# Cloud Storage Defense Against Advanced Persistent Threats: A Prospect Theoretic Study

Liang Xiao, *Senior Member, IEEE*, Dongjin Xu, Caixia Xie, Narayan B. Mandayam, *Fellow, IEEE*, and H. Vincent Poor, *Fellow, IEEE*

***Abstract*— Cloud storage is vulnerable to advanced persistent threats (APTs), in which an attacker launches stealthy, continuous, and targeted attacks on storage devices. In this paper, prospect theory (PT) is applied to formulate the interaction between the defender of a cloud storage system and an APT attacker who makes subjective decisions that sometimes deviate from the results of expected utility theory, which is a basis of traditional game theory. In the PT-based cloud storage defense game with pure strategy, the defender chooses a scan interval for each storage device and the subjective APT attacker chooses his or her interval of attack against each device. A mixed-strategy subjective storage defense game is also investigated, in which each subjective defender and APT attacker acts under uncertainty about the action of its opponent. The Nash equilibria (NEs) of both games are derived, showing that the subjective view of an APT attacker can improve the utility of the defender. A Q-learning-based APT defense scheme that the storage defender can apply without being aware of the APT attack model or the subjectivity model of the attacker in the dynamic APT defense game is also proposed. Simulation results show that the proposed defense scheme suppresses the attack motivation of subjective APT attackers and improves the utility of the defender, compared with the benchmark greedy defense strategy.**

***Index Terms*— Cloud storage, advanced persistent threats, game theory, prospect theory, Q-learning.**

## I. Introduction

CLOUD storage is vulnerable to advanced persistent threats (APTs), in which an attacker launches sophisticated, stealthy, continuous, and targeted attacks. By applying multiple sophisticated attack methods, APT attackers aim to steal information from a target cyber system including cloud storage over an extended period of time without being noticed. APT attackers usually take multiple attack phases and study the defense policy of the target system in advance, making it challenging to detect APTs and estimate the attack duration. According to [1], more than 65% of the organizations responding to the survey in 2014 witnessed an increase of APT attacks, and the current doctrine against APTs is to detect them as early as possible [2].

Game theory is an important tool for studying APT attacks. In the seminal work in [3], the interaction between an APT attacker and a defender was formulated as a stealthy takeover game. Most existing game theoretic studies of APT attacks are based on expected utility theory (EUT), in which each player chooses its strategy to maximize the expected utility. However, as human beings, APT attackers are not always rational as assumed in traditional game theoretic models and they sometimes make subjective decisions under uncertainties that deviate from the results of expected utility theory, such as risk seeking, loss aversion and the nonlinear weighting of gains and losses [4], as illustrated by Allais paradox described in [5]. Similarly, the defenders also are subject to such subjective traits in decision-making, thus making the model here amenable to use of prospect theory (PT).

By using the probability weighting function and value function, prospect theory can model the subjective decision-making processes of end-users and successfully explain the deviations of their decisions from EUT-based results [6]. Prospect theory has been successfully applied to study the interactions between people in many areas, such as social sciences [7], [8], communication networks [9]–[15], and smart energy management [16], [17].

In this paper, prospect theory is applied to study cloud storage defense against advanced persistent threats and investigate the impact of end-user subjectivity on storage defense. More specifically, we formulate a cloud storage defense game, in which a subjective attacker chooses his or her interval to launch APT attacks to compromise storage devices and a defender chooses its scan interval to recapture the compromised storage devices. Prelec's probability weighting function [18] is applied to model the subjective decision-making of the attacker and defender under uncertain attack durations in the pure-strategy game or uncertain action of their opponent in the mixed-strategy game. The Nash equilibria (NEs) of both subjective games are derived to investigate the impact of end-user subjectivity on the APT defense games.

Manuscript received April 30, 2016; revised September 30, 2016; accepted November 28, 2016. Date of publication January 26, 2017; date of current version April 26, 2017. This work was supported in part by the National Natural Science Foundation of China under Grant 61671396 and Grant 61271242, in part by the U.S. National Science Foundation under Grant CMMI-1435778, Grant ECCS-1549881, Grant CNS-1421961, and Grant ACI-1541069, and in part by the CCF-Venustech Hongyan Research Initiative 2010–2016.

L. Xiao is with the Department of Communication Engineering, Xiamen University, Xiamen 361005, China, and with the Key Laboratory of Underwater Acoustic Communication and Marine Information Technology Ministry of Education, Xiamen University, Xiamen, China, and also with the Beijing Key Laboratory of IoT Information Security Technology, Institute of Information Engineering, Chinese Academy of Sciences, Beijing 100000, China (e-mail: lxiao@xmu.edu.cn).

D. Xu and C. Xie are with the Department of Communication Engineering, Xiamen University, Xiamen, China.

N. B. Mandayam is with the Wireless Information Network Laboratory, Department Electrical and Computer Engineering, Rutgers University, New Brunswick, NJ 08816 USA (e-mail: narayan@winlab.rutgers.edu).

H. V. Poor is with the Department of Electrical Engineering, Princeton University, Princeton, NJ 08544 USA (e-mail: poor@princeton.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSAC.2017.2659418

A Q-learning based APT defense strategy is proposed for the cloud storage defender who is unaware of the attack model or the subjectivity model of the APT attacker to derive the optimal storage scan policy via trials in a dynamic form of the game. Based on the iterative Bellman equation, the Q-learning algorithm, as a model-free reinforcement learning technique, is convenient to implement and can achieve the optimal policy in the Markov decision process (MDP). Simulations are performed to evaluate the performance of the Q-learning based APT defense scheme, showing that it can suppress the attack motivation of subjective APT attackers and improve the utility of the defender.

The main contributions of this work can be summarized as follows:

- We formulate a PT-based cloud storage defense game, in which both the APT attacker and the storage defender hold subjective views to choose their attack or scan interval at each cloud storage device under uncertain attack durations in the pure-strategy game or action of the opponents in the mixed-strategy game. We derive the NEs of the PT-based storage defense games and provide the conditions under which the equilibria exist, showing that a subjective APT attacker tends to attack less frequently.
- We propose a Q-learning based APT defense scheme for the cloud storage defender to derive the optimal scan interval policy without knowing the APT attack model or subjectivity model in the dynamic storage defense games against subjective APT attackers.

The remainder of the paper is organized as follows. We review related work in Section II and present the system model in Section III. We present a static subjective storage defense game with pure-strategy in Section IV and investigate the mixed-strategy PT-based game in Section V. We propose the Q-learning based APT defense schemes in dynamic storage defense games in Section VI. We provide simulation results in Section VII and conclude in Section VIII.

## II. Related Work

Game theoretic approaches for modeling and studying APT attacks have received considerable attention. In the seminal work of [3], a Flipit game was proposed to formulate the stealthy and continuous attacks of APT. The game between an overt defender and a stealthy attacker was investigated in [19], showing that the periodic defense strategy is the best response against a non-adaptive attacker. A cyber-physical signaling game among an APT attacker, a cloud defender and a mobile device was formulated in [20], in which the mobile device decides whether to trust the commands from the cloud under APTs. The defense based on the dynamic programming algorithm proposed in [21] provides a nearly optimal solution against APT attacks. The two-layer APT defense game formulated in [22] studies the joint threats from an APT attacker and insiders in the cyber system.

Prospect theory has been applied to study wireless communications and network security. For instance, a random access game formulated in [9] applies prospect theory to study channel access between two subjective end-users in wireless networks. The impact of user subjectivity on both the wireless random access and data pricing games was identified in [10] based on prospect theory. The spectrum investment of subjective secondary operators was investigated in [11], and a PT-based sensing and leasing method was derived. The PT-based pricing and resource allocation scheme proposed in [15] improves the revenue of service providers in presence of subjective users. A PT-based anti-jamming transmission game formulated in [23] investigates the impact of the subjectivity of end-users and jammers on the throughput in cognitive radio networks.

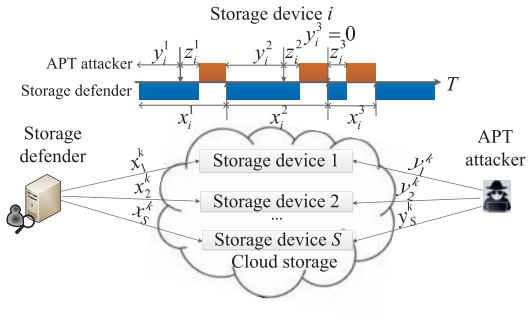



Fig. 1. Illustration of a cloud storage defense game, in which the defender scans storage device $i$ at $x_i^k$ interval, while the APT attacker takes a duration $z_i^k$ to complete the $k$-th attack against device $i$ after attack interval $y_i^k$, with $1 \le i \le S$ and $k > 0$.

Game theory can help develop security mechanisms for cloud computing. For example, the game theoretic study of coresident attacks in [24] develops a semi-supervised learning based defense strategy to increase the attack costs. In the PT-based storage defense game against subjective APT attacks as presented in [25], we derived the NE of the game under uncertain APT attack durations. In this paper, we consider the generic APT scenarios with multiple storage devices and multiple attack duration levels, instead of the special case with a single device as assumed in [25]. We also present a dynamic storage defense game with mixed-strategy and present a Q-learning based defense strategy to resist subjective APT attacks under uncertain device scan intervals.

## III. System Model

We consider a cloud storage system consisting of $S$ storage devices that are threatened by a subjective APT attacker ($A$) and are protected by a storage defender ($D$), as shown in Fig. 1. The defender chooses the time interval to perform the $k$-th detection at storage device $i$ against APT attacks, denoted by $x_i^k$, with $1 \le i \le S$. It is clear that $x_i^k > 0$, because the defender has to take time to scan a storage device to detect APT attacks. Upon detecting APT attacks, the defender restores a compromised storage device and provides privacy for the data stored on the device. The defender is unaware of whether a storage device is compromised unless the device is monitored.

According to the APT model as given in [21], the APT attacker can apply advanced and sophisticated methods and inject multiple types of malware to estimate the defense strategy of the target system. The attacker can also determine whether the attack successfully controls the target storage device according to the data stolen from the device, and observe the size of the stolen data to determine when the attack

TABLE I
SUMMARY OF SYMBOLS AND NOTATION

| Notation | Definition |
|---|---|
| $S$ | Number of storage devices |
| $\alpha_{A/D}$ | Objective weight of the attacker/defender |
| $x_i^k/y_i^k$ | Defense/attack interval at time $k$ against device $i$ |
| $z_i^k$ | Duration to complete the $k$-th attack against device $i$ |
| $G_i$ | Defense gain of device $i$ |
| $C_i$ | Attack cost against device $i$ |
| $L$ | Number of non-zero attack duration levels |
| $M$ | Number of detection interval levels |
| $N$ | Number of non-zero attack interval levels |
| $\mathbf{p}/\mathbf{q}$ | Mixed-strategy of the defender/attacker |

is detected and stopped by the defender. The attacker waits $y_i^k$ time before launching the $k$-th APT attack against storage device $i$, once the defender detects attacks and restores that storage device. The duration for the attacker to complete its $k$-th attack at storage device $i$, denoted by $z_i^k$, is in general a positive random variable that is unknown to both players. The defender is assumed to take charge of all the $S$ storage devices at the beginning.

We use the Prelec function in [18] to explain how a subjective attacker or defender over-weighs low-probability events and under-weighs outcomes having a high probability. Being easy to analyze, the Prelec function has been used to explain the human decision deviations from EUT results in network security [12], [16]. Therefore, we apply this probability weighting function to model the subjective probability of the attacker (or defender), denoted by $w_A$ (or $w_D$), and given by

$$w_r(p) = \exp\big(-(-\ln p)^{\alpha_r}\big), \tag{1}$$

where $\alpha_r \in (0, 1]$ as the objective weight of player $r$ represents the distortion that subjectivity causes in making decisions. For example, if $\alpha_A = 1$, the attacker is objective and $w_A(p) = p$. Table 1 summarizes the notation used in the paper.

## IV. SUBJECTIVE STORAGE DEFENSE GAME WITH PURE-STRATEGY

The interaction between an APT attacker and a storage defender over $S$ storage devices is formulated as a subjective cloud storage defense game with pure-strategy, denoted by $\mathbb{G}$. In this game, the storage defender chooses the scan interval $x_i$ for storage device $i$, and the attacker decides his or her attack interval $y_i$ against storage device $i$. The defense interval and the attack interval are normalized for simplicity of analysis. According to the maximum scan interval of the defender denoted by $T$, the attacker and defender compete to take charge of the $S$ storage devices, with $0 < x_i \leq 1$ and $0 \leq y_i \leq 1$, $\forall 1 \leq i \leq S$. If the attack interval denoted by $T_a$ is greater than $T$, the game can be divided into $K = \lceil T_a/T \rceil$ interactions, with $y_i = 1, \forall i < K$ and $y_K = \mod(T_a, T)$, where $\lceil\rceil$ is the ceiling function.

The gain of the defender for a longer scan interval at storage device $i$ is denoted by $G_i$, and the attack cost against device $i$ is denoted by $C_i$. As shown in Fig. 1, the time interval during which storage device $i$ is not compromised and the data is safe is $\min((y_i + z_i)/x_i, 1)$. Therefore, the attack rate denoted by $R$ is defined as the normalized "bad" interval during which data privacy is at risk averaged over $S$ storage devices, and is given by

$$R = 1 - \frac{1}{S}\sum_{i=1}^{S} \min\left(\frac{y_i + z_i}{x_i}, 1\right). \tag{2}$$

The utility of the defender depends on the normalized "good" interval during which each storage device is protected by the defender, i.e., $\min((y_i + z_i)/x_i, 1)$, and the gain of a longer defense interval. Similar to the game model presented in [21], the utility of the defender denoted by $u_D$ is defined as

$$u_D(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{S}\left(\min\left(\frac{y_i + z_i}{x_i}, 1\right) + x_i G_i\right). \tag{3}$$

The utility of the attacker denoted by $u_A$ is defined as

$$u_A(\mathbf{x}, \mathbf{y}) = -\sum_{i=1}^{S}\left(\min\left(\frac{y_i + z_i}{x_i}, 1\right) + \mathrm{I}(y_i < x_i)C_i\right), \tag{4}$$

where the indicator function $\mathrm{I}(\xi) = 1$ if $\xi$ is true and 0 otherwise. The desire of the attacker to steal information from the storage device is modeled by $-\min\big((y_i + z_i)/x_i, 1\big)$.

The time interval for the attacker to successfully launch an APT attack against storage device $i$ $z_i$ is difficult to estimate and is quantized into $L$ non-zero levels following the distribution $[P_l^i]_{0 \leq l \leq L}$, where $P_l^i = \Pr(z_i = l/L)$, $\forall 0 \leq l \leq L$ and $1 \leq i \leq S$. By definition, we have $P_l^i \geq 0$ and $\sum_{l=0}^{L} P_l^i = 1$. The expected utilities of the defender and the attacker over the realizations of attack duration $z_i$, denoted by $U_D^{EUT}$ and $U_A^{EUT}$, respectively, are given by (3) and (4) as

$$U_D^{EUT}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{S}\left(\sum_{l=0}^{L} P_l^i \min\left(\frac{y_i L + l}{x_i L}, 1\right) + x_i G_i\right) \tag{5}$$

$$U_A^{EUT}(\mathbf{x}, \mathbf{y}) = -\sum_{i=1}^{S}\left(\sum_{l=0}^{L} P_l^i \min\left(\frac{y_i L + l}{x_i L}, 1\right) + \mathrm{I}(y_i < x_i)C_i\right). \tag{6}$$

The Prelec probability weighting function in (1) is used to model the subjective decision-making of the players under uncertain attack durations. The PT-based utilities of the subjective defender and the attacker, denoted by $U_D^{PT}$ and $U_A^{PT}$, respectively, are given by replacing the objective probability $P_l^i$ in (5) and (6) with the subjective probability $w(P_l^i)$, i.e.,

$$U_D^{PT}(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^{S}\left(\sum_{l=0}^{L} w_D(P_l^i) \min\left(\frac{y_i L + l}{x_i L}, 1\right) + x_i G_i\right) \tag{7}$$

$$U_A^{PT}(\mathbf{x}, \mathbf{y}) = -\sum_{i=1}^{S}\left(\sum_{l=0}^{L} w_A(P_l^i) \min\left(\frac{y_i L + l}{x_i L}, 1\right) + \mathrm{I}(y_i < x_i)C_i\right). \tag{8}$$

A Nash equilibrium of the PT-based storage defense game $\mathbb{G}$, denoted by $(\mathbf{x}^*, \mathbf{y}^*)$, consists of the best response of the player in terms of the PT-based utility, if the opponent uses the NE strategy. By definition, we have

$$\mathbf{x}^* = \arg\max_{\mathbf{x}} U_D^{PT}(\mathbf{x}, \mathbf{y}^*), \quad \forall \mathbf{x} \tag{9}$$

$$\mathbf{y}^* = \arg\max_{\mathbf{y}} U_A^{PT}(\mathbf{x}^*, \mathbf{y}), \quad \forall \mathbf{y}. \tag{10}$$

The objective weight of the attacker $\alpha_A$ can be estimated by the defender according to the defense history or provided by security agents. Similarly, the attacker can obtain both $\alpha_D$ and $\alpha_A$ according to the attack history against the target storage system. On the other hand, if $\alpha_A$ and $\alpha_D$ are unknown, the defender can apply the Q-learning based defense strategy to derive its best defense policy which converges to the NEs, as described in Section VI.

We first evaluate the NE of the static cloud storage defense game $\mathbb{G}$ with a single storage device and two non-zero attack duration levels, i.e., the probability mass function of $z$ is given by $[P_0, P_1, 1 - P_0 - P_1]$. The index $i$ in the superscript is omitted when no confusion occurs.

*Theorem 1: The subjective storage defense game $\mathbb{G}$ with $S = 1$ and $L = 2$ has an NE $(x^*, y^*) = (\frac{1}{2}, 0)$, if*

$$I_1: \begin{cases} G \le \exp\Big(-(-\ln P_1)^{\alpha_D}\Big) & \text{(11a)} \\ C \le \exp\Big(-(-\ln P_0)^{\alpha_A}\Big); & \text{(11b)} \end{cases}$$

*$(x^*, y^*) = (1, 0)$, if*

$$I_2: \begin{cases} G > \exp\Big(-(-\ln P_1)^{\alpha_D}\Big) & \text{(12a)} \\ C < \exp\Big(-(-\ln P_0)^{\alpha_A}\Big) \\ \quad + 0.5\exp\Big(-(-\ln P_1)^{\alpha_A}\Big); & \text{(12b)} \end{cases}$$

*and $(x^*, y^*) = (1, 1)$, if*

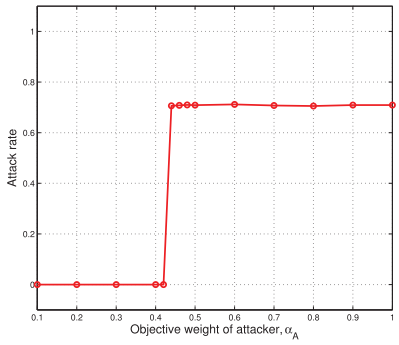
$$I_3: C > \exp\Big(-(-\ln P_0)^{\alpha_A}\Big) + 0.5\exp\Big(-(-\ln P_1)^{\alpha_A}\Big). \tag{13}$$

*Proof:* By (1) and (8), we see that if $0 \le y < \frac{1}{2}$,

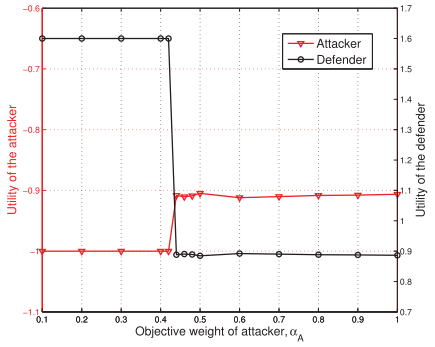
$$\begin{aligned} U_A^{PT}\big(\frac{1}{2}, 0\big) &= -w_A(P_1) - w_A(1 - P_0 - P_1) - C \\ &\ge -2y w_A(P_0) - w_A(P_1) \\ &\quad - w_A(1 - P_0 - P_1) - C \\ &= U_A^{PT}\big(\frac{1}{2}, y\big). \end{aligned} \tag{14}$$

Similarly, if (11b) holds, $\forall \frac{1}{2} \le y \le 1$, we have

$$\begin{aligned} U_A^{PT}\big(\frac{1}{2}, 0\big) &= -w_A(P_1) - w_A(1 - P_0 - P_1) - C \\ &\ge -w_A(P_0) - w_A(P_1) - w_A(1 - P_0 - P_1) \\ &= U_A^{PT}\big(\frac{1}{2}, y\big). \end{aligned} \tag{15}$$




(a) Attack rate




(b) Utility

Fig. 2. Performance of the static PT-based storage defense game $\mathbb{G}$ at the NEs, with $C = 0.62$, $G = 0.6$, $P_0 = 0.46$, $P_1 = 0.5$, $\alpha_D = 1$ and $L = 2$.

Thus, (10) holds for $(x^*, y^*) = (\frac{1}{2}, 0)$. By (7), if $0 < x \le \frac{1}{2}$, we have

$$U_D^{PT}(x, 0) = w_D(P_1) + w_D(1 - P_0 - P_1) + xG, \tag{16}$$

which increases linearly with $x$ and is maximized at $x = \frac{1}{2}$.

Similarly, if $\frac{1}{2} < x \le 1$, we have

$$U_D^{PT}(x, 0) = \frac{1}{2x} w_D(P_1) + w_D(1 - P_0 - P_1) + xG, \tag{17}$$

and

$$\left.\frac{\partial^2 U_D^{PT}}{\partial x^2}\right|_{y=0} = \frac{1}{x^3} w_D(P_1) \ge 0, \tag{18}$$

indicating that $U_D^{PT}(x, 0)$ is concave and maximized at $x = \frac{1}{2}$ or 1.

If (11a) holds, by (7), we have

$$\begin{aligned} U_D^{PT}\big(\frac{1}{2}, 0\big) &= w_D(P_1) + w_D(1 - P_0 - P_1) + \frac{1}{2}G \\ &\ge \frac{1}{2} w_D(P_1) + w_D(1 - P_0 - P_1) + G \\ &= U_D^{PT}(1, 0). \end{aligned} \tag{19}$$

Thus, (9) holds for $(x^*, y^*) = (1/2, 0)$, which is an NE of the game $\mathbb{G}$. Similarly, we can prove $(1, 0)$ and $(1, 1)$ are also NEs of the game. ■

Under a low attack cost, as shown in (11b) and (12b), the APT attack is launched immediately and the scan interval is maximized to save energy, as shown in (12a). Otherwise, if the attack cost is high, as shown in (13), a subjective APT attacker has no motivation to launch attacks against the storage device. As a concrete example, we evaluate the performance of the storage defense game $\mathbb{G}$ with $C = 0.62$, $G = 0.6$,

$P_0 = 0.46$, $P_1 = 0.5$, $\alpha_D = 1$ and $L = 2$. As shown in Fig. 2, the attack rate has a sharp increase from 0 to 0.7, as the attacker's objective weight $\alpha_A$ changes at around 0.42, because a subjective APT attacker tends to overweigh his or her attack cost. The objective weight of attacker $\alpha_A = 0.42$ is a turning point from Condition $I_3$ to $I_2$, i.e., the utility of the defender decreases sharply from 1.6 to 0.89.

Next, we consider the storage defense game with a single storage device and three non-zero attack duration levels, i.e., the distribution of the attack duration follows $[P_0, P_1, P_2, 1 - P_0 - P_1 - P_2]$.

*Theorem 2: The subjective storage defense game* $\mathbb{G}$ *with* $S = 1$ *and* $L = 3$ *has an NE* $(x^*, y^*) = (1/3, 0)$*, if*

$$I_4: \begin{cases} G < \min\left\{\frac{3}{2}\exp\left(-(-\ln P_1)^{\alpha_D}\right), \exp\left(-(-\ln P_1)^{\alpha_D}\right) + \frac{1}{2}\exp\left(-(-\ln P_2)^{\alpha_D}\right)\right\} & (20a) \\ C < \exp\left(-(-\ln P_0)^{\alpha_A}\right); & (20b) \end{cases}$$

$(x^*, y^*) = (2/3, 0)$*, if*

$$I_5: \begin{cases} \frac{3}{2}\exp\left(-(-\ln P_1)^{\alpha_D}\right) < G < \frac{1}{2}\exp\left(-(-\ln P_1)^{\alpha_D}\right) + \exp\left(-(-\ln P_2)^{\alpha_D}\right) & (21a) \\ C < \exp\left(-(-\ln P_0)^{\alpha_A}\right) + \frac{1}{2}\exp\left(-(-\ln P_1)^{\alpha_A}\right); & (21b) \end{cases}$$

$(x^*, y^*) = (1, 0)$*, if*

$$I_6: \begin{cases} G > \max\left(\exp\left(-(-\ln P_1)^{\alpha_D}\right) + \frac{1}{2}\exp\left(-(-\ln P_2)^{\alpha_D}\right), \frac{1}{2}\exp\left(-(-\ln P_1)^{\alpha_D}\right) + \exp\left(-(-\ln P_2)^{\alpha_D}\right)\right) & (22a) \\ C < \exp\left(-(-\ln P_0)^{\alpha_A}\right) + \frac{2}{3}\exp\left(-(-\ln P_1)^{\alpha_A}\right) + \frac{1}{3}\exp\left(-(-\ln P_2)^{\alpha_A}\right); & (22b) \end{cases}$$

*and* $(x^*, y^*) = (1, 1)$*, if*

$$I_7: C > \exp\left(-(-\ln P_0)^{\alpha_A}\right) + \frac{2}{3}\exp\left(-(-\ln P_1)^{\alpha_A}\right) + \frac{1}{3}\exp\left(-(-\ln P_2)^{\alpha_A}\right). \quad (23)$$

*Proof:* The proof is given in the appendix. ∎

Under a low attack cost, as shown in (20b), (21b) and (22b), the attacker launches an attack immediately against the storage device and the defender maximizes its detection interval to save energy, as shown in (22a). If the attack cost is high, as in (23), the attacker has no motivation to launch an attack.

Now we consider the case with two storage devices that have the same detection gain, i.e., $G_1 = G_2 = G$, and the same attack duration distribution, i.e., $P_l^1 = P_l^2 = P_l$.

*Theorem 3: If* $S = 2$ *and* $L = 2$*, then the subjective storage defense game* $\mathbb{G}$ *has an NE* $(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{1/2}, \mathbf{0})$*, if*

$$I_8: \begin{cases} G \le \exp\left(-(-\ln P_1)^{\alpha_D}\right) & (24a) \\ \max(C_1, C_2) \le \exp\left(-(-\ln P_0)^{\alpha_A}\right); & (24b) \end{cases}$$

$(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{1}, \mathbf{0})$*, if*

$$I_9: \begin{cases} G > \exp\left(-(-\ln P_1)^{\alpha_D}\right) & (25a) \\ \max(C_1, C_2) \le \exp\left(-(-\ln P_0)^{\alpha_A}\right) + \frac{1}{2}\exp\left(-(-\ln P_1)^{\alpha_A}\right); & (25b) \end{cases}$$

$(\mathbf{x}^*, \mathbf{y}^*) = ([1, \frac{1}{2}], [1, 0])$*, if*

$$I_{10}: \begin{cases} G \le \exp\left(-(-\ln P_1)^{\alpha_D}\right) & (26a) \\ C_1 > \exp\left(-(-\ln P_0)^{\alpha_A}\right) + \frac{1}{2}\exp\left(-(-\ln P_1)^{\alpha_A}\right) & (26b) \\ C_2 \le \exp\left(-(-\ln P_0)^{\alpha_A}\right); & (26c) \end{cases}$$

$(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{1}, [1, 0])$*, if*

$$I_{11}: \begin{cases} G > \exp\left(-(-\ln P_1)^{\alpha_D}\right) & (27a) \\ C_2 < \exp\left(-(-\ln P_0)^{\alpha_A}\right) + \frac{1}{2}\exp\left(-(-\ln P_1)^{\alpha_A}\right) < C_1; & (27b) \end{cases}$$

*and* $(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{1}, \mathbf{1})$*, if*

$$I_{12}: \min(C_1, C_2) > \exp\left(-(-\ln P_0)^{\alpha_A}\right) + \frac{1}{2}\exp\left(-(-\ln P_1)^{\alpha_A}\right). \quad (28)$$

*Proof:* By (8), if $0 \le y_1, y_2 < \frac{1}{2}$, we have

$$\begin{aligned} U_A^{PT}(\mathbf{\frac{1}{2}}, \mathbf{0}) &= -2w_A(P_1) - 2w_A(1 - P_0 - P_1) \\ &\quad - C_1 - C_2 \ge -(2y_1 + 2y_2)\, w_A(P_0) - 2w_A(P_1) \\ &\quad - 2w_A(1 - P_0 - P_1) - C_1 - C_2 = U_A^{PT}(\mathbf{\frac{1}{2}}, \mathbf{y}). \end{aligned} \quad (29)$$

If $0 \le y_2 < \frac{1}{2} \le y_1 \le 1$, and $C_1 < w_A(P_0)$, we have

$$\begin{aligned} U_A^{PT}(\mathbf{\frac{1}{2}}, \mathbf{0}) &= -2w_A(P_1) - 2w_A(1 - P_0 - P_1) - C_1 - C_2 \\ &\ge -(1 + 2y_2)\, w_A(P_0) - 2w_A(P_1) - 2w_A(1 - P_0 - P_1) \\ &\quad - C_2 = U_A^{PT}(\mathbf{\frac{1}{2}}, \mathbf{y}). \end{aligned} \quad (30)$$

Similarly, (10) also holds, if $0 \le y_1 < \frac{1}{2} \le y_2 \le 1$. Thus, (10) holds for $(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{\frac{1}{2}}, \mathbf{0})$.

By (7), if $0 < x_1, x_2 \le \frac{1}{2}$, we have

$$\begin{aligned} U_D^{PT}(\mathbf{\frac{1}{2}}, \mathbf{0}) &= 2w_D(P_1) + 2w_D(1 - P_0 - P_1) + G \\ &\ge 2w_D(P_1) + 2w_D(1 - P_0 - P_1) + (x_1 + x_2)G \\ &= U_D^{PT}(\mathbf{x}, \mathbf{0}). \end{aligned} \quad (31)$$

If $0 \le x_2 < \frac{1}{2} \le x_1 \le 1$, and $G \le w_D(P_1)$, we have

$$\begin{aligned} U_D^{PT}(\mathbf{\frac{1}{2}}, \mathbf{0}) &= 2w_D(P_1) + 2w_D(1 - P_0 - P_1) + G \\ &\ge \left(\frac{1}{2x_1} + 1\right) w_D(P_1) + 2w_D(1 - P_0 - P_1) \\ &\quad + (x_1 + x_2)G = U_D^{PT}(\mathbf{x}, \mathbf{0}). \end{aligned} \quad (32)$$

Similarly, (9) holds for the other cases, indicating that $(\mathbf{\frac{1}{2}}, \mathbf{0})$ is an NE of the game. We can prove the other NEs of the game similarly. ∎

If the attack cost is low, i.e., (24b), the attacker launches APT attacks and the defender scans the two devices at the

same frequency. If (28) holds with a high attack cost, the attack motivation is suppressed and the defender maximizes the scan interval.

*Theorem 4: An NE of the subjective storage defense game $\mathbb{G}$ with S storage devices and L non-zero attack duration levels is given by $(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{1}, \mathbf{1})$, if*

$$C_i > \sum_{l=0}^{L} \frac{L-l}{L} \exp\Big(-\Big(-\ln P_l^i\Big)^{\alpha_A}\Big), \quad \forall 1 \le i \le S. \tag{33}$$

*Proof:* If (33) holds, by (8), $\forall 0 \le y_i \le 1$ we have

$$\begin{aligned} U_A^{PT}(\mathbf{1},\mathbf{1}) &= -\sum_{i=1}^{S}\sum_{l=0}^{L} \exp\Big(-\Big(-\ln P_l^i\Big)^{\alpha_A}\Big) \\ &\ge -\sum_{i=1}^{S}\sum_{l=0}^{L} \exp\Big(-\Big(-\ln P_l^i\Big)^{\alpha_A}\Big) \min\left(\frac{y_i L + l}{L}, 1\right) \\ &\quad -\sum_{i=1}^{S} C_i = U_A^{PT}(\mathbf{1},\mathbf{y}). \end{aligned} \tag{34}$$

Thus, (10) holds for $(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{1}, \mathbf{1})$.

By (7), $\forall 0 < x_i \le 1$, we have

$$\begin{aligned} U_D^{PT}(\mathbf{1},\mathbf{1}) &= \sum_{i=1}^{S}\sum_{l=0}^{L} \exp\Big(-\Big(-\ln P_l^i\Big)^{\alpha_A}\Big) + \sum_{i=1}^{S} G_i \\ &\ge \sum_{i=1}^{S}\sum_{l=0}^{L} \exp\Big(-\Big(-\ln P_l^i\Big)^{\alpha_A}\Big) + \sum_{i=1}^{S} x_i G_i \\ &= U_D^{PT}(\mathbf{x},\mathbf{1}). \end{aligned} \tag{35}$$

Thus, (9) holds for $(\mathbf{x}^*, \mathbf{y}^*) = (\mathbf{1}, \mathbf{1})$, indicating that $(\mathbf{1}, \mathbf{1})$ is an NE of the game. ∎

The subjective attacker has no motivation to launch an attack and the scan interval is maximized if (33) holds for a high attack cost.

## V. Subjective PT-Based Storage Defense Game With Mixed-Strategy

In the subjective cloud storage defense game with mixed-strategy, denoted by $\mathbb{G}'$, the defender quantizes the scan interval into $M$ levels, i.e., $x_i \in \{m/M\}_{1\le m\le M}$, and chooses $x_i$ according to the mixed strategy denoted by $\mathbf{p} = [p_m^i]_{1\le m\le M, 1<i\le S}$, where $p_m^i = \Pr(x_i = m/M)$. The APT attacker quantizes his or her non-zero attack interval into $N$ non-zero levels, i.e., $y_i \in \{n/N\}_{0\le n\le N}$, and determines the mixed strategy denoted by $\mathbf{q} = [q_n^i]_{0\le n\le N, 1<i\le S}$, where $q_n^i = \Pr(y_i = n/N)$. By definition, we have $p_m^i \ge 0, q_n^i \ge 0, \sum_{m=1}^{M} p_m^i = 1$ and $\sum_{n=0}^{N} q_n^i = 1$.

For simplicity, we assume a known and constant time $z_i$ in the mixed-strategy game $\mathbb{G}'$ to focus on the impact of uncertain opponent actions. The expected utilities of the defender and the attacker in the mixed-strategy game $\mathbb{G}'$ are given by (3) and (4) as

$$\begin{aligned} U_D^{EUT}(\mathbf{p},\mathbf{q}) &= \sum_{i=1}^{S}\sum_{m=1}^{M}\sum_{n=0}^{N} p_m^i q_n^i \\ &\quad \times \left(\min\left(\frac{nM + z_i MN}{mN}, 1\right) + \frac{mG_i}{M}\right) \end{aligned} \tag{36}$$

$$\begin{aligned} U_A^{EUT}(\mathbf{p},\mathbf{q}) &= \sum_{i=1}^{S}\sum_{m=1}^{M}\sum_{n=0}^{N} p_m^i q_n^i \\ &\quad \times \left(-\min\left(\frac{nM + z_i MN}{mN}, 1\right) \right. \\ &\quad \left. -\mathrm{I}\left(\frac{n}{N} < \frac{m}{M}\right) C_i\right). \end{aligned} \tag{37}$$

In the PT-based game $\mathbb{G}'$, the subjective defender and the attacker make decisions to maximize their PT-based utilities, given by

$$\begin{aligned} U_D^{PT}(\mathbf{p},\mathbf{q}) &= \sum_{i=1}^{S}\sum_{m=1}^{M}\sum_{n=0}^{N} p_m^i w_D(q_n^i) \\ &\quad \times \left(\min\left(\frac{nM + z_i MN}{mN}, 1\right) + \frac{mG_i}{M}\right) \end{aligned} \tag{38}$$

$$\begin{aligned} U_A^{PT}(\mathbf{p},\mathbf{q}) &= \sum_{i=1}^{S}\sum_{m=1}^{M}\sum_{n=0}^{N} w_A(p_m^i) q_n^i \\ &\quad \times \left(-\min\left(\frac{nM + z_i MN}{mN}, 1\right) \right. \\ &\quad \left. -\mathrm{I}\left(\frac{n}{N} < \frac{m}{M}\right) C_i\right). \end{aligned} \tag{39}$$

By definition, an NE of the PT-based mixed-strategy storage defense game $\mathbb{G}'$, denoted by $(\mathbf{p}^*, \mathbf{q}^*)$ is given by

$$\begin{cases} \mathbf{p}^* = \arg\max_{\mathbf{p}} U_D^{PT}(\mathbf{p}, \mathbf{q}^*) & \text{(40a)} \\ \mathbf{q}^* = \arg\max_{\mathbf{q}} U_A^{PT}(\mathbf{p}^*, \mathbf{q}) & \text{(40b)} \\ \sum_{m=1}^{M} p_m^i = 1, \quad \mathbf{p} \succeq \mathbf{0} & \text{(40c)} \\ \sum_{n=0}^{N} q_n^i = 1, \quad \mathbf{q} \succeq \mathbf{0}, \quad 1 \le i \le S. & \text{(40d)} \end{cases}$$

*Theorem 5: The NE of the subjective storage defense game $\mathbb{G}'$ is given by*

$$\begin{cases} \left[u_D^i(\frac{m}{M}, \frac{n}{N})\right]_{1\le m\le M, 0\le n\le N} \left[w_D\left(q_k^{i*}\right)\right]_{0\leqslant k\leqslant N}^{T} \\ \qquad = \lambda_D^i \mathbf{1_{N+1}} & \text{(41a)} \\ \left[u_A^i(\frac{m}{M}, \frac{n}{N})\right]_{1\le m\le M, 0\le n\le N}^{T} \left[w_A\left(p_k^{i*}\right)\right]_{1\leqslant k\leqslant M}^{T} \\ \qquad = \lambda_A^i \mathbf{1_M} & \text{(41b)} \\ \sum_{m=1}^{M} p_m^{i*} = 1, \quad \mathbf{p} \succeq \mathbf{0}, \quad 1 \le i \le S & \text{(41c)} \\ \sum_{n=0}^{N} q_n^{i*} = 1, \quad \mathbf{q} \succeq \mathbf{0}, \quad 1 \le i \le S & \text{(41d)} \\ \lambda_D^i \ge 0, \quad \lambda_A^i \le 0, & \text{(41e)} \end{cases}$$

*if the solution exists, where $\mathbf{1}_\eta$ is the $\eta$-dimensional all-1 column vector, and*

$$u_D^i(\frac{m}{M}, \frac{n}{N}) = \min\left(\frac{nM + z_i MN}{mN}, 1\right) + \frac{mG_i}{M} \tag{42}$$

$$u_A^i(\frac{m}{M}, \frac{n}{N}) = -\min\left(\frac{nM + z_i MN}{mN}, 1\right) - \mathrm{I}\left(\frac{\mathrm{n}}{\mathrm{N}} < \frac{\mathrm{m}}{\mathrm{M}}\right) \mathrm{C_i}. \tag{43}$$

*Proof:* The Karush-Kuhn-Tucker (KKT) conditions of (40) are given by

$$\begin{cases} L_D = U_D^{PT}\left(\mathbf{p}, \mathbf{q}^*\right) - \varphi\left(\sum_{m=1}^{M} p_m^i - 1\right) + \sum_{m=1}^{M} \mu_m^i p_m^i \\ \frac{\partial L_D}{\partial p_m^i} = 0 \\ -p_m^i \le 0, \mu_m^i \ge 0, \mu_m^i p_m^i = 0, 1 \le m \le M \\ \sum_{m=1}^{M} p_m^i - 1 = 0. \end{cases} \tag{44}$$

According to (38), we apply the complementary slackness for (44) to obtain

$$\begin{cases} \sum_{n=0}^{N} u_D^i\left(\frac{k}{M}, \frac{n}{N}\right) w_D(q_n^{i*}) - \lambda_D^i = 0, 1 \le k \le M \\ \sum_{m=1}^{M} p_m^i = 1 \\ \lambda_D^i \ge 0, \end{cases} \tag{45}$$

and yield (41a). Similarly, we have (41b). ■

*Corollary 1: If* $S = 1$, $M = 2$, $N = 1$,

$$\frac{u_D(\frac{1}{2}, 0) - u_D(1, 0)}{u_D(1, 1) - u_D(\frac{1}{2}, 1)} > 1, \tag{46}$$

$$\frac{u_A(\frac{1}{2}, 1) - u_A(\frac{1}{2}, 0)}{u_A(1, 0) - u_A(1, 1)} > 1, \tag{47}$$

*the subjective storage defense game* $\mathbb{G}'$ *has a unique NE given by*

$$\ln\left(\frac{u_A(\frac{1}{2}, 1) - u_A(\frac{1}{2}, 0)}{u_A(1, 0) - u_A(1, 1)}\right) + \left(-\ln\left(1 - p_1^*\right)\right)^{\alpha_A} - \left(-\ln\left(p_1^*\right)\right)^{\alpha_A} = 0 \tag{48}$$

$$\ln\left(\frac{u_D(\frac{1}{2}, 0) - u_D(1, 0)}{u_D(1, 1) - u_D(\frac{1}{2}, 1)}\right) + \left(-\ln\left(1 - q_0^*\right)\right)^{\alpha_D} - \left(-\ln\left(q_0^*\right)\right)^{\alpha_D} = 0. \tag{49}$$

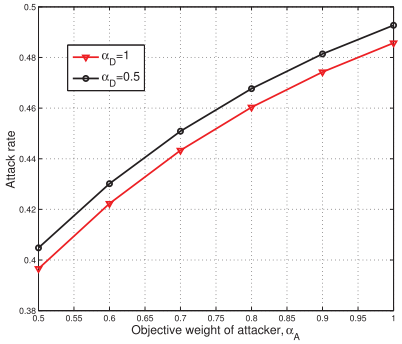
*Proof:* According to (1), (41a) and (46), we have (49). Similarly, we can obtain (48) by (1), (41b) and (47). Next, we prove the uniqueness of $q_0^*$. As $f(x) = (-\ln(x))^{\alpha}$ monotonically decreases with $x$, by (46) and (49) we have $f(q_0^*) > f(1 - q_0^*)$, yielding $0 < q_0^* < 1 - q_0^* < 1$. Thus we have $0 < q_0^* < 1/2$. If $0 < x < 1/2$, we have

$$\frac{d\big(f(1-x) - f(x)\big)}{dx} = f'(1-x) - f'(x) > 0, \tag{50}$$

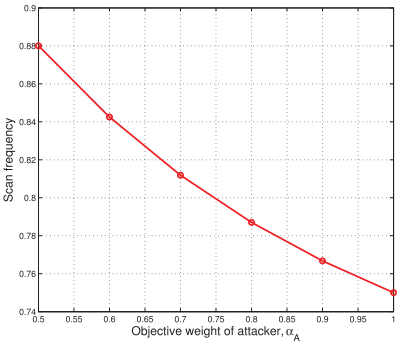
indicating that $f(1-x) - f(x)$ increases with $x$. Therefore, (49) has a unique solution. Similarly, (48) has a unique solution. ■

According to Corollary 1, the NE of the EUT-based storage defense game is given by

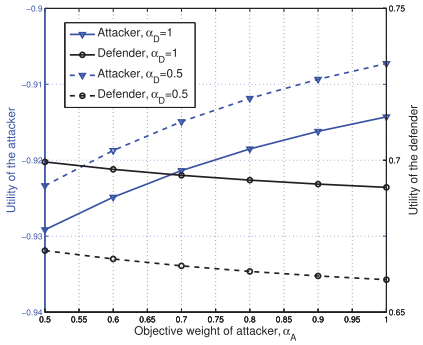
$$(p_1^*, q_0^*) = \left(\frac{1 - z - C}{\min(2z, 1) - z}, \frac{G}{2\min(2z, 1) - 2z}\right). \tag{51}$$




(a) Attack rate




(b) Scan frequency, $p_1^*$, with $\alpha_D = 1$




(c) Utility

Fig. 3. Performance of the static subjective storage defense game with mixed-strategy $\mathbb{G}'$ at the NE, with $C = 0.5$, $G = 0.1$ and $z = 0.2$.

As shown in Fig. 3, the attack rate of the subjective storage defense game $\mathbb{G}'$ decreases with $\alpha_D$, e.g., it decreases by 1.04%, as $\alpha_D$ changes from 0.5 to 1, because a subjective defender scans less frequently. Consequently, the utility of the defender increases from 0.66 to 0.69, if $\alpha_D$ changes from 0.5 to 1. In addition, the scan frequency decreases by 15% as $\alpha_A$ changes from 0.5 to 1, and thus the utility of the defender decreases from 0.7 to 0.69.

## VI. Dynamic PT-Based Storage Defense Game

If the defender is unaware of the APT attack model and the subjective view model in the dynamic subjective cloud storage defense game, the storage defender can apply the Q-learning technique, a model-free and widely-used reinforcement learning technique, to derive an optimal action-selection

**Algorithm 1** APT Defense in a Dynamic Game With Pure-Strategy

Initialize $\gamma = 0.7$, $\delta = 0.7$, $y^0$, $z^0$, $Q(s,x) = 0$, $V(s) = 0, \forall x, s$.
For $k = 1, 2, 3, ...$
  $s^k = y^{k-1} + z^{k-1}$
  Choose $x^k$ via (54)
  Scan the storage device after time $x^k$
  Observe $u_D$ and $y^k + z^k$
  Update $Q(s^k, x^k)$ via (52)
  Update $V(s^k)$ via (53)
End for

**Algorithm 2** APT Defense in a Dynamic Game With Mixed-Strategy

Initialize $\gamma = 0.7$, $\delta = 0.7$, $\Phi^0$, $Q(\mathbf{s}, \mathbf{p}) = 0$, $V(\mathbf{s}) = 0, \forall \mathbf{s}, \mathbf{p}$.
For $k = 1, 2, 3, ...$
  Update $\mathbf{s}^k = \Phi^{k-1}$
  Choose $\mathbf{p}^k$ with the $\epsilon$-greedy algorithm
  Scan the storage device according to strategy $\mathbf{p}^k$
  Observe $u_D$ and $\Phi^k$
  Update $Q(\mathbf{s}^k, \mathbf{p}^k)$ via (55)
  Update $V(\mathbf{s}^k)$ via (56)
End for

policy in the Markov decision process. The Q-learning based defense strategy updates the quality function, denoted by $Q(s,x)$, which is the expected long-term discounted reward with action $x$ at system state $s$, which consists of the action of the opponent and the parameters of the environment.

### A. Dynamic Storage Defense Game with Pure-Strategy

The dynamic PT-based game with pure-strategy, denoted by $\mathcal{G}$, consists of a storage defender and a subjective APT attacker under uncertain attack duration against a storage device.

The system state observed at time $k$ is the total attack duration in the last slot $z^{k-1} + y^{k-1}$. The value function $V(s)$ provides the maximum expected reward of the defender at system state $s$. The defender updates the $Q$-function based on the immediate utility $u_D$ and the value function as follows:

$$Q\left(s^k, x^k\right) \leftarrow (1-\gamma) Q\left(s^k, x^k\right) + \gamma\left(u_D\left(s^k, x^k\right) + \delta V\left(s^{k+1}\right)\right) \quad (52)$$

$$V\left(s^k\right) = \max_{x \in \mathbf{x}} Q\left(s^k, x\right), \quad (53)$$

where $\delta \in [0,1]$ is the discount factor regarding the future reward, and $\gamma \in (0,1]$ is the learning rate of the current experience.

By applying the $\epsilon$-greedy policy, the defender chooses its scan interval $x^k$ to maximize its current $Q$-function as

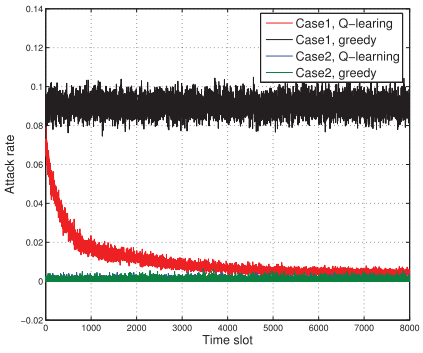
$$\Pr(x^k = \tilde{x}) = \begin{cases} 1-\epsilon, & \tilde{x} = \arg\max_{\mathbf{x}} Q(s^k, x) \\ \dfrac{\epsilon}{M-1}, & \text{o.w.} \end{cases} \quad (54)$$

The Q-learning based storage defense algorithm is summarized in Algorithm 1.

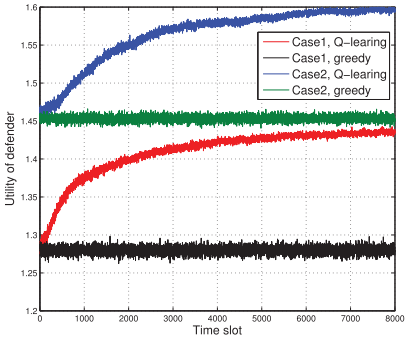
### B. PT-Based Dynamic Game With Mixed-Strategy

In the dynamic PT-based storage defense game with mixed-strategy, denoted by $\mathcal{G}'$, the defender chooses the detection interval distribution $\mathbf{p} = [p_m]_{1 \le m \le M}$, while the attacker determines the attack interval distribution $\mathbf{q} = [q_n]_{0 \le n \le N}$, where $p_m$ and $q_n$ are quantized into $\zeta$ levels, with $1 \le m \le M$, and $0 \le n \le N$.

The system state at time $k$ is defined as the total attack duration distribution in the last time slot, denoted by $\Phi^{k-1}$.




(a) Attack rate




(b) Utility of the defender

Fig. 4. Performance of the dynamic storage defense game with pure-strategy $\mathcal{G}$ averaged over 1000 runs, with $L = 5$, $C = 0.4$ and $\alpha_A = 0.8$ in Case 1, and $L = 2$, $C = 0.62$ and $\alpha_A = 0.3$ in Case 2.

Let $Q(\mathbf{s}, \mathbf{p})$ denote the Q-function with mixed-strategy $\mathbf{p}$, and $V(\mathbf{s})$ be the value function. Based on the iterative Bellman equation, the Q-function can be updated with

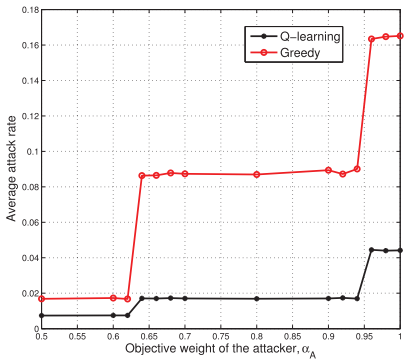
$$Q\left(\mathbf{s}^k, \mathbf{p}^k\right) \leftarrow (1-\gamma) Q\left(\mathbf{s}^k, \mathbf{p}^k\right) + \gamma\left(u_D + \delta V\left(\mathbf{s}^{k+1}\right)\right) \quad (55)$$

$$V\left(\mathbf{s}^k\right) = \max_{\mathbf{p}} Q\left(\mathbf{s}^k, \mathbf{p}\right). \quad (56)$$

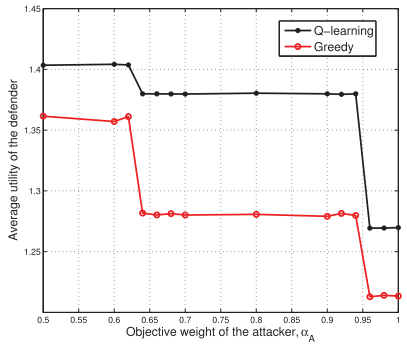
The mixed-strategy is chosen based on the $\epsilon$-greedy algorithm in terms of the Q-function in (55). The algorithm is summarized in Algorithm 2.

## VII. Simulation Results

Simulations have been performed to evaluate the performance of the Q-learning based APT detection scheme in the

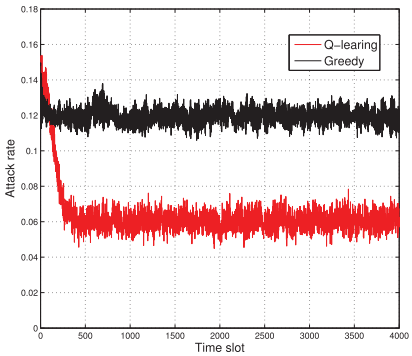



(a) Average attack rate

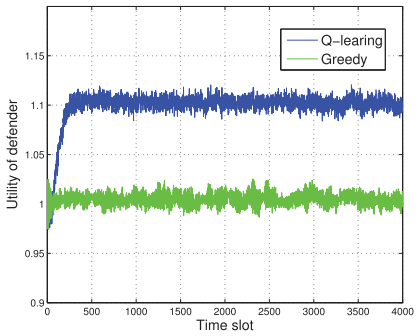



(b) Average utility of the defender

Fig. 5. Performance of the dynamic storage defense game with pure-strategy $\mathcal{G}$ averaged over 4000 time slots, with $L = 5$, $C = 0.4$, $G = 0.6$ and $\epsilon = 0.1$.




(a) Attack rate



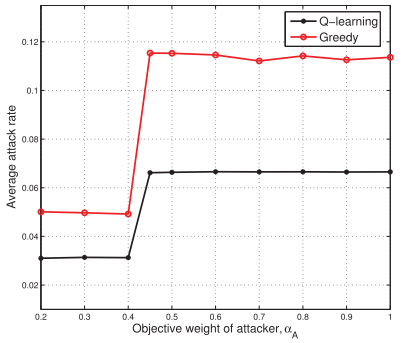

(b) Utility of the defender

Fig. 6. Performance of the dynamic storage defense game with mixed-strategy $\mathcal{G}'$ averaged over 1000 runs, with $z = 0.3$, $C = 0.6$, $G = 0.25$, $M = 2$, $N = 1$, $\alpha_D = 1$, $\alpha_A = 0.8$ and $\epsilon = 0.1$.

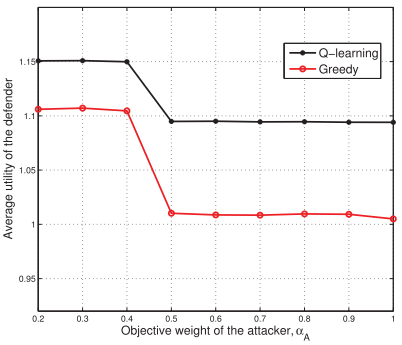
PT-based dynamic games $\mathcal{G}$ and $\mathcal{G}'$. If not specified otherwise, we set $\alpha_D = 1$ to maximize the utility of the defender, $\alpha_A = 0.8$ to represent a typical subjective attacker, $\gamma = 0.7$, $\delta = 0.7$, and $\epsilon = 0.1$ to achieve good performance. We chose typical attack and defense parameters, $G = 0.6$ and $z = 0.3$, and used a greedy detection strategy as a benchmark, in which the scan interval is chosen to maximize the estimated immediate utility based on the previous attack interval. The attack strategy is chosen to maximize the PT utility of the attacker according to the attack history in the last time slot.

The performance of the PT-based dynamic game with pure-strategy $\mathcal{G}$ is shown in Fig. 4, with $L = 5$, $C = 0.4$ and $\alpha_A = 0.8$ in Case 1, and $L = 2$, $C = 0.62$ and $\alpha_A = 0.3$ in Case 2. The attack rate decreases over time, from 8% at the beginning of the game to 0.5% after 6000 time slots in Case 1, about 93.7% lower than the convergent attack rate of the benchmark strategy. Consequently, as shown in Fig. 4 (b), the utility of the defender increases over time from 1.29 at the beginning to 1.43 after 6000 time slots at convergence, about 10.9% higher than the benchmark strategy. In Case 2, the utility of the defender converges to 1.6 after 8000 times slots, which matches the result of the NE given by Theorem 1.

As shown in Fig. 5, the attack rate increases with the objective weight of the attacker, e.g., $R$ increases about four times if $\alpha_A$ changes from 0.5 to 1. The attack rate at convergence is 4.4%, which is 70.6% lower than the benchmark strategy, with $\alpha_A = 1$. Consequently, the utility of the defender decreases from 1.39 to 1.27, if $\alpha_A$ changes from 0.5 to 1. The utility




(a) Average attack rate




(b) Average utility of the defender

Fig. 7. Performance of the dynamic storage defense game with mixed-strategy $\mathcal{G}'$ averaged over 2500 time slots, with $z = 0.3$, $C = 0.6$, $G = 0.25$, $M = 2$, $N = 1$ and $\epsilon = 0.1$.

changes most significantly if $\alpha_A$ changes between 0.7 and 0.9, because the attack interval changes most significantly due to the probability distortion of the subjective attacker, and the

turning points at $\alpha_A = 0.64$ and 0.94 match the results in Theorem 1.

The performance of the PT-based mixed-strategy game $\mathcal{G}'$ in Fig. 6 shows that the attack rate decreases with time from 15% at the beginning of the game to 6% after 500 time slots, which is only half of that of the benchmark strategy if $\alpha_A = 1$. Thus, the utility of the defender increases over time, e.g., from 1 to around 1.1 after 500 time slots, and is 10% higher than that of the benchmark strategy at time slot 500.

As shown in Fig. 7, the attack rate $R$ increases from 3% to 6.5% , if $\alpha_A$ changes from 0.2 to 1, and the attack rate is only half of that of the benchmark strategy at $\alpha_A = 1$. Consequently, the utility of the defender decreases from 1.15 to 1.095, if $\alpha_A$ changes from 0.2 to 1.

## VIII. Conclusion

In this work, we have formulated PT-based cloud storage defense games to investigate the impact of the subjective view of APT attackers under uncertain attack durations in the pure-strategy game or uncertain scan interval of the defender in the mixed-strategy game. The NEs of the PT-based games have been provided, showing that a subjective attacker tends to overweight his or her attack cost and thus increases the scan interval, yielding a higher utility of the defender. A Q-learning based APT resistance scheme has been proposed to improve the performance of the dynamic storage defense game, e.g., in our simulation examples, the attack rate decreases by 50% and the utility of the defender increases by 10% against subjective APT attackers compared with the benchmark greedy strategy.

## Appendix
## Proof of Theorem 2

By (8), if $0 \le y < 1/3$, we have

$$\begin{aligned} U_A^{PT}\left(\frac{1}{3},0\right) &= -w_A(P_1) - w_A(P_2) \\ &\quad - w_A(1-P_0-P_1-P_2) - C \ge -3y w_A(P_0) - w_A(P_1) \\ &\quad - w_A(P_2) - w_A(1-P_0-P_1-P_2) - C \\ &= U_A^{PT}\left(\frac{1}{3}, y\right). \end{aligned} \tag{57}$$

If (20b) holds, by (8), $\forall 1/3 \le y \le 1$, we have

$$\begin{aligned} U_A^{PT}\left(\frac{1}{3},0\right) &= -w_A(P_1) - w_A(P_2) \\ &\quad - w_A(1-P_0-P_1-P_2) - C \ge -w_A(P_0) \\ &\quad - w_A(P_1) \\ &\quad - w_A(P_2) - w_A(1-P_0-P_1-P_2) \\ &= U_A^{PT}\left(\frac{1}{3}, y\right). \end{aligned} \tag{58}$$

Thus, (10) holds for $(x^*, y^*) = (1/3, 0)$.

By (7), if $0 < x < 1/3$, we see that $U_D^{PT}(x,0)$ increases linearly with $x$ and is maximized at $1/3$.

According to (7), if $1/3 < x < 2/3$, we have

$$\begin{aligned} U_D^{PT}(x,0) &= \frac{1}{3x} w_D(P_1) + w_D(P_2) \\ &\quad + w_D(1-P_0-P_1-P_2) + xG, \end{aligned} \tag{59}$$

and

$$\left.\frac{\partial^2 U_D^{PT}}{\partial x^2}\right|_{y=0} = \frac{2}{3x^3} w_D(P_1) \ge 0, \tag{60}$$

indicating that $U_D^{PT}(x,0)$ is concave and is maximized at $1/3$ or $2/3$.

Similarly, if $2/3 < x < 1$, we see that $U_D^{PT}(x,0)$ is concave and is maximized at $2/3$ or 1.

By (7), if (20a) holds, we have

$$\begin{aligned} U_D^{PT}(1/3,0) &= w_D(P_1) + w_D(P_2) \\ &\quad + w_D(1-P_0-P_1-P_2) + \frac{1}{3}G \\ &\ge \max\left\{\frac{1}{3}w_D(P_1) + \frac{2}{3}w_D(P_2) + w_D(1-P_0-P_1-P_2) + G, \right. \\ &\quad \left. \frac{1}{2}w_D(P_1) + w_D(P_2) + w_D(1-P_0-P_1-P_2) + \frac{2}{3}G\right\} \\ &= \max\left\{U_D^{PT}(1,0), U_D^{PT}(2/3,0)\right\}. \end{aligned} \tag{61}$$

Thus, (9) holds for $(x^*, y^*) = (1/3, 0)$. Similarly, we can prove the other NEs in the subjective APT game.

## References


[1] R. B. Sagi. (May 2014). *The Economic Impact of Advanced persistent Threats, IBM Research Intelligence*. [Online]. Available: http://www-01.ibm.com

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

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

[4] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," *Econ., J. Econ. Soc.*, vol. 47, no. 2, pp. 263–291, 1979.

[5] A. Tversky and D. Kahneman, "Advances in prospect theory: Cumulative representation of uncertainty," *J. Risk Uncertainty*, vol. 5, no. 4, pp. 297–323, 1992.

[6] W. Saad, A. L. Glass, N. B. Mandayam, and H. V. Poor, "Toward a consumer-centric grid: A behavioral perspective," *Proc. IEEE*, vol. 104, no. 4, pp. 865–882, Apr. 2016.

[7] S. Gao, E. Frejinger, and M. Ben-Akiva, "Adaptive route choices in risky traffic networks: A prospect theory approach," *Transp. Res. C, Emerg. Technol.*, vol. 18, no. 5, pp. 727–740, Oct. 2010.

[8] G. W. Harrison and E. E. Rutström, "Expected utility theory and prospect theory: One wedding and a decent funeral," *Experim. Econ.*, vol. 12, no. 2, pp. 133–158, Jun. 2009.

[9] T. Li and N. B. Mandayam, "Prospects in a wireless random access game," in *Proc. 46th Annu. Conf. Inf. Sci. Syst.*, Mar. 2012, pp. 1–6.

[10] T. Li and N. B. Mandayam, "When users interfere with protocols: Prospect theory in wireless networks using random access and data pricing as an example," *IEEE Trans. Wireless Commun.*, vol. 13, no. 4, pp. 1888–1907, Apr. 2014.

[11] J. Yu, M. H. Cheung, and J. Huang, "Spectrum investment with uncertainty based on prospect theory," in *Proc. IEEE Int. Conf. Commun.*, Sydney, NSW, Australia, Jun. 2014, pp. 1620–1625.

[12] L. Xiao, J. Liu, Y. Li, N. B. Mandayam, and H. V. Poor, "Prospect theoretic analysis of anti-jamming communications in cognitive radio networks," in *Proc. IEEE Global Commun. Conf.*, Austin, TX, USA, Dec. 2014, pp. 746–751.

[13] Y. Yang and N. B. Mandayam, "Impact of end-user decisions on pricing in wireless networks," in *Proc. IEEE 48th Annu. Conf. Inf. Sci. Syst. (CISS)*, Princeton, NJ, USA, Mar. 2014, pp. 1–6.

[14] Y. Yang and N. B. Mandayam, "Impact of end-user decisions on pricing in wireless networks under a multiple-user-single-provider setting," in *Proc. 52nd Annu. Allerton Conf. Commun. Control Comput.*, Monticello, IL, USA, Sep. 2014, pp. 206–212.

[15] Y. Yang, L. T. Park, N. B. Mandayam, I. Seskar, A. L. Glass, and N. Sinha, "Prospect pricing in cognitive radio networks," *IEEE Trans. Cogn. Commun. Netw.*, vol. 1, no. 1, pp. 56–70, Mar. 2015.

[16] L. Xiao, N. B. Mandayam, and H. V. Poor, "Prospect theoretic analysis of energy exchange among microgrids," *IEEE Trans. Smart Grid*, vol. 6, no. 1, pp. 63–72, Jan. 2015.
[17] Y. Wang, W. Saad, N. B. Mandayam, and H. V. Poor, "Integrating energy storage into the smart grid: A prospect theoretic approach," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process.*, Florence, Italy, May 2014, pp. 7779–7783.
[18] D. Prelec, "The probability weighting function," *Econometrica*, vol. 66, no. 3, pp. 497–527, May 1998.
[19] M. Zhang, Z. Zheng, and N. B. Shroff, "Stealthy attacks and observable defenses: A game theoretic model under strict resource constraints," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Atlanta, GA, USA, Dec. 2014, pp. 813–817.
[20] J. Pawlick, S. Farhang, and Q. Zhu, "Flip the cloud: Cyber-physical signaling games in the presence of advanced persistent threats," in *Decision Game Theory for Security*, vol. 9406. Berlin, Germany: Springer, Nov. 2015, pp. 289–308.
[21] M. Zhang, Z. Zheng, and N. B. Shroff, "A game theoretic model for defending against stealthy attacks with limited resources," in *Decision Game Theory for Security*, vol. 940. Berlin, Germany: Springer, Nov. 2015, pp. 93–112.
[22] P. Hu, H. Li, H. Fu, D. Cansever, and P. Mohapatra, "Dynamic defense strategy against advanced persistent threat with insiders," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Hong Kong, China, May 2015, pp. 747–755.
[23] 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.
[24] Y. Han, T. Alpcan, J. Chan, C. Leckie, and B. I. P. Rubinstein, "A game theoretical approach to defend against co-resident attacks in cloud computing: Preventing co-residence using semi-supervised learning," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 3, pp. 556–570, Mar. 2016.
[25] D. Xu, Y. Li, L. Xiao, N. B. Mandayam, and H. V. Poor, "Prospect theoretic study of cloud storage defense against advanced persistent threats," in *Proc. IEEE Global Commun. Conf.*, Washington, DC, USA, Dec. 2016, pp. 1–6.

**Liang Xiao** (M'09–SM'13) received the B.S. degree in communication engineering from the Nanjing University of Posts and Telecommunications, Nanjing, China, in 2000, the M.S. degree in electrical engineering from Tsinghua University, Beijing, China, in 2003, and the Ph.D. degree in electrical engineering from Rutgers University, New Brunswick, NJ, USA, in 2009. She is currently a Professor with the Department of Communication Engineering, Xiamen University, Xiamen, China. Her current research interests include smart grids, network security, and wireless communications.

**Dongjin Xu** received the B.S. degree in communication engineering from Xiamen University, Xiamen, China, in 2016, where she is currently pursuing the M.S. degree with the Department of Communication Engineering. Her research interests include network security and wireless communications.

**Caixia Xie** received the B.S. degree in communication engineering from Xiamen University, Xiamen, China, in 2015, where she is currently pursuing the M.S. degree with the Department of Communication Engineering. Her research interests include network security and wireless communications.

**Narayan B. Mandayam** (S'89–M'94–SM'99–F'09) received the B.Tech. degree (Hons.) from IIT Kharagpur, in 1989, and the M.S. and Ph.D. degrees from Rice University TX, USA, in 1991 and 1994, respectively, all in electrical engineering.

He was a Visiting Faculty Fellow with the Department of Electrical Engineering, Princeton University, in 2002, and a Visiting Faculty member with the Indian Institute of Science, Bengaluru, India, in 2003. Using constructs from game theory, communications, and networking, his work has focused on system modeling, information processing, and resource management for enabling cognitive wireless technologies to support various applications. Since 1994, he has been with Rutgers University, New Brunswick, NJ, USA, where he is currently a Distinguished Professor and the Chair of the Electrical and Computer Engineering Department. He also serves as an Associate Director with WINLAB. He is currently involved in the use of prospect theory in understanding the psychophysics of data pricing for wireless networks and the smart grid. His recent interests also include privacy in IoT, and modeling and analysis of trustworthy knowledge creation on the Internet. He has co-authored the book *Principles of Cognitive Radio* (Cambridge University Press, 2012) and *Wireless Networks: Multiuser Detection in Cross-Layer Design* (Springer, 2004).

Dr. Mandayam was a co-recipient of the 2015 IEEE Communications Society Advances in Communications Award for his seminal work on power control and pricing, the 2014 IEEE Donald G. Fink Award for his IEEE Proceedings paper "Frontiers of Wireless and Mobile Communications," and the 2009 Fred W. Ellersick Prize from the IEEE Communications Society for his work on dynamic spectrum access models and spectrum policy. He was a recipient of the Peter D. Cherasia Faculty Scholar Award from Rutgers University in 2010, the National Science Foundation CAREER Award in 1998, and the Institute Silver Medal from IIT in 1989. He has served as an Editor of journals, including the IEEE COMMUNICATION LETTERS and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS. He has also served as a Guest Editor of the IEEE JSAC special issues on adaptive, spectrum agile and cognitive radio networks in 2007 and game theory in Communication Systems in 2008. He is currently a Distinguished Lecturer of the IEEE.

**H. Vincent Poor** (S'72–M'77–SM'82–F'87) received the Ph.D. degree in EECS from Princeton University in 1977. From 1977 to 1990, he was on the faculty of the University of Illinois at Urbana–Champaign. Since 1990, he has been on the faculty at Princeton University, where he is currently the Michael Henry Strater University Professor of Electrical Engineering. From 2006 to 2016, he served as the Dean of Princeton's School of Engineering and Applied Science. His research interests are in the areas of information theory, statistical signal processing and stochastic analysis, and their applications in wireless networks and related fields. Among his publications in these areas is the book *Mechanisms and Games for Dynamic Spectrum Allocation* (Cambridge University Press, 2014).

Dr. Poor is a member of the National Academy of Engineering and the National Academy of Sciences, and is a foreign member of the Royal Society. He is also a fellow of the American Academy of Arts and Sciences, the National Academy of Inventors, and other national and international academies. He received the Marconi and Armstrong Awards of the IEEE Communications Society in 2007 and 2009, respectively. Recent recognition of his work includes the 2016 John Fritz Medal, the 2017 IEEE Alexander Graham Bell Medal, a Doctor of Science *honoris causa* from Syracuse University in 2017, and Honorary Professorships at Peking University and Tsinghua University, conferred in 2016.