# Adversarial Attacks on Continuous Authentication Security: A Dynamic Game Approach[⋆]

Serkan Sarıtaş[1][0000−0001−5638−3213], Ezzeldin Shereen[1][0000−0002−9988−9545], Henrik Sandberg[2][0000−0003−1835−2963], and György Dán[1][0000−0002−4876−0223]

[1] Division of Network and Systems Engineering
KTH Royal Institute of Technology, SE-10044, Stockholm, Sweden
{saritas,eshereen,gyuri}@kth.se
[2] Division of Decision and Control Systems
KTH Royal Institute of Technology, SE-10044, Stockholm, Sweden
hsan@kth.se

**Abstract.** Identity theft through phishing and session hijacking attacks has become a major attack vector in recent years, and is expected to become more frequent due to the pervasive use of mobile devices. Continuous authentication based on the characterization of user behavior, both in terms of user interaction patterns and usage patterns, is emerging as an effective solution for mitigating identity theft, and could become an important component of defense-in-depth strategies in cyber-physical systems as well. In this paper, the interaction between an attacker and an operator using continuous authentication is modeled as a stochastic game. In the model, the attacker observes and learns the behavioral patterns of an authorized user whom it aims at impersonating, whereas the operator designs the security measures to detect suspicious behavior and to prevent unauthorized access while minimizing the monitoring expenses. It is shown that the optimal attacker strategy exhibits a threshold structure, and consists of observing the user behavior to collect information at the beginning, and then attacking (rather than observing) after gathering enough data. From the operator's side, the optimal design of the security measures is provided. Numerical results are used to illustrate the intrinsic trade-off between monitoring cost and security risk, and show that continuous authentication can be effective in minimizing security risk.

**Keywords:** Continuous authentication · Dynamic stochastic game · Markov decision process.

[⋆] This work was partly funded by the Swedish Civil Contingencies Agency (MSB) through the CERCES project and has received funding from the European Institute of Innovation and Technology (EIT). This body of the European Union receives support from the European Union's Horizon 2020 research and innovation programme.

# 1 Introduction

Online identity theft and session hijacking are widely used for performing cyber-attacks against online payment systems. As tools for performing identity theft and session hijacking are becoming widely available, the incidence of such attacks is expected to rise in the future. Furthermore, with the proliferation of bring your own device (BYOD) policies, identity theft and session hijacking could be an important attack vector in compromising not only online transactions but also critical infrastructures. Addressing these attack is thus crucial in mitigating advanced persistent threats (APT).

Continuous authentication based on behavioral authentication is emerging as a promising technology in detecting identity theft and session hijacking. Continuous authentication typically relies on a machine learning model trained based on recorded user input, e.g., movement patterns of pointing devices, keystroke patterns, transaction characteristics, which is used for detecting anomalous user input in real-time [2, 3]. User input that is classified as anomalous is typically rejected, and may result in the need for user re-authentication. Clearly, a high incidence of false positives is detrimental to the usability of the system, and thus it should be kept low. A lower false positive rate at the same time implies a higher false negative rate, i.e, lower probability of detection. Finding the optimal parameters for continuous authentication is thus a challenging problem, especially if continuous authentication is used in combination with other solutions for incident detection, such as intrusion detection services (IDS).

In this paper we address this problem. We formulate a model of a system that uses an IDS and continuous authentication for mitigating APT. We then formulate the optimization problem faced by the attacker and by the defender as a dynamic leader-follower game. We characterize the optimal attack strategy, and show that it has a threshold structure. We then provide a characterization of the impact of the parameters of continuous authentication and of the IDS on the cost of the defender, so as to facilitate their joint optimization. We provide numerical results to illustrate the attacker strategy and the impact of the defender's strategy on the attacker's expected cost.

The rest of the paper is organized as follows. After presenting the related literature in Section 2, the problem formulation is provided in Section 3. The optimal attack and defense strategies are discussed in Section 4 and Section 5, respectively. In Section 6, we provide numerical examples and comparative analyses. Section 7 concludes the paper.

# 2 Background and Related Work

Continuous authentication has received increasing attention lately both from industry and academia. Authors in [3] demonstrated the use of keystroke dynamics, mouse movements, and application usage for continuously authenticating users on workstations. Their results showed that keystroke dynamics proved to be the best indicator of user identity. Continuous authentication for smartphone users and users of other wearable electronic devices was considered recently in [2], based on behavioral information of touch gestures like pressure, location, and timing. Authors in [8] demonstrated the potential of using other behavioral information like hand movement, orientation and grasp (HMOG) information for continuously authenticating mobile users. Similarily, authors in [7] demonstrated continuous authentication for wearable glasses, such as Google glass. Authors in [4] showed that car owners or office workers could be continuously authenticated by sensors on their seats. Similar ideas have been proposed for military and battlefield applications for continuously authenticating soldiers by their weapons and suits [1].

Related to our work are previous works that used game theoretic approaches for modeling network security problems and for proposing security solutions. Cooperative authentication in Mobile Ad-hoc Networks (MANETs) was considered in [11], where many selfish mobile nodes need to cooperate in authenticating messages from other nodes while not sacrificing their location privacy. In [9], a game was used to model the process of physical layer authentication in wireless networks, where the defender adjusts its detection threshold in hypothesis testing while the attacker adjusts how often it attacks. The problem of secret (password) picking and guessing was modeled in [6] as a game between a defender (the picker) and an attacker (the guesser). Slightly similar to our work is [10], where the authors consider a game between monitoring nodes and monitored nodes in wireless sensor networks, where the monitoring nodes decide the duration of behavioral monitoring, and the monitored nodes decide when to cooperate and when not to cooperate. Nonetheless, to the best of our knowledge, our work is the first to propose a game theoretic approach for secure risk management considering continuous authentication.

## 3 Problem Formulation

We consider a system that consists of an organization that maintains a corporate network (e.g., a critical infrastructure operator), an employee $u$ of the organization that uses resources on the corporate network, and an attacker denoted by $a$. Our focus is on the interaction between the organization and the attacker, which we model as a dynamic discrete stochastic game with imperfect information. Following common practice in game theoretic models of security, we assume that the attacker is aware of the strategy of the defender (operator), while the defender (operator) is not aware of the actions taken by the attacker over time, and hence of the attacker's knowledge. In this section, we first describe the system model, then define the actions of the operator and the attacker.

### 3.1 User Behavior

For ease of exposition, we consider that time is slotted, and use $t$ for indexing time-slots. We focus on a user $u$ that interacts with the operator's resources (e.g., servers, control systems, etc.) through generating data traffic, and focus on one resource ($r$) for ease of exposition. We denote by $\mathit{\Lambda}_u(t)$ the amount of traffic generated by user $u$ in time-slot $t$, and we assume it is Poisson distributed with parameter $\lambda_u$ (this is equivalent to the common assumption that arrivals can be modeled by a Poisson process, with intensity $\lambda_u/\iota$, where $\iota$ is the length of the time-slot). The successful interaction of the user with the resource in a time-slot generates immediate reward $v_r$ for the operator.

### 3.2 Intrusion Detection and Continuous Authentication

We consider that the operator maintains or buys an intrusion detection service (IDS) in its infrastructure. Motivated by state-of-the-art IDSs, we consider that the intrusion detection service detects anomalous behavior in hosts and in the network, detection thus requires attacker activity. A detection by the IDS is followed by an investigation by a security threat analyst, which implies that a potential attacker would be detected and eliminated. We denote by $m$ the per time-slot operation cost of the IDS, which determines its ability to detect an attacker (e.g., $m$ determines the number of security threat analysts that can be hired), as discussed later.

In addition to the IDS, in order to mitigate identity theft, e.g., through session hijacking and remote access tool-kits, the operator uses continuous authentication (also referred to as behavioral authentication) for verifying that the traffic received from user $u$ is indeed generated by user $u$. Behavioral authentication is based on a characterization of the user behavior, e.g., through training a machine learning model. For simplicity, we consider that the user behavior can be described by a Gaussian distribution $\mathcal{B}_u \sim \mathcal{N}(b_u, \sigma_u)$, with mean $b_u$ and variance $\sigma_u$. While this model is admittedly simple, it allows for analytical tractability.

We consider that continuous authentication is used on a per time-slot basis, that is, the user behavior during the time-slot is verified at the end of every time-slot, and a decision is made based on the match between the user behavior model and the actual behavior of the user during the corresponding time-slot. If the user fails the test then the user is blocked from accessing the resources. We assume that for an appropriate cut-off point $c$, the test result is positive if $\mathcal{B}_u > c$, and negative otherwise. Note that even if there is no attacker, the user could be blocked due to a false positive (FP). We denote by $\eta_u$ the false positive rate of the continuous authentication security system. This is equivalent to saying that the system applies a detection threshold of $c = \mathit{\Phi}_u^{-1}(1-\eta_u)$, where $\mathit{\Phi}_u$ is the cumulative distribution function (CDF) of $\mathcal{B}_u$.

Thus, without an attacker, the system $S$ can be in two different states[1]: the blocking state (BL) and the unblocking state (UB). In state BL, the user can not interact with resources and hence cannot generate reward $v_r$, while in state UB, it is authorized to interact with the resources and thus it can generate reward $v_r$. If the user fails continuous authentication in a time-slot, then the state of the system switches from UB to BL. Note that this could happen due

[1] In the case of an attacker, the third state AD (attacker is detected) is introduced in Section 3.4.

to a FP or due to a true positive (TP), i.e., input generated by an attacker as discussed later. Furthermore, to allow productivity, we consider that a user that is blocked in time-slot $t$ is unblocked in time-slot $t+1$ with probability $q$; i.e., $\mathsf{Pr}\big(S(t+1) = \mathrm{UB} \,|\, S(t) = \mathrm{BL}\big) = q$.

The above assumptions imply that if there is no input from the user in time-slot $t$ and the system was in state UB, then it will stay in state UB, as no false alarm is generated in the case of the lack of user activity. Hence, without an attacker, we can model the continuous authentication security system as a discrete time Markov chain with state space $\{\mathrm{UB}, \mathrm{BL}\}$, and the state transition probabilities are

$$\begin{aligned}
\mathsf{Pr}\big(S(t+1) = \mathrm{UB} \,\big|\, S(t) = \mathrm{UB}\big) &= \mathrm{e}^{-\lambda_u} + (1-\mathrm{e}^{-\lambda_u})(1-\eta_u) \triangleq P_{uu}\,, \\
\mathsf{Pr}\big(S(t+1) = \mathrm{BL} \,\big|\, S(t) = \mathrm{UB}\big) &= \eta_u(1-\mathrm{e}^{-\lambda_u}) = 1 - P_{uu}\,, \\
\mathsf{Pr}\big(S(t+1) = \mathrm{UB} \,\big|\, S(t) = \mathrm{BL}\big) &= q\,, \\
\mathsf{Pr}\big(S(t+1) = \mathrm{BL} \,\big|\, S(t) = \mathrm{BL}\big) &= 1 - q\,.
\end{aligned} \tag{1}$$

### 3.3 Attack Model

Motivated by recent security incidents caused by identity theft and session hijacking, we consider an attacker that compromises a system component at cost $C_a$, e.g., the user's computer, which allows it to observe the traffic generated by user $u$ and to craft packets that appear to originate from user $u$. We refer to observing the user traffic as *listening*, and to crafting packets as *attacking* in the following. In addition, the attacker can decide not do anything during a time-slot, which we refer to as *waiting*.

Consequently, in every time-slot, the attacker can choose between three actions: wait ($l(t) = 0$, $a(t) = 0$), listen ($l(t) = 1$, $a(t) = 0$), and attack ($l(t) = 0$, $a(t) = 1$), where $l(t) = 1$ stands for listening and $a(t) = 1$ stands for attacking. The purpose of listening is to collect behavioral information about the user, so as to learn to imitate legitimate user behavior that would pass continuous authentication. The purpose of attacking is to execute a rogue command on the resource, but in order for the attack to be successful, the system has to be in state UB and the attacker generated input should pass continuous authentication. If in time-slot $t$ the attack is successful, then the attacker obtains immediate reward $c_r$, which is a penalty for the defender. Motivated by that many attacks have a monetary reward, we consider that the future reward of the attacker is discounted by a discount factor $\rho$. In what follows we first define the actions of the attacker at time-slot $t$, then we will provide expressions for the attacker's reward in Section 4.2.

**Listening ($\boldsymbol{l(t) = 1}$, $\boldsymbol{a(t) = 0}$).** The attacker observes the behavior of the user during a time-slot in order to learn it and imitate the user for a successful attack. Learning during time-slot $t$ is determined by the traffic $\Lambda_u(t)$ generated by the user and by the learning rate $\gamma$. The total amount of observation of the attacker about the user until time-slot $t$ can be expressed as $L(t) = \sum_{\tau=0}^{t-1} \mathbb{1}_{\{l(\tau)=1\}} \Lambda_u(\tau)$, where $\mathbb{1}_{\{D\}}$ is the indicator function of an event $D$.

At the same time, since listening requires activity from the attacker, the IDS could detect the attacker in the time-slot. We denote by $\delta_l(m)$ the probability that the IDS detects the attacker in a time slot when it is listening. We make the reasonable assumption that $\delta_l(m)$ is a concave function of $m$, $\delta_l(0) = 0$ and $\lim_{m\to\infty} \delta_l(m) = 1$, where $m$ is the per time-slot operation cost of the IDS, as defined previously.

**Attacking ($l(t) = 0$, $a(t) = 1$).** The attacker generates and sends rouge input to the resource, trying to impersonate the legitimate user. How well the attacker can imitate the user depends on the amount of observation $L(t)$ that it has collected about the user. We consider that given $L(t)$ amount of information the attacker can generate input following a Gaussian distribution, $\hat{\mathcal{B}}_u(L(t)) \sim \mathcal{N}(\hat{b}_u(L(t)), \hat{\sigma}_u(L(t)))$, where $\hat{b}_u(L(t)) = b_u(1 + \mathrm{e}^{-\gamma L(t)})$ and $\hat{\sigma}_u(L(t)) = \sigma_u(1 + \mathrm{e}^{-\gamma L(t)})$. Since the user behavior is a Gaussian r.v., $\mathcal{B}_u \sim \mathcal{N}(b_u, \sigma_u)$, and $\hat{\mathcal{B}}_u \sim \mathcal{N}(\hat{b}_u, \hat{\sigma}_u)$ is the random variable generated by the attacker, we can use the binormal method [5] for expressing the Receiver Operating Characteristic (ROC) curve of the continuous authentication security system as

$$\mathrm{ROC}(\eta_u, L(t)) = \Phi(a + b\Phi^{-1}(\eta_u)) \,, \tag{2}$$

where $\eta_u$ is the FP rate, $\Phi(\cdot)$ is the CDF of the standard normal distribution, $a = \frac{\hat{b}_u(L(t)) - b_u}{\hat{\sigma}_u(L(t))}$, and $b = \frac{\sigma_u}{\hat{\sigma}_u(L(t))}$. Note that $\mathrm{ROC}(\eta_u, L(t))$ is the TP rate of the detector (i.e., the conditional probability of classifying rogue input as such).

By inspecting (2), and substituting $\omega = L(t)$, we can observe that

$$\mathrm{ROC}(\eta_u, \omega) = \Phi(a + b\Phi^{-1}(\eta_u)) = \Phi\Bigg( \underbrace{\frac{b_u}{\sigma_u} - \frac{b_u - \sigma_u \Phi^{-1}(\eta_u)}{\sigma_u(1 + \mathrm{e}^{-\gamma\omega})}}_{\triangleq \xi_\omega} \Bigg) = \Phi(\xi_\omega) \,.$$

Normally, one would expect that as the number of observations increases, the attacker can imitate the real behavior of the user more successfully; i.e., its input is harder to distinguish from a real user input. Hence, we can safely assume that $\mathrm{ROC}(\eta_u, \omega) = \Phi(\xi_\omega)$ should be a non-increasing function of $\omega$, or equivalently, it must hold that $b_u \geq \sigma_u \Phi^{-1}(\eta_u)$.

Similar to listening, attacking requires activity from the attacker, and thus the IDS could detect the attacker in the time-slot. We denote by $\delta_a(m)$ the probability that the IDS detects the attacker in a time slot when it is attacking. Similar to $\delta_l(m)$, we assume that $\delta_a(m)$ is concave, $\delta_a(0) = 0$ and $\lim_{m\to\infty} \delta_a(m) = 1$.

**Waiting ($l(t) = 0$, $a(t) = 0$).** If the attacker chooses to wait, it neither learns nor attacks, hence it cannot be detected but cannot learn or obtain a reward either.

### 3.4 Continuous Authentication Game

We can now informally introduce the continuous authentication game. In the game the defender (the operator) is the leader, and chooses a defense strategy $(m, \eta_u)$. The defense strategy is known to the attacker, i.e., the follower, who in turn decides whether or not to invest in compromising the system at cost $C_a$, and if it decides to compromise the system, in every time-slot it decides whether to wait, listen, or attack. The game ends when the attacker is detected (AD) by the IDS, i.e., when $S(t) = \text{AD}$. AD is thus an absorbing state. The attacker is interested in maximizing its utility (reward), while the operator is interested in maximizing its average utility. In what follows we formulate the Markov decision process (MDP) faced by the attacker and the optimization problem faced by the defender.

## 4 Optimal Attack Strategy

We start with describing the state space and the state transitions as a function of the attacker's policy and of the defender's strategy. We then derive the optimal attack policy for given defender strategy.

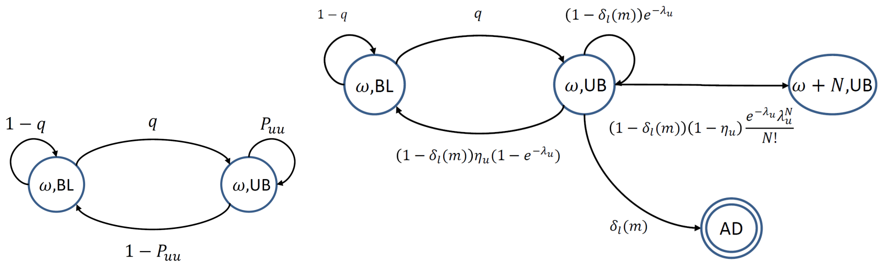
### 4.1 States and Actions

In order to formulate the MDP faced by the attacker, observe that the state of the system from the perspective of the attacker depends on whether the system is blocked ($S(t) = \text{BL}$) or unblocked ($S(t) = \text{UB}$), and on the amount of observations $L(t)$ it has collected so far. Clearly, the state transition probabilities are affected by the actions of the attacker, hence the optimization problem faced by the attacker can be formulated as an MDP. In the following we provide the state transition probabilities, depending on the action chosen by the attacker.

**Waiting ($l(t) = 0$, $a(t) = 0$).** When waiting, the attacker does not observe the user traffic, neither does it attempt to attack, hence the state transition probabilities are determined by the FP rate in state UB, and by unblocking in state BL. As depicted in Fig. 1-(a), the state transition probabilities are thus

$$\mathsf{Pr}\big(S(t+1) = \text{UB}, L(t+1) = \omega \,\big|\, S(t) = \text{UB}, L(t) = \omega, l(t) = 0, a(t) = 0\big) = P_{uu}\,,$$
$$\mathsf{Pr}\big(S(t+1) = \text{BL}, L(t+1) = \omega \,\big|\, S(t) = \text{UB}, L(t) = \omega, l(t) = 0, a(t) = 0\big) = 1 - P_{uu}\,,$$
$$\mathsf{Pr}\big(S(t+1) = \text{UB}, L(t+1) = \omega \,\big|\, S(t) = \text{BL}, L(t) = \omega, l(t) = 0, a(t) = 0\big) = q\,,$$
$$\mathsf{Pr}\big(S(t+1) = \text{BL}, L(t+1) = \omega \,\big|\, S(t) = \text{BL}, L(t) = \omega, l(t) = 0, a(t) = 0\big) = 1 - q\,.$$

Although seemingly unimportant, waiting is preferred by the attacker when the system is in state BL, since the user can not interact with the resources, the attacker cannot increase its total number of observation about the user, and thus listening is not an optimal action for the attacker due to the possibility



(a) State transitions when waiting.

(b) State transitions when listening.

Fig. 1: State transitions and corresponding probabilities when the attacker is (a) waiting and (b) listening.

of being detected. Similarly, when the system is in state BL, an attack would be blocked, but the attacker could be detected. Therefore, when the system is in state BL, the attacker would prefer waiting. Since the attacker is completely passive while waiting, the IDS cannot detect the attacker. Therefore, practically, as depicted in Fig. 1-(a), Fig. 1-(b) and Fig. 2, it is not possible to switch to state AD from state BL.

**Listening ($l(t) = 1$, $a(t) = 0$).** If listening, the attacker can be detected by the IDS with probability $\delta_l(m)$. If the user does not generate traffic (i.e., $N = 0$), then a FP cannot be triggered, but the attacker's amount of observation does not change. On the contrary, if the user generates traffic (i.e., $N \geq 1$), then the attacker can observe and learn, as long as the user generated traffic does not cause a FP. Thus, as depicted in Fig. 1-(b), the transition probabilities are

$$
\begin{aligned}
\mathsf{Pr}\big(S(t+1) = \mathrm{AD} \,\big|\, S(t) = \mathrm{UB}, L(t) = \omega, l(t) = 1, a(t) = 0\big) &= \delta_l(m)\,, \\
\mathsf{Pr}\big(S(t+1) = \mathrm{UB}, L(t+1) = \omega \,\big|\, S(t) = \mathrm{UB}, L(t) = \omega, l(t) = 1, a(t) = 0\big)& \\
= (1 - \delta_l(m))\mathrm{e}^{-\lambda_u}\,,& \\
\mathsf{Pr}\big(S(t+1) = \mathrm{UB}, L(t+1) = \omega + N \,\big|\, S(t) = \mathrm{UB}, L(t) = \omega, l(t) = 1, a(t) = 0\big)& \\
= (1 - \delta_l(m))(1 - \eta_u)\frac{\mathrm{e}^{-\lambda_u}\lambda_u^N}{N!}\,,\ \text{for } N = 1, 2, \ldots\,,& \\
\mathsf{Pr}\big(S(t+1) = \mathrm{BL}, L(t+1) = \omega \,\big|\, S(t) = \mathrm{UB}, L(t) = \omega, l(t) = 1, a(t) = 0\big)& \\
= (1 - \delta_l(m))\eta_u(1 - \mathrm{e}^{-\lambda_u})\,.&
\end{aligned}
$$

**Attacking ($l(t) = 0$, $a(t) = 1$).** If attacking, the attacker can be detected by the IDS with probability $\delta_a(m)$. If the attacker is not detected, then for a successful attack the attacker generated input must pass continuous authentication (false negative) and the user traffic must not cause a FP. If any of these two does not hold, the system switches to state BL. As depicted in Fig. 2, we thus have

$$\begin{aligned}\mathsf{Pr}\big(S(t+1)=\mathrm{AD}\,\big|\,S(t)=\mathrm{UB},L(t)=\omega,l(t)=0,a(t)=1\big)&=\delta_a(m)\,,\\ \mathsf{Pr}\big(S(t+1)=\mathrm{UB},L(t+1)=\omega\,\big|\,S(t)=\mathrm{UB},L(t)=\omega,l(t)=0,a(t)=1\big)&\\ =(1-\delta_a(m))P_{uu}(1-\Phi(\xi_\omega))\,,&\\ \mathsf{Pr}\big(S(t+1)=\mathrm{BL},L(t+1)=\omega\,\big|\,S(t)=\mathrm{UB},L(t)=\omega,l(t)=0,a(t)=1\big)&\\ =(1-\delta_a(m))(1-P_{uu}(1-\Phi(\xi_\omega)))\,.&\end{aligned}$$

### 4.2 Attacker Reward as a Dynamic Programming Recursion

Let the total observation of the attacker about the user at the beginning of the time-slot $t$ be $L(t)=\omega$. Further, let us denote by $J_t(L(t)=\omega,S(t)=\mathrm{UB})$ and $J_t(L(t)=\omega,S(t)=\mathrm{BL})$ the total reward of the attacker starting from the time-slot $t$ when $S(t)=\mathrm{UB}$ and $S(t)=\mathrm{BL}$, respectively. For notational convenience, we will use $J(\omega,\mathrm{UB})$ and $J(\omega,\mathrm{BL})$ henceforth[2]. Clearly, the total reward of the attacker corresponds to $J(0,\mathrm{UB})$. Then, depending on the states, actions and corresponding probabilities described above, and accounting for the discount factor $\rho$, the dynamic programming recursion of the attacker reward can be established as

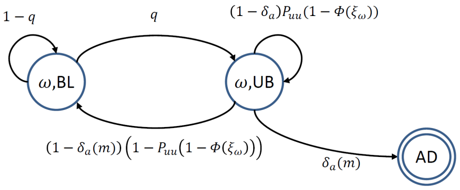

$1-q$
$q$
$(1-\delta_a)P_{uu}(1-\Phi(\xi_\omega))$
$\omega$,BL
$\omega$,UB
$(1-\delta_a(m))\big(1-P_{uu}(1-\Phi(\xi_\omega))\big)$
$\delta_a(m)$
AD

Fig. 2: State transitions and probabilities when the attacker choose to attack.

$$\begin{aligned}&J(\omega,\mathrm{BL})=\rho qJ(\omega,\mathrm{UB})+\rho(1-q)J(\omega,\mathrm{BL})\\ &\Rightarrow J(\omega,\mathrm{BL})=\frac{\rho q}{1-\rho(1-q)}J(\omega,\mathrm{UB})\,,\end{aligned}\tag{3}$$

$$J(\omega,\mathrm{UB})=\begin{cases}\rho P_{uu}J(\omega,\mathrm{UB})+\rho(1-P_{uu})J(\omega,\mathrm{BL}) & l=0\,,\ a=0\\ \rho(1-\delta_l(m))\big(O(\omega)+\eta_u(1-\mathrm{e}^{-\lambda_u})J(\omega,\mathrm{BL})\big) & l=1\,,\ a=0\,,\\ \rho(1-\delta_a(m))A(\omega) & l=0\,,\ a=1\end{cases}\tag{4}$$

where

$$O(\omega)=\mathrm{e}^{-\lambda_u}J(\omega,\mathrm{UB})+(1-\eta_u)\sum_{n=1}^{\infty}P(\Lambda_u=n)J(\omega+\Lambda_u,\mathrm{UB})\,,$$

$$A(\omega)=\underbrace{P_{uu}\big(1-\Phi(\xi_\omega)\big)}_{\triangleq\tilde{q}_\omega}\left(\frac{c_r}{\rho}+J(\omega,\mathrm{UB})\right)+\Big(1-\underbrace{P_{uu}\big(1-\Phi(\xi_\omega)\big)}_{\triangleq\tilde{q}_\omega}\Big)J(\omega,\mathrm{BL})$$

[2] For ease of exposition, $\omega$ denotes $L(t)=\omega$. Notice the time dependency of $\omega$ (even though it is not explicitly stated in the notation).

$$= \widetilde{q}_\omega \frac{c_r}{\rho} + \widetilde{q}_\omega J(\omega, \mathrm{UB}) + (1 - \widetilde{q}_\omega) J(\omega, \mathrm{BL})\,.$$

Note that, since $\mathrm{ROC}(\eta_u, \omega) = \varPhi(\xi_\omega)$ is a non-increasing function of $\omega$, the parameter $\widetilde{q}_\omega$ is a non-decreasing function of $\omega$.

To simplify the expressions and to obtain structural insight, let us substitute (3) into (4), thus for $l = 0$ and $a = 0$ we obtain

$$J(\omega, \mathrm{UB}) = \rho \frac{P_{uu}(1-\rho) + \rho q}{1 - \rho + \rho q} J(\omega, \mathrm{UB})\,. \tag{5}$$

This allows us to formulate the following proposition.

**Proposition 1.** *Let $\rho < 1$. Then waiting cannot be optimal in state $UB$.*

*Proof.* By (5), if waiting is to be optimal in state UB then its reward must be $J(\omega, \mathrm{UB}) = 0$, which cannot be optimal. □

As a consequence, if the system is in state UB then the attacker prefers either listening or attacking during the time-slot. On the contrary, waiting is the optimal action in state BL as discussed in Section 4.1.

Let us now consider that the attacker prefers listening in state UB. We can again substitute (3), to obtain for $l = 1$ and $a = 0$,

$$\begin{aligned}
J(\omega, \mathrm{UB}) &= \rho(1 - \delta_l(m)) \left( \mathrm{e}^{-\lambda_u} J(\omega, \mathrm{UB}) + (1 - \eta_u) \sum_{n=1}^{\infty} P(\varLambda_u = n) J(\omega + \varLambda_u, \mathrm{UB}) \right) \\
&\quad + \rho(1 - \delta_l(m)) \eta_u (1 - \mathrm{e}^{-\lambda_u}) \frac{\rho q}{1 - \rho(1-q)} J(\omega, \mathrm{UB}) \\
&= \underbrace{\rho(1 - \delta_l(m)) \left( \mathrm{e}^{-\lambda_u} + \eta_u (1 - \mathrm{e}^{-\lambda_u}) \frac{\rho q}{1 - \rho + \rho q} \right)}_{\triangleq U} J(\omega, \mathrm{UB}) \\
&\quad + \underbrace{\rho(1 - \delta_l(m))(1 - \eta_u) \sum_{n=1}^{\infty} P(\varLambda_u = n) J(\omega + \varLambda_u, \mathrm{UB})}_{\triangleq K_\omega} \\
&= U J(\omega, \mathrm{UB}) + K_\omega\,.
\end{aligned} \tag{6}$$

Using the same substitution, if the attacker prefers attacking in state UB, i.e., for $l = 0$ and $a = 1$, we obtain

$$\begin{aligned}
J(w, \mathrm{UB}) &= \rho(1 - \delta_a(m)) \left( \widetilde{q}_\omega \frac{c_r}{\rho} + \widetilde{q}_\omega J(\omega, \mathrm{UB}) + (1 - \widetilde{q}_\omega) \frac{\rho q}{1 - \rho(1-q)} J(\omega, \mathrm{UB}) \right) \\
&= \underbrace{(1 - \delta_a(m)) \widetilde{q}_\omega c_r}_{\triangleq C_\omega} + \underbrace{\rho(1 - \delta_a(m)) \frac{\widetilde{q}_\omega (1-\rho) + \rho q}{1 - \rho + \rho q}}_{\triangleq T_\omega} J(\omega, \mathrm{UB}) \\
&= T_\omega J(\omega, \mathrm{UB}) + C_\omega\,.
\end{aligned} \tag{7}$$

Based on (6) and (7), the attacker reward in state UB can be expressed as

$$J_\omega \triangleq J(\omega, \mathrm{UB}) = \begin{Bmatrix} UJ_\omega + K_\omega & l=1\,,\; a=0 \\ T_\omega J_\omega + C_\omega & l=0\,,\; a=1 \end{Bmatrix} = \begin{cases} \frac{K_\omega}{1-U} & l=1\,,\; a=0 \\ \frac{C_\omega}{1-T_\omega} & l=0\,,\; a=1 \end{cases} .$$

Note that since the attacker aims for the maximum reward as $J_\omega = \max\left\{\frac{K_\omega}{1-U}, \frac{C_\omega}{1-T_\omega}\right\}$, her optimal policy must satisfy the following:

$$\begin{cases} J_\omega = \frac{K_\omega}{1-U}\,,\; l=1\,,\; a=0 & \text{if} \quad \frac{K_\omega}{1-U} > \frac{C_\omega}{1-T_\omega} \\ J_\omega = \frac{C_\omega}{1-T_\omega}\,,\; l=0\,,\; a=1 & \text{if} \quad \frac{K_\omega}{1-U} \leq \frac{C_\omega}{1-T_\omega} \end{cases} . \tag{8}$$

Based on the above analysis, we can summarize the parameters of the attacker reward in (8) as follows.

$$\begin{aligned}
K_\omega &= \rho(1-\delta_l(m))(1-\eta_u)\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}J_{\omega+n} \\
U &= \rho(1-\delta_l(m))\left(\mathrm{e}^{-\lambda_u} + \eta_u(1-\mathrm{e}^{-\lambda_u})\frac{\rho q}{1-\rho+\rho q}\right) \\
C_\omega &= (1-\delta_a(m))\widetilde{q}_\omega c_r \\
T_\omega &= \rho(1-\delta_a(m))\frac{\widetilde{q}_\omega(1-\rho)+\rho q}{1-\rho+\rho q} \\
\widetilde{q}_\omega &= P_{uu}\big(1-\mathrm{ROC}(\eta_u,\omega)\big) = P_{uu}\left(1-\Phi(\xi_\omega)\right) \\
P_{uu} &= \mathrm{e}^{-\lambda_u} + (1-\mathrm{e}^{-\lambda_u})(1-\eta_u) = 1-\eta_u+\eta_u\mathrm{e}^{-\lambda_u} \\
\xi_\omega &= \frac{b_u}{\sigma_u} - \frac{b_u-\sigma_u\Phi^{-1}(\eta_u)}{\sigma_u(1+\mathrm{e}^{-\gamma\omega})}
\end{aligned}$$

### 4.3 Listening Reward vs. Attacking Reward

Observe that since $\widetilde{q}_\omega$ is a non-decreasing function of $\omega$, the reward of attacking $\frac{C_\omega}{1-T_\omega}$ is a non-decreasing function of $\omega$; i.e., more observation is always at least as good for the attacker. Thus, it is interesting for the attacker to analyze the advantage of attacking with more observation:

**Proposition 2.** *Let* $\mathcal{L}_\omega \triangleq \frac{\frac{C_{\omega+1}}{1-T_{\omega+1}}}{\frac{C_\omega}{1-T_\omega}}$ *be the ratio between the attacking rewards of two consecutive amounts of observation. Then,* $\mathcal{L}_\omega$ *is a monotonic decreasing function of* $\omega$ *and* $\lim_{\omega\to\infty}\mathcal{L}_\omega = 1$.

*Proof.* $\mathcal{L}_\omega$ can be expanded as follows:

$$\mathcal{L}_\omega = \frac{\frac{(1-\delta_a(m))\widetilde{q}_{\omega+1}c_r}{1-\rho(1-\delta_a(m))\frac{\widetilde{q}_{\omega+1}(1-\rho)+\rho q}{1-\rho+\rho q}}}{\frac{(1-\delta_a(m))\widetilde{q}_\omega c_r}{1-\rho(1-\delta_a(m))\frac{\widetilde{q}_\omega(1-\rho)+\rho q}{1-\rho+\rho q}}} = \frac{\widetilde{q}_{\omega+1}}{\widetilde{q}_\omega}\,\frac{1-\rho(1-\delta_a(m))\frac{\widetilde{q}_\omega(1-\rho)+\rho q}{1-\rho+\rho q}}{1-\rho(1-\delta_a(m))\frac{\widetilde{q}_{\omega+1}(1-\rho)+\rho q}{1-\rho+\rho q}} . \tag{9}$$

Since $\widetilde{q}_\omega$ is a non-decreasing function of $\omega$; i.e., $\widetilde{q}_{\omega+1} \geq \widetilde{q}_\omega$, it can be obtained from (9) that $\mathcal{L}_\omega \geq 1$. Furthermore, since $\lim_{\omega\to\infty} \widetilde{q}_\omega = (1-\eta_u+\eta_u \mathrm{e}^{-\lambda_u})\,(1-\eta_u)$, it holds that $\lim_{\omega\to\infty} \mathcal{L}_\omega = 1$. We note that $\frac{\mathrm{d}\mathcal{L}_\omega}{\mathrm{d}\omega} < 0$ can be proved analytically for a continuous extension of $\mathcal{L}_\omega$; i.e., assuming $\omega \in [0,\infty)$ rather than $\omega \in \{0,1,\ldots\}$. □

In order to compare the listening and attacking rewards for some amount of observation $\omega$, let us define the incremental observation gain

$$\chi_\omega \triangleq \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U}\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}\left(\prod_{i=0}^{n-1}\mathcal{L}_{\omega+i}\right). \tag{10}$$

**Lemma 1.** *$\chi_\omega$ is a decreasing function of $\omega$. Furthermore, $\lim_{\omega\to\infty}\chi_\omega < 1$.*

*Proof.* The first part of the lemma follows from that $\mathcal{L}_\omega$ is a decreasing function of $\omega$. To prove the second part of the lemma, since $\lim_{\omega\to\infty}\mathcal{L}_\omega = 1$, observe that

$$\lim_{\omega\to\infty}\chi_\omega = \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-\rho(1-\delta_l(m))\left(\mathrm{e}^{-\lambda_u}+\eta_u(1-\mathrm{e}^{-\lambda_u})\frac{\rho q}{1-\rho+\rho q}\right)}(1-\mathrm{e}^{-\lambda_u}) < 1\,. \quad \square$$

As a consequence of the above result we can state the following.

**Corollary 1.** *If $\chi_{\omega=0} > 1$ then there exists a critical value $\widetilde{\omega}$ such that $\chi_{\omega=\widetilde{\omega}-1} > 1$ and $\chi_{\omega=\widetilde{\omega}} \leq 1$. Otherwise; i.e., if $\chi_{\omega=0} \leq 1$ then $\widetilde{\omega} = 0$.*

Note that $\widetilde{\omega}$ is independent of time and can be calculated (before the game-play) for a given set of parameters[3]. We are now ready to prove that the attacker policy is indeed a threshold policy.

**Theorem 1.** *The attacker prefers listening over attacking for $\omega < \widetilde{\omega}$.*

*Proof.* In order to compare the listening reward $\frac{K_\omega}{1-U}$ and the attacking reward $\frac{C_\omega}{1-T_\omega}$ observe that

$$\begin{aligned}
\frac{K_\omega}{1-U} &= \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U}\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}J_{\omega+n}\\
&\geq \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U}\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}\frac{C_{\omega+n}}{1-T_{\omega+n}}\\
&= \frac{C_\omega}{1-T_\omega}\frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U}\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}\frac{\frac{C_{\omega+n}}{1-T_{\omega+n}}}{\frac{C_\omega}{1-T_\omega}}\\
&= \frac{C_\omega}{1-T_\omega}\underbrace{\frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U}\sum_{n=1}^{\infty}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}\left(\prod_{i=0}^{n-1}\mathcal{L}_{\omega+i}\right)}_{\chi_\omega}.
\end{aligned} \tag{11}$$

[3] The sum in (10) can be partitioned into $\sum_{n=1}^{\mathcal{C}}$ and $\sum_{n=\mathcal{C}+1}^{\infty}$ for any arbitrary $\mathcal{C}$, and $\chi_\omega$ can be approximated from below by utilizing $\mathcal{L}_\omega \geq 1$ in the latter one. Then, the corresponding $\widetilde{\omega}$ can be calculated accordingly.

Thus, listening is preferred over attacking when $\chi_\omega > 1$; i.e., $\omega < \widetilde{\omega}$, which proves the theorem. □

Note that, after the critical value of $\omega \geq \widetilde{\omega}$, since $\chi_\omega \leq 1$, we cannot compare the attacking and listening rewards based on (11).

### 4.4 Listening or Attacking (by Value Iteration)

Due to Theorem 1, listening is optimal for $\omega < \widetilde{\omega}$. An optimal strategy; i.e., listening or attacking, for $\omega \geq \widetilde{\omega}$ will be our focus in this part.

Note that the attacker gets an immediate reward $c_r$ only when the attack is successful. The attacker gets a (discounted) reward by listening on account of the successful attacks in the future. Therefore, the attacker gets zero reward if she only listens, which implies that for any amount of observation $\widehat{\omega} \geq \widetilde{\omega}$, there must be some $\overline{\omega} \geq \widehat{\omega}$, in which attacking is optimal.

Since a backward induction through the Bellman optimality equations for the attacker reward is already established in (8), we are ready to apply the value iteration method to obtain the optimal attacker strategy for $\omega \geq \widetilde{\omega}$ (note that since $\frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\lambda_u} \lambda_u^n}{n!} < 1$, the Bellman update/operator in (8) is a contraction mapping, which guarantees the existence and the uniqueness of an optimal point, that is achievable by the value iteration method).

**Theorem 2.** *The attacker prefers attacking over listening for $\omega \geq \widetilde{\omega}$.*

*Proof.* For the initial values of the rewards, we assign zero reward for every $\omega$; i.e, $J_\omega^{(0)} = 0 \ \forall \omega$. Regarding the first iteration of the value updates, since

$$J_{\omega,L}^{(1)} = \frac{K_\omega^{(0)}}{1-U} = \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\lambda_u} \lambda_u^n}{n!} J_{\omega+n}^{(0)} = 0\,; \tag{12}$$

i.e., all listening rewards are zero, attacking would be the optimal choice for every $\omega$. Then, $J_{\omega,L}^{(1)} = 0$ and $J_{\omega,A}^{(1)} = \frac{C_\omega}{1-T_\omega}$ hold $\forall \omega$, which implies $J_{\omega,*}^{(1)} = \frac{C_\omega}{1-T_\omega} \ \forall \omega$.

In the second iteration, the attacking rewards do not change; i.e., $J_{\omega,A}^{(2)} = \frac{C_\omega}{1-T_\omega} \ \forall \omega$. Regarding the listening rewards, for every $\omega$, similar to (11), we obtain

$$\begin{aligned} J_{\omega,L}^{(2)} &= \frac{K_\omega^{(1)}}{1-U} = \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\lambda_u} \lambda_u^n}{n!} \frac{C_{\omega+n}}{1-T_{\omega+n}} \\ &= \frac{C_\omega}{1-T_\omega} \underbrace{\frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\lambda_u} \lambda_u^n}{n!} \left( \prod_{i=0}^{n-1} \mathcal{L}_{\omega+i} \right)}_{\chi_\omega} . \end{aligned} \tag{13}$$

In (13), if $\chi_\omega \leq 1$, then $J_{\omega,L}^{(2)} \leq \frac{C_\omega}{1-T_\omega} = J_{\omega,A}^{(2)}$, which implies that attacking is the optimal strategy. Since $\lim_{\omega \to \infty} \chi_\omega < 1$, and $\chi_\omega$ is a decreasing function of $\omega$, after the critical value $\omega \geq \widetilde{\omega}$, attacking is always preferred over listening.

Similarly, if $\chi_\omega > 1$, or equivalently if $\omega < \widetilde{\omega}$, since $J^{(2)}_{\omega,L} > \frac{C_\omega}{1-T_\omega} = J^{(2)}_{\omega,A}$, listening is preferred over attacking. Thus, at the end of the second iteration, the following holds regarding the value update $J^{(2)}_{\omega,*}$ and the corresponding strategy:

$$\left.\begin{aligned} J^{(2)}_{\omega,L} &= \tfrac{K^{(1)}_\omega}{1-U}\ \forall\omega \\ J^{(2)}_{\omega,A} &= \tfrac{C_\omega}{1-T_\omega}\ \forall\omega \end{aligned}\right\} \quad \Rightarrow \quad J^{(2)}_{\omega,*} = \begin{cases} J^{(2)}_{\omega,L} & \omega < \widetilde{\omega} \\ J^{(2)}_{\omega,A} & \omega \geq \widetilde{\omega} \end{cases} .$$

In the third iteration, the attacking rewards are the same again; i.e., $J^{(3)}_{\omega,A} = \frac{C_\omega}{1-T_\omega}\ \forall\omega$. For the listening rewards, since

$$J^{(3)}_{\omega,L} = \frac{K^{(2)}_\omega}{1-U} = \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!} J^{(2)}_{\omega+n,*}$$

holds, we have $J^{(3)}_{\omega,L} = J^{(2)}_{\omega,L}$ for $\omega \geq \widetilde{\omega}$. Regarding $\omega < \widetilde{\omega}$, observe the following:

$$\begin{aligned} J^{(3)}_{\omega,L} &= \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \left( \sum_{n=1}^{\widetilde{\omega}-\omega-1} \frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!} J^{(2)}_{\omega+n,L} + \sum_{\widetilde{\omega}-\omega}^{\infty} \frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!} J^{(2)}_{\omega+n,A} \right) \\ &> \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \left( \sum_{n=1}^{\widetilde{\omega}-\omega-1} \frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!} J^{(2)}_{\omega+n,A} + \sum_{\widetilde{\omega}-\omega}^{\infty} \frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!} J^{(2)}_{\omega+n,A} \right) \\ &= J^{(2)}_{\omega,L} > J^{(2)}_{\omega,A} = J^{(3)}_{\omega,A}\,, \end{aligned}$$

which implies that, for $\omega < \widetilde{\omega}$, listening would be the optimal strategy, as in the previous iteration. Furthermore, the listening rewards are greater than or equal to the ones from the previous iteration; i.e., $J^{(3)}_{\omega,L} > J^{(2)}_{\omega,L}$ for $\omega < \widetilde{\omega} - 1$ and $J^{(3)}_{\omega,L} = J^{(2)}_{\omega,L}$ for $\omega = \widetilde{\omega} - 1$.

Note that the optimal attacker reward $J_{\omega,*}$ is obtained partially at each iteration. In particular, $J_{\omega,*}$ is obtained for $\omega \geq \widetilde{\omega}$ in the second iteration, $J_{\omega,*}$ is obtained for $\omega = \widetilde{\omega}-1$ in the third iteration, $J_{\omega,*}$ can be obtained for $\omega = \widetilde{\omega}-2$ in the forth iteration, and so on. By iterating further in this way, we observe that listening is optimal for $\omega < \widetilde{\omega}$ and attacking is optimal for $\omega \geq \widetilde{\omega}$, and we can obtain the optimal reward and strategy of the attacker in at most $\widetilde{\omega}+2$ number of iterations. Moreover, $J^{(n)}_{\omega,A} = \frac{C_\omega}{1-T_\omega}$ and $J^{(n)}_{\omega,L} \geq J^{(n-1)}_{\omega,L}$ hold $\forall\omega$ (in particular, $J^{(n)}_{\omega,L} > J^{(n-1)}_{\omega,L}$ for $\omega < \widetilde{\omega}$ and $J^{(n)}_{\omega,L} = J^{(n-1)}_{\omega,L}$ for $\omega \geq \widetilde{\omega}$). □

Based on the results and observations above, the optimal attack strategy can be summarized as follows:

$$\begin{cases} \text{Waiting}\ \ (l(t)=0\,,\ a(t)=0) & \text{if } S(t)=\mathrm{BL}\,,\ L(t) \text{ arbitrary} \\ \text{Listening}\ \ (l(t)=1\,,\ a(t)=0) & \text{if } S(t)=\mathrm{UB}\,,\ L(t) < \widetilde{\omega} \\ \text{Attacking}\ \ (l(t)=0\,,\ a(t)=1) & \text{if } S(t)=\mathrm{UB}\,,\ L(t) \geq \widetilde{\omega} \end{cases} \tag{14}$$

## 5 Optimal Defense Strategy

Due to the leadership role, for any defense strategy ($m$, $\eta_u$), the defender (operator) can anticipate the optimal attacker strategy; i.e., the critical amount of observation $\widetilde{\omega}$ (as a function of $m$ and $\eta_u$), and corresponding decisions (i.e., listening is optimal for $\omega < \widetilde{\omega}$ and attacking is optimal for $\omega \geq \widetilde{\omega}$). Knowing $\widetilde{\omega}$, the defender can make use of the transition probabilities described in Section 4.1[4].

From the perspective of the defender, the system state consists of the triplets $(t, L(t) = \omega, S(t))$ which evolve over time as an MDP, and the corresponding transition probabilities are provided in Section 4.1. Note that the evolution of the triplets $(t, L(t) = \omega, S(t))$ starting from $(0, 0, \mathrm{UB})$ can be represented as an infinite directed graph with countably many vertices (note that $t$ and $w$ are discrete, and there are three possible states $S(t)$). We will make use of the vertices as states/rewards, and the edges as the transition probabilities as follows. Firstly, let us define the total defender reward until the time-slot $t$ as $V_t(\omega, \mathrm{UB})$, $V_t(\omega, \mathrm{BL})$, and $V_t(\omega, \mathrm{AD})$ when $L(t) = \omega$, and $S(t) = \mathrm{UB}$ (the unblocking state), $S(t) = \mathrm{BL}$ (the blocking state), and $S(t) = \mathrm{AD}$ (the attacker is detected by the IDS and the game ends), respectively. Starting from $V_0(0, \mathrm{UB}) = 0$ and for $\omega < \widetilde{\omega}$, a forward recursion on the total defender reward can be established as

$$\begin{aligned}
V_{t+1}(\omega, \mathrm{BL}) &= (1-q)V_t(\omega, \mathrm{BL}) + (1-\delta_l(m))\eta_u(1-\mathrm{e}^{-\lambda_u})V_t(\omega, \mathrm{UB})\,, \\
V_{t+1}(\omega, \mathrm{UB}) &= qV_t(\omega, \mathrm{BL}) + (1-\delta_l(m))\mathrm{e}^{-\lambda_u}V_t(\omega, \mathrm{UB}) \\
&\quad + (1-\delta_l(m))(1-\eta_u)\sum_{n=1}^{\omega}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}\Big(v_r + V_t(\omega-n, \mathrm{UB})\Big)\,, \\
V_{t+1}(\omega, \mathrm{AD}) &= \delta_l(m)V_t(\omega, \mathrm{UB})\,. 
\end{aligned} \tag{15}$$

Similarly, for $\omega \geq \widetilde{\omega}$, since the attacker prefers attacking, the recursion becomes

$$\begin{aligned}
V_{t+1}(\omega, \mathrm{BL}) &= (1-q)V_t(\omega, \mathrm{BL}) + (1-\delta_a(m))\Big(1 - P_{uu}\big(1-\Phi(\xi_\omega)\big)\Big)V_t(\omega, \mathrm{UB})\,, \\
V_{t+1}(\omega, \mathrm{UB}) &= qV_t(\omega, \mathrm{BL}) + (1-\delta_a(m))P_{uu}\big(1-\Phi(\xi_\omega)\big)(-c_r + V_t(\omega, \mathrm{UB})) \\
&\quad + (1-\delta_l(m))(1-\eta_u)\sum_{n=\omega-\widetilde{\omega}+1}^{\omega}\frac{\mathrm{e}^{-\lambda_u}\lambda_u^n}{n!}\Big(v_r + V_t(\omega-n, \mathrm{UB})\Big)\,, \\
V_{t+1}(\omega, \mathrm{AD}) &= \delta_a(m)V_t(\omega, \mathrm{UB})\,. 
\end{aligned} \tag{16}$$

The game continues until the attacker is detected, i.e., until $S(t) = \mathrm{AD}$ for some $t$, and the defender is interested in maximizing its average utility through all possible realizations of the events/transitions. However, since the defender does not know $L(t) = \omega$, stochastic averaging is needed to obtain the average

[4] As depicted in Fig. 1-(b), the average amount of observation learnt by the attacker is $\sum_{N=1}^{\infty}(1-\delta_l(m))(1-\eta_u)\frac{\mathrm{e}^{-\lambda_u}\lambda_u^N}{N!}N = (1-\delta_l(m))(1-\eta_u)\lambda_u$. Thus, after $\frac{\widetilde{\omega}}{(1-\delta_l(m))(1-\eta_u)\lambda_u}$ time-slots, the defender can assume that the attacker has learned enough information to imitate the user; i.e., attacking is optimal for the attacker.

defender reward. Thus, the goal of the defender can be expressed as

$$\overline{V} = \sup_{m,\eta_u} \left( \left( \sum_{t,\omega} \frac{V_t(\omega, \mathrm{AD})}{t} \right) - m \right), \tag{17}$$

where $V_t(\omega, \mathrm{AD})$ satisfies the recursions in (15) and (16) (depending on $\omega$) with the initial condition $V_0(0, \mathrm{UB}) = 0$, and $m$ stands for the per time-slot operation cost of the IDS. In the next section, after presenting the illustrations regarding the averaging part $\sum_{t,\omega} \frac{V_t(\omega,\mathrm{AD})}{t}$, we provide a corresponding discussion on (17).

## 6 Numerical Results

In the following we show results from extensive simulations to illustrate the optimal attacker strategy, optimal attacker reward, and the optimal average defender reward. Unless otherwise noted, we use the default parameter values shown in Table 1. We do not make use of a default value for $m$ and $C_a$, but we provide a related discussion at the end of the section.

Table 1: Default parameters.

| | |
|---|---|
| $\lambda_u$ | 10 |
| $\mathcal{B}_u$ | $\mathcal{N}(100, 3)$ |
| $\eta_u$ | 0.01 |
| $v_r$ | 0.1 |
| $c_r$ | 1 |
| $\delta_l(m)$ | 0.1 |
| $\delta_a(m)$ | 0.2 |
| $q$ | 0.7 |
| $\rho$ | 0.98 |
| $\gamma$ | 0.1 |
| $m$ | |
| $C_a$ | |

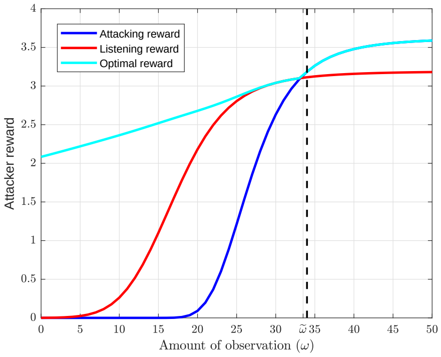
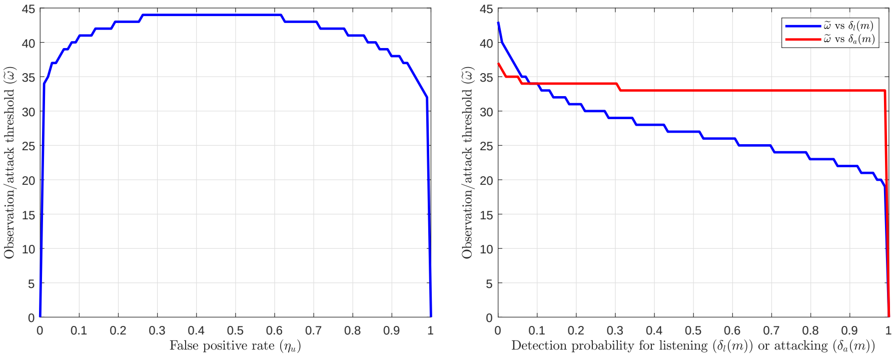


Fig. 3: Attacker reward vs. amount of observation ($\omega$) under different strategies.

Fig. 3 shows the attacker's reward as a function of the amount of observation ($\omega$) for various attacker strategies: an attacker that always attacks (blue), i.e., $J_\omega = \frac{C_\omega}{1-T_\omega}$, listening reward assuming that attacking is optimal for all future $\omega$ (red); i.e., $J_\omega = \frac{K_\omega}{1-U} = \frac{\rho(1-\delta_l(m))(1-\eta_u)}{1-U} \sum_{n=1}^{\infty} \frac{\mathrm{e}^{-\lambda_u} \lambda_u^n}{n!} \frac{C_{\omega+n}}{1-T_{\omega+n}}$, and the optimal choice (green). The observation/attack threshold ($\widetilde{\omega} = 34$) is represented by a vertical line. As it can be observed from Fig. 3, the optimal strategy performs better than the other strategies. After $\omega \geq \widetilde{\omega}$, since attacking is optimal, the optimal and attacking curves coincide. Before $\omega < \widetilde{\omega}$, since it is not optimal to attack; i.e., the attacker does not gather enough amount of observation to attack, optimal strategy is always listening.

Fig. 4 shows the observation/attack threshold $\widetilde{\omega}$ (which can only take integer values) to start attacking as a function of detection parameters $\eta_u$, $\delta_l(m)$, and $\delta_a(m)$. Fig. 4-(a) illustrates that since the success probability of attack is higher, the attacking is more desirable when the FP rate is low. Furthermore, for the higher FP rates, since the system stays in the blocking state longer, attacking would be the preferred action. Therefore, for the lower and higher rates of FP, the attacker is urged to attack, as depicted in Fig. 4-(a). Note that since the attacking reward in (7) is not dependent on the amount of observation $\omega$ for the extreme cases (i.e., $\widetilde{q}_\omega = 1$ for $\eta_u = 0$, and $\widetilde{q}_\omega = 0$ for $\eta_u = 1$), the attacker prefers attacking without listening so that $\widetilde{\omega} = 0$. Fig. 4-(b) shows that the attacker is more willing to attack (instead of listening) as the probability of being detected increases, so that $\widetilde{\omega}$ is non-increasing.



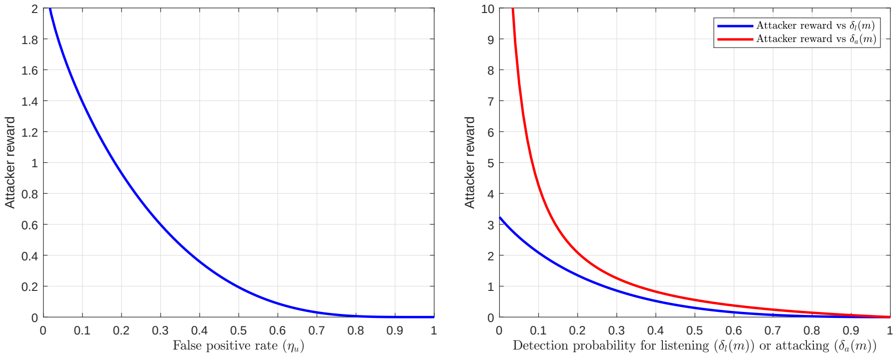
(a) Threshold ($\widetilde{\omega}$) vs. $\eta_u$.

(b) Threshold ($\widetilde{\omega}$) vs. $\delta_l(m)$ and $\delta_a(m)$.

Fig. 4: Observation/attack threshold ($\widetilde{\omega}$) vs. detection parameters

Fig. 5-(a) shows that the reward of the attacker is a convex decreasing function of $\eta_u$. This is because a higher false positive rate implies a higher true positive rate, and thus the system stays in the blocking state longer. Similarly, Fig. 5-(b) shows that the attacker reward is also a convex decreasing function of the detection probabilities.

Note that the optimization of the defender reward in (17) consists of taking the expectation over infinitely many states $(t, \omega, \mathrm{AD})$ and infinitely many different paths, as discussed in Section 5. In order to be able to cover all most-likely states $(t, \omega, \mathrm{AD})$ and the corresponding average rewards $\frac{V_t(\omega,\mathrm{AD})}{t}$ (excluding the per time-slot threat monitoring cost $m$ in (17)), we implemented Monte Carlo simulations that run the dynamic stochastic game with corresponding states and transition probabilities in (15) and (16) $1 \cdot 10^6$ times for a given parameter set of the defender ($\eta_u$, $\delta_l(m)$, $\delta_a(m)$), and then we take the average of all results to obtain the average defender reward (which corresponds to $\sum_{t,\omega} \frac{V_t(\omega,\mathrm{AD})}{t}$ in (17)) and the average detection time (i.e., the average length of the game).



(a) Attacker reward vs. false positive rate ($\eta_u$).

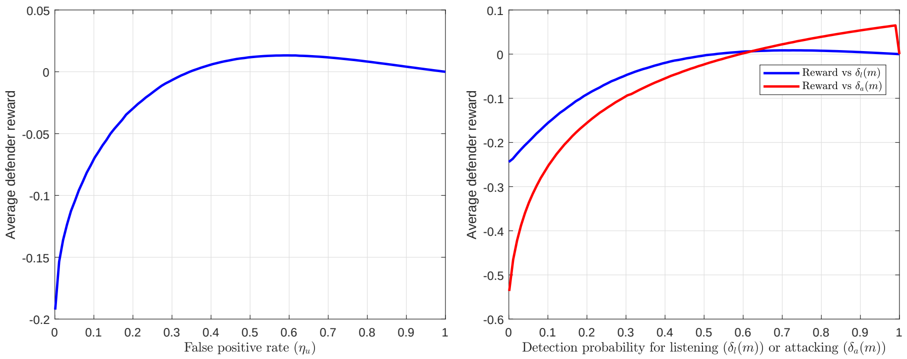
(b) Attacker reward vs. $\delta_l(m)$ and $\delta_a(m)$.

Fig. 5: Attacker reward vs. detection parameters

Fig. 6 shows the average reward of the defender as a function of detection parameters ($\eta_u$, $\delta_l(m)$, $\delta_a(m)$). Fig. 6-(a) shows that the average defender reward is a concave function of the false positive rate $\eta_u$. This is due to that as the FP rate increases, the amount of immediate reward the defender obtains decreases, while the attacker prefers listening longer. At the same time, the TP rate increases with $\eta_u$, and thus the ratio of unsuccessful attacks increases, which reduces the damage caused to the defender. In Fig. 6-(b), as $\delta_l(m)$ increases, the trade-off between the decreasing utilization of the system (which reduces the gain for the defender) and decreasing success probability of attack (which reduces the damage to the defender since the attacker attacks with less amount of observation) can be observed. Also note that when the attacker prefers attacking, the reward of the defender is always negative since the system is either in the blocking state (which has zero reward for the defender) or the attack is successful (which has a cost to the defender). Therefore, reducing the attack rate makes the defender better off. However, when $\delta_a(m) = 1$, as illustrated in Fig. 4-(b), the attacker prefers attacking without observing any input from the user. Thus, both players get zero reward.

Fig. 7 illustrates the relation between the average length of the game (i.e., the average detection time for the defender) and the detection parameters ($\eta_u$, $\delta_l(m)$, $\delta_a(m)$). Fig. 7-(a) shows that the average length of the game increases (except for the extreme case) as $\eta_u$ increases. This is because as the TP and FP rates increase, the attacker spends more time listening. Furthermore, Fig. 7-(b) shows that as the probability of being detected increases, the average length of the game decreases, but with a marginal gain.

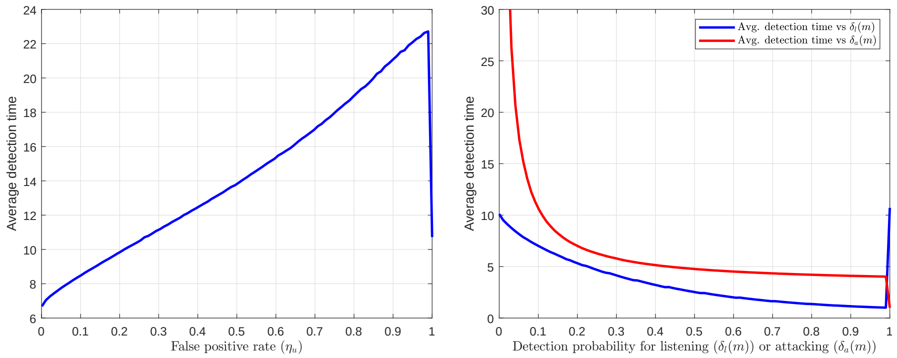
Note that when the attacker reward is less than the cost of compromising the system; i.e., if $J_{\omega,*} = J(0, \mathrm{UB}) \leq C_a$, a rational attacker would refrain from attacking. In this case the state evolution can be described by (1), i.e., the case of no attacker, and the system can be modeled as a renewal-reward process, whose



(a) Average defender reward vs. false positive rate ($\eta_u$).

(b) Average defender reward vs. $\delta_l(m)$ and $\delta_a(m)$.

Fig. 6: Average defender reward vs. detection parameters



(a) Average detection time vs. false positive rate ($\eta_u$).

(b) Average detection time vs. $\delta_l(m)$ and $\delta_a(m)$.

Fig. 7: Average detection time vs. detection parameters

average reward (defender cost) can be easily calculated. Thus, if the defender knows $C_a$, it can optimize the parameters ($\eta_u$, $\delta_l(m)$, $\delta_a(m)$) by formulating a constrained maximization problem with objective $\left(\sum_{t,\omega} \frac{V_t(\omega,\mathrm{AD})}{t}\right) - m$ subject to $J_{0,\mathrm{UB}} \leq C_a$.

## 7 Conclusion

The problem of security risk management using continuous authentication was modeled as a dynamic discrete stochastic leader-follower game with imperfect information between an attacker and a defender (an operator). The optimal attacker reward was derived as a backward dynamic programming recursion, and the corresponding optimal strategy was obtained by the value iteration algorithm. Then, based on the optimal strategy of the attacker, the average reward of the defender was expressed as a forward recursion. Extensive simulations illustrate the relations between the optimal strategies/rewards and the defender parameters, and show that continuous authentication can be very efficient for security risk reduction, if combined with appropriate incident detection.

## References

1. Castiglione, A., Raymond Choo, K., Nappi, M., Ricciardi, S.: Context aware ubiquitous biometrics in edge of military things. IEEE Cloud Computing **4**(6), 16–20 (Nov 2017)
2. Dee, T., Richardson, I., Tyagi, A.: Continuous transparent mobile device touchscreen soft keyboard biometric authentication. In: International Conference on VLSI Design (VLSID). pp. 539–540 (Jan 2019)
3. Deutschmann, I., Nordström, P., Nilsson, L.: Continuous authentication using behavioral biometrics. IT Professional **15**(4), 12–15 (July 2013)
4. Ferro, M., Pioggia, G., Tognetti, A., Mura, G.D., De Rossi, D.: Event related biometrics: Towards an unobtrusive sensing seat system for continuous human authentication. In: International Conference on Intelligent Systems Design and Applications. pp. 679–682 (Nov 2009)
5. Goncalves, L., Subtil, A., Oliveira, R.M., de Zea Bermudez, P.: ROC curve estimation: An overview. Revstat - Statistical Journal **12**, 1–20 (Mar 2014)
6. Khouzani, M.H.R., Mardziel, P., Cid, C., Srivatsa, M.: Picking vs. guessing secrets: A game-theoretic analysis. In: IEEE Computer Security Foundations Symposium. pp. 243–257 (July 2015)
7. Peng, G., Zhou, G., Nguyen, D.T., Qi, X., Yang, Q., Wang, S.: Continuous authentication with touch behavioral biometrics and voice on wearable glasses. IEEE Transactions on Human-Machine Systems **47**(3), 404–416 (June 2017)
8. Sitová, Z., Šeděnka, J., Yang, Q., Peng, G., Zhou, G., Gasti, P., Balagani, K.S.: HMOG: New behavioral biometric features for continuous authentication of smartphone users. IEEE Transactions on Information Forensics and Security **11**(5), 877–892 (May 2016)
9. Xiao, L., Li, Y., Han, G., Liu, G., Zhuang, W.: Phy-layer spoofing detection with reinforcement learning in wireless networks. IEEE Transactions on Vehicular Technology **65**(12), 10037–10047 (Dec 2016)
10. Yang, L., Lu, Y., Liu, S., Guo, T., Liang, Z.: A dynamic behavior monitoring game-based trust evaluation scheme for clustering in wireless sensor networks. IEEE Access **6**, 71404–71412 (2018)
11. Yunchuan, G., Lihua, Y., Licai, L., Binxing, F.: Utility-based cooperative decision in cooperative authentication. In: IEEE INFOCOM. pp. 1006–1014 (Apr 2014)