] M. Zhang, Z. Zheng, and N. B. Shroff, "A game theoretic model for defending against stealthy attacks with limited resources," in *Proc. Int. Conf. Decision and Game Theory for Security*, pp. 93–112, London, UK, Nov. 2015.

[7] P. Hu, H. Li, H. Fu, D. Cansever, and P. Mohapatra, "Dynamic defense strategy against advanced persistent threat with insiders," in *Proc. IEEE Int. Conf. Comput. Commun. (INFOCOM)*, pp. 747–755, HK, May 2015.

[8] L. Xiao, D. Xu, Y. Li, N. B. Mandayam, and H. V. Poor, "Cloud storage defense against advanced persistent threats: A prospect theoretic study," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Washington, DC, Dec. 2016, in press.

[9] L. Xiao, D. Xu, Y. Li, N. B. Mandayam, and H. V. Poor, "Cloud storage defense against advanced persistent threats: A prospect theoretic study," *IEEE J. Sel. Areas Commun*, in press.

[10] L. Xiao, J. Liu, Q. Li, N. B. Mandayam, and H. V. Poor, "User-centric view of jamming games in cognitive radio networks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 12, pp. 2578–2590, Dec. 2015.

[11] B. Roberson, "The Colonel Blotto game," *Economic Theory*, vol. 29, no. 1, pp. 1–24, Sep. 2006.

[12] C. Thomas, "N-dimensional Colonel Blotto game with asymmetric battlefield values," Working paper, The University of Texas at Austin. 2012.

[13] S. Hart, "Discrete Colonel Blotto and General Lotto games," *International Journal of Game Theory*, vol. 36, no. 3-4, pp. 441–460, 2008.

[14] M. Hajimirsadeghi, G. Sridharan, W. Saad, and N. B. Mandayam, "Inter-network dynamic spectrum allocation via a Colonel Blotto game," in *Proc. IEEE Annu. Conf. Inf. Sci. Syst. (CISS)*, pp. 252–257, Princeton, NJ, Mar. 2016.

[15] M. Labib, S. Ha, W. Saad, and J. H. Reed, "A Colonel Blotto game for anti-jamming in the Internet of Things," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, pp. 1–6, San Diego, CA, Dec. 2015.

[16] N. Namvar, W. Saad, N. Bahadori, and B. Kelley, "Jamming in the Internet of Things: A game-theoretic perspective," *arXiv preprint arXiv:1607.06255*, 2016.

[17] M. Bowling and M. Veloso, "Rational and convergent learning in stochastic games," in *Proc. Int. Joint Conf. Artificial Intelligence*, pp. 1021–1026, Seattle, Washhington, Aug. 2001.