Int. J. Inf. Secur. (2018) 17:221–242
https://doi.org/10.1007/s10207-017-0364-2

REGULAR CONTRIBUTION

A game-theoretic approach for integrity assurance in resource-bounded systems

Aron Laszka[1] · **Yevgeniy Vorobeychik**[2] · **Xenofon Koutsoukos**[2]

Published online: 31 January 2017

Abstract Assuring communication integrity is a central problem in security. However, overhead costs associated with cryptographic primitives used toward this end introduce significant practical implementation challenges for resource-bounded systems, such as cyber-physical systems. For example, many control systems are built on legacy components which are computationally limited, but have strict timing constraints. If integrity protection is a binary decision, it may simply be infeasible to introduce into such systems; without it, however, an adversary can forge malicious messages, which can cause significant physical or financial harm. To bridge the gap between such binary decisions, we propose a stochastic message authentication approach that can explicitly trade computational cost off for security. We introduce a formal game-theoretic framework for optimal stochastic message authentication, providing provable guarantees for resource-bounded systems based on an existing message authentication scheme. We use our framework to investigate attacker deterrence, as well as optimal stochastic message authentication when deterrence is impossible, in both short-term and long-term equilibria. Additionally, we propose two schemes for implementing stochastic message authentication in practice, one for saving computation only at the receiver and one for saving computation at both ends, and demonstrate the associated computational savings using an actual implementation.

Keywords Message authentication · Game theory · Economics of security · Resource-bounded system

1 Introduction

Ensuring communication integrity in networked systems is a fundamental problem in security research, one with an abundance of solutions that typically rely on cryptographic primitives. For example, if the sender and receiver share a secret key, message integrity can be guaranteed (in a typical cryptographic sense) by using message authentication codes (MAC). In a MAC scheme, for each outgoing message $\boldsymbol{m}$, the sender generates an authentication tag $t = \mathrm{MAC}(K, \boldsymbol{m})$ using the key K and attaches it to the message. Then, for each incoming message $(\boldsymbol{m}, t)$, the receiver also computes the tag as $\mathrm{MAC}(K, \boldsymbol{m})$ and verifies whether it matches the tag attached to the message.

Message authentication schemes are typically based on cryptographic primitives, such as cryptographic hash functions or block ciphers. Unfortunately, these can be expensive to compute. In numerous applications, the overhead of cryptographic routines is negligible, for example, when these run on state-of-the-art desktop computers. Many applications, however, particularly those of relevance in cyber-physical systems (such as supervisory control systems), involve a myriad of legacy, embedded, or battery-powered devices, such as smart cards, RFID tags, and sensors [1,6,8,14]. For example, many devices used in electric power grids are decades old [7], and these legacy devices were not designed with security in mind and are often connected by "unsecured" avenues [5,27]. The severely limited computational power of these devices

✉ Aron Laszka
laszka@berkeley.edu

Yevgeniy Vorobeychik
yevgeniy.vorobeychik@vanderbilt.edu

Xenofon Koutsoukos
xenofon.koutsoukos@vanderbilt.edu

[1] University of California, Berkeley, CA, USA

[2] Vanderbilt University, Nashville, TN, USA

makes cryptographic computation prohibitive, particularly when there are tight timing and/or energy requirements. Since upgrading such systems can entail prohibitive costs, security is often compromised in favor of performance. Given the importance of systems composed of such resource-bounded devices, from the electric power grid to nuclear power plants, lack of assured integrity can be devastating, as an attacker can introduce arbitrary messages into the system [6].

Numerous approaches for "lightweight" cryptography have previously been proposed to address this problem [8,20,23] (see related work in Sect. 6). However, these have the same fundamental limitation: a decision to secure a system is binary; either security is employed, incurring some associated overhead, or it is not. Thus, if the computational requirements for a given lightweight security primitive are too high for a particular system, one is simply out of luck. Furthermore, most of the recently proposed lightweight cryptographic schemes have not seen widespread deployment, which means that their security has not been put to a real-world test.

Consequently, assuring communication integrity in resource-bounded systems poses an interesting technical challenge. Without adequate resources to authenticate every important message using strong cryptography, we must carefully choose what to authenticate. On the other hand, compared to confidentiality, providing integrity presents an opportunity since adversaries cannot tell what is authenticated (i.e., if we provided confidentiality, a man-in-the-middle attacker could see what is encrypted and what is plaintext). Hence, we can authenticate a random subset of messages without revealing which subset we have chosen. However, an adversary may anticipate the strategy that we use to choose random subsets, which can make finding an optimal strategy a challenging computational problem.

We address the problem of assuring integrity in resource-bounded devices by creating a general purpose framework for explicitly trading off security requirements and computational constraints of the device. Our approach can thus be applied to an arbitrary resource-bounded device, with associated formal guarantees about achieved integrity. In this paper, we target systems in which individual modified or spoofed messages alone are non-catastrophic, but each message may lead to some loss in the system, and we express guarantees about integrity in terms of the maximum loss that a rational attacker may cause. For example, in a smart grid, tampering with sensor data coming from a household consumer may result in only small variations in the aggregate consumption and, consequently, cannot cause a breakdown [11]. However, tampered data will result in inefficient power control and, if these data are used for billing, in financial losses. As another example, in intelligent traffic control, tampering with traffic flow data cannot cause accidents in practical systems due to hardware-based failsafes [13]. However, distorted traffic data will lead to suboptimal traffic control and traffic light schedules, which will result in wasted time and increased environmental impact.

For these systems, we propose a stochastic message authentication framework, which authenticates messages randomly in a way that abides by the resource constraints of the system. We introduce a game-theoretic model to achieve two ends: first, provide algorithmic means to compute an optimal stochastic authentication strategy, accounting for the relative importance of messages, and second, to provide formal guarantees about the extent that system integrity is preserved, as well as expected damage when it is not. We consider the case when there are not enough computational resources to provide integrity protection for all important messages, which means that some important messages have to be sent and received without authentication. The stochastic nature of our approach prevents attackers from predicting which messages can be modified without risking detection. By taking the relative importance of messages into account, we can devise optimal stochastic strategies for selecting which messages to protect, which result in lower expected damage than selecting all important messages with uniform probability.

Our main contributions are:

- Based on our threat model and objectives (Sect. 2), we formulate stochastic message authentication as an attacker–defender game (Sect. 3), considering both short-term and long-term conflicts.
- We study the adversary's best responses (Sect. 4.1), characterize when the adversary can be deterred from attacking (Sect. 4.2.1), discuss finding an optimal defense when deterrence is impossible (Sect. 4.2.2), study the defender's best responses (Sect. 4.3.1), and characterize when the game has a Nash equilibrium (Sect. 4.3.2).
- We propose two schemes for implementing stochastic authentication in practice (Sects. 5.2 and 5.3) and demonstrate their viability using experiments (Sect. 5.4).

This paper is a significant extension of our previous publication [18], in which we (1) restricted our analysis to short-term conflicts and (2) considered only the problem of saving computation at the receiving end of communication, without considering the sender's computational costs. Besides extending our analysis to long-term conflicts, we also introduce a novel scheme in this paper, called stochastic generation, which allows saving computation at both the sender and the receiver. As a key part of the new contributions, we study the security properties that this scheme has to satisfy, and we propose algorithms for implementing it in practice.

The significance of this extension lies not only in the new theoretical results, but also in substantially broadening the

range of problems to which our approach can be applied. For example, monitoring spatially distributed cyber-physical systems usually entails operating a large number of low-power sensor devices. Since these devices act primarily as senders of data, the scheme considered in our previous publication cannot lead to significant savings in computation; however, the extension presented in this paper enables us to substantially decrease the computational load of these sensor devices.

2 Threat model and objectives

We assume that the adversary is capable of modifying or fabricating messages sent to the receiver, but she is not able to generate correct authentication tags. From our point of view, modified and fabricated messages are equivalent (i.e., both have malicious content and incorrect authentication tags); consequently, we will use the word modify exclusively for the remainder of the paper. We assume that the adversary's goal is to cause damage or loss by modifying messages, while remaining undetected. Finally, we assume that the adversary cannot change traffic patterns substantially, since anomalies, such as substantially increased amount of traffic, would be detected.

Our goal is to reduce the computational cost of a given MAC-based message authentication scheme, while maintaining an acceptable probability of detecting modified messages (see Sect. 3 for details). Consequently, we do *not* intend to provide security features that are not already provided by the MAC scheme, such as thwarting replay attacks.

3 Game-theoretic model

Now, we introduce our game-theoretic model of stochastic message authentication. For ease of presentation, we consider only *stochastic verification* in our model. In stochastic verification, the receiver saves computation by verifying only a random subset of incoming messages, but the sender has to compute a correct authentication tag for every outgoing message. In Sect. 5, we will then introduce *stochastic generation*, which enables the sender to also save computation by computing correct authentication tags for only a random subset of outgoing messages. Since optimal strategies for stochastic verification and generation are the same, we can simplify the presentation of our results by restricting it to verification for now and then discuss how our results can be implemented for generation in Sect. 5.3.

We model the problem as a two-player, non-deterministic, nonzero-sum game between an adversary, who tries to cause damage or loss by modifying some messages, and a defender, who tries to detect the presence of the adversary by verifying the authenticity of some messages (see Definition 1 below). For a list of symbols used in the model, see Table 1.

Table 1 List of symbols

Symbol	Description
C	Number of message classes
L_c	Amount of loss that a message of class c can cause
F	Adversary's punishment for getting caught
T_c	Traffic (i.e., amount of messages) of class c
B	Computational budget of the defender
p_c	Probability that the defender verifies a message of class c
a_c	Number of messages of class c modified by the adversary

We begin by discussing the properties of the potentially malicious messages that can be received by the defender. Each received message—regardless of whether it has been modified by the adversary or not—is assigned to one of C classes based on the amount of damage or loss it could cause if it were malicious. For example, messages that control the air conditioning system of a car obviously belong to a less dangerous class, while messages that control the brakes belong to a more dangerous class. As another example, messages containing abnormally high or low values (e.g., measurements of critical pressure values in a pressure vessel) may belong to a more dangerous class than messages containing normal values. We denote the amount of loss that a modified message of class $c \in \{1, \ldots, C\}$ can potentially cause by $L_c > 0$. Further, we assume that these losses are additive. Formally, if a_c messages have been modified for each class $c \in \{1, \ldots, C\}$, then the cumulative loss sustained by the system is assumed to be

$$\sum_{c=1}^{C} a_c L_c \tag{1}$$

if the attack remains undetected.

The defender represents the receiver of the messages, who has the ability to verify any given message and tell whether it has been modified or not. We assume that this verification is perfect, that is, it can always tell whether a message has been modified or if it is authentic. In other words, we assume that the underlying cryptographic primitives are secure.

The defender's strategic choice is to select, for each class $c \in \{1, \ldots, C\}$, the probability p_c that a message belonging to class c is verified upon its reception. Since any temporal correlation may give a statistical edge to the attacker, we assume that the decision to verify a message is made independently from the other messages. Now, if the defender were able to verify every message (i.e., if she could select $p_c = 1$

for every class c), then she would be able to always detect any attack. However, verifying a message has some computational cost (e.g., computing a cryptographic hash of the message), and the defender has only a limited computational budget, which does not allow her to verify every single message. Formally, we assume that the defender can choose a strategy $\boldsymbol{p}$ only if it satisfies

$$\sum_{c=1}^{C} p_c T_c \leq B \ , \tag{2}$$

where T_c is the amount of traffic for message class c and B is the defender's computational budget.

Note that this budget constraint formulation can be used with messages of varying verification costs as well; in this case, we simply let T_c be the expected computational cost of verifying every message of class c. For the defender, the challenge lies in finding a strategy that maximizes the probability of detection while being feasible with respect to the computational budget limit.

The adversary represents an attacker or a malware that has penetrated the system, and who is now trying to cause damage or loss by modifying messages. The adversary's strategic choice is to select, for each class $c \in \{1, \ldots, C\}$, the number of messages $a_c \in \mathbb{N}$ that she modifies. Using this notation, the probability of an attack remaining undetected is

$$\prod_{c=1}^{C} (1 - p_c)^{a_c}. \tag{3}$$

The adversary's goal is to maximize both the probability of remaining undetected and the cumulative loss sustained by the system when she succeeds in remaining undetected. The former is important not only because of the success of the attack, but also because the adversary sustains a punishment of value $F > 0$ when she is detected. For the adversary, the challenge arises from these two goals being opposite.

Finally, since the adversary cannot change traffic patterns substantially, her strategy has only negligible effect on T_c for every class c. Consequently, the defender knows in advance which strategies will be feasible with respect to her computational budget, and which strategies will be infeasible.

Now, we define our game formally.

Definition 1 The *message authentication game* has two players, called the *defender* and the *adversary*, and it is played as follows:

1. First, the defender selects a strategy $\boldsymbol{p} = (p_1, \ldots, p_C) \in [0, 1]^C$ satisfying $\sum_c p_c T_c \leq B$.
2. Then, the adversary selects a strategy $\boldsymbol{a} = (a_1, \ldots, a_C) \in \mathbb{N}^C$.
3. Finally, nature chooses outcome *undetected* with probability $\prod_{c=1}^{C} (1 - p_c)^{a_c}$ and outcome *detected* with probability $1 - \prod_{c=1}^{C} (1 - p_c)^{a_c}$.
4. For a given outcome, the players' payoffs are given by the following table:

		Outcome	
		undetected	*detected*
Payoff for	defender	$-\sum_{c=1}^{C} a_c L_c$	0
	adversary	$\sum_{c=1}^{C} a_c L_c$	$-F$

We assume symmetry between the defender's loss and the attacker's gain for two reasons: firstly, to consider the worst-case attacker, who tries to maximize damage, as is common in security, and secondly, to minimize the number of model parameters. Note that our model and results generalize to asymmetry in a relatively straightforward manner.

In our analysis, we assume that both players try to maximize their respective expected payoffs. For a given strategy profile $(\boldsymbol{p}, \boldsymbol{a})$, the defender's expected payoff (i.e., expected inverse loss) can be expressed as

$$\mathcal{U}_D(\boldsymbol{p}, \boldsymbol{a}) = -\prod_{c=1}^{C} (1 - p_c)^{a_c} \sum_{c=1}^{C} a_c L_c \ , \tag{4}$$

and the adversary's expected payoff can be expressed as

$$\mathcal{U}_A(\boldsymbol{p}, \boldsymbol{a}) = \prod_{c=1}^{C} (1-p_c)^{a_c} \sum_{c=1}^{C} a_c L_c - \left(1 - \prod_{c=1}^{C} (1 - p_c)^{a_c}\right) F \tag{5}$$

$$= \prod_{c=1}^{C} (1 - p_c)^{a_c} \left(\sum_{c=1}^{C} a_c L_c + F\right) - F. \tag{6}$$

Note that we will refer to expected payoff and expected loss simply as payoff and loss whenever usage is unambiguous.

Finally, we will consider two different solution concepts, Stackelberg and Nash equilibrium, which correspond to different assumptions on the time span of the conflict.

Stackelberg equilibrium Firstly, suppose that the conflict is short term, and the game is played only once. Following Kerckhoffs's principle, we assume that the attacker knows the defender's algorithms, implementation, etc., and can thus compute the defender's strategy. On the other hand, the defender cannot observe and respond to the attacker's strategy. Hence, in this case, we have to find the adversary's best-response and the defender's optimal strategies, which are defined formally as follows.

Definition 2 A player's strategy is a *best response* if it maximizes the player's payoff, taking the other player's strategy as given.

As is typical in the security literature, we consider a refinement of subgame perfect equilibria, called *strong Stackelberg* equilibria [15]. We will refer to the defender's equilibrium strategies as optimal strategies for the remainder of the paper.

Definition 3 We call a defense strategy *optimal* if it maximizes the defender's payoff given that the adversary always plays a best response with tie-breaking in favor of the defender. Formally, strategy $\boldsymbol{p}$ is optimal if it maximizes

$$\max_{\boldsymbol{a}^* \in \operatorname{argmax}_{\boldsymbol{a}} \mathcal{U}_A(\boldsymbol{p},\boldsymbol{a})} \mathcal{U}_D(\boldsymbol{p}, \boldsymbol{a}^*). \tag{7}$$

Note that the effect of the tie-breaking rule is negligible in practice, and its sole purpose is to avoid pathological mathematical cases where no optimal strategy would exist.

Nash equilibrium Secondly, suppose that the conflict is long term, and the game is played multiple times. In this case, both players may observe and respond to their opponents' strategies. Hence, we have to find both players' best-response strategies and—if it exists—the *Nash equilibrium* that they form, which is defined formally for our game as follows.

Definition 4 A strategy profile $(\boldsymbol{p}, \boldsymbol{a})$ forms a Nash equilibrium if

- strategy $\boldsymbol{p}$ is a best response against strategy $\boldsymbol{a}$
- and strategy $\boldsymbol{a}$ is a best response against strategy $\boldsymbol{p}$.

4 Analysis

In this section, we present theoretical results on our message authentication game. First, in Sect. 4.1, we discuss the adversary's best-response strategies, on which both the Stackelberg and the Nash equilibria depend. Then, we study the defender's optimal (i.e., Stackelberg equilibrium) strategies in Sects. 4.2.1 and 4.2.2. In Sect. 4.2.1, we characterize those instances of the message authentication game where the defender's optimal payoff is zero, while in Sect. 4.2.2, we study the instances where the optimal payoff is nonzero. Finally, we characterize the defender's best-response strategies and the existence of Nash equilibria in Sects. 4.3.1 and 4.3.2, respectively.

We let $\mathbf{1}$ and $\mathbf{0}$ denote vectors of ones and zeros, respectively (their sizes are not indicated, as they are never ambiguous).

4.1 Adversary's best response

We begin our analysis with characterizing the adversary's best responses. Being able to characterize and compute the adversary's best responses is of key importance, since this allows us to quantify how secure a given defense is (i.e., compute the defender's expected loss for a given strategy).

4.1.1 Continuous relaxation

First, we study a continuous relaxation of the problem. Notice that detection probability, cumulative loss, and the players' payoffs remain well defined if we allow $\boldsymbol{a}$ to be an arbitrary vector of non-negative real numbers, instead of integers. Hence, we can easily define a continuous relaxation of the model as follows.

Definition 5 The *continuous relaxation* of the message authentication game is played as the original game, except that the adversary can select a strategy $(a_1, \ldots, a_C) \in \mathbb{R}^C_{\geq 0}$.

Although the relaxed model has no practical interpretation, it will play an important role in facilitating the analysis of the original model and finding optimal defense strategies in computationally challenging instances. The following lemma provides a necessary condition on the best responses in the relaxed model.

Lemma 1 *Let $\boldsymbol{a} \in \mathbb{R}^C_{\geq 0}$ be a best-response strategy against some defense strategy $\boldsymbol{p}$. Then, for every class $i \in \{1, \ldots, C\}$,*

- *either $a_i = 0$*
- *or $\frac{L_i}{\ln(1-p_i)} = -F - \sum_{c=1}^{C} a_c L_c$ must hold.*

The proof of Lemma 1 can be found in "Appendix".

The above lemma implies that, in a best-response strategy, the ratio $\frac{L_c}{\ln(1-p_c)}$ has to be uniform over those classes c for which the number of modified messages is nonzero. Since this ratio depends only on the defender's strategy, we can divide the classes into groups based on their ratios and readily have that the adversary will modify messages from only a single group.

In order to characterize the adversary's best-response strategies, we have to answer two questions. The first question asks which group is selected by a best response (i.e., which ratio maximizes the adversary's payoff), while the second one asks which classes are selected from the payoff-maximizing group. The following lemma can help us answer both questions.

Lemma 2 *Let $\boldsymbol{a} \in \mathbb{R}^C_{\geq 0}$ be an adversarial strategy, let $\boldsymbol{p} < \mathbf{1}$ be a defense strategy, and assume that $\frac{L_i}{\ln(1-p_i)} \geq \frac{L_j}{\ln(1-p_j)}$. Then, if we decrease a_i by Δ (where $\Delta \leq a_i$) and increase*

a_j by $\Delta\frac{L_i}{L_j}$, the adversary's payoff does not decrease. Furthermore, the adversary's payoff increases if and only if the inequality between the ratios is strict.

The proof of Lemma 2 can be found in "Appendix".

Intuitively, the above lemma says that any two classes having the same ratio are "payoff-equivalent", that is, we can increase the number of modified messages for one class and decrease it for the other class, without changing the adversary's payoff. Furthermore, the adversary can achieve higher payoff by attacking classes with lower ratios.[1] Using the above lemma, we can characterize the adversary's best-response strategies as follows (please recall that we can disregard classes c with $p_c = 1$, since a best response never modifies messages of such classes).

Theorem 1 *Given a defense strategy $\boldsymbol{p} < 1$, the adversary's best-response strategy modifies messages of only those classes i for which the ratio $\frac{L_i}{\ln(1-p_i)}$ is minimal. Furthermore, there always exists a best-response strategy which modifies messages of at most one class only.*

Proof First, we show that a best response modifies messages of classes with minimal ratios only. For the sake of contradiction, suppose that the claim does not hold for some best-response strategy $\boldsymbol{a}^*$, that is, there exists a class i with non-minimal ratio such that $a_i^* > 0$. Then, let j be some class with minimal ratio, and consider the strategy $\hat{\boldsymbol{a}}$ defined as follows: $\hat{a}_i = 0$, $\hat{a}_j = a_i^* + a_j^*$, and $\hat{a}_c = a_c^*$ for every $c \neq i, j$. From Lemma 2, we readily have that the adversary's payoff is strictly higher for strategy $\hat{\boldsymbol{a}}$ than for strategy $\boldsymbol{a}^*$; however, this contradicts our initial assumption that $\boldsymbol{a}^*$ is a best-response strategy. Therefore, the first claim of the theorem has to hold.

Second, we show how to construct a best-response strategy which modifies messages of at most one class only. Let $\boldsymbol{a}^*$ be an arbitrary best-response strategy, and let M be the set of classes c for which $a_c^* > 0$. If $|M| \leq 1$, then strategy $\boldsymbol{a}^*$ already satisfies the condition, so we are ready. Otherwise, let class i be an arbitrary element of the set M, and consider the strategy $\hat{\boldsymbol{a}}$ defined as follows: $\hat{a}_i = \sum_{c \in M} a_c^*$, and $\hat{a}_c = 0$ for every $c \neq i$. Now, from the first claim of the theorem, we already have that classes in M all have minimal ratios. Consequently, it follows from Lemma 2 that the adversary's payoff for strategy $\hat{\boldsymbol{a}}$ is the same as for strategy $\boldsymbol{a}^*$, which implies that $\hat{\boldsymbol{a}}$ is a best response. Since $\hat{\boldsymbol{a}}$ also satisfies the condition that it modifies messages of at most one class only (i.e., of class i), we have proved the existence of such a best response. □

Unfortunately, this result does not apply to the original, integral model, since the adversary cannot choose arbitrary, non-integral combinations of message numbers in the original model. For an example, see Fig. 2a later.

[1] Note that, since the ratios are always negative, this means that the adversary will attack classes with ratios of higher absolute value.

4.1.2 Special case of a single message class

We continue our analysis of the adversary's best-response strategies with the special case of a single message class (i.e., $C = 1$) in the original, integral model (Definition 1). The following lemma characterizes the adversary's best responses.

Lemma 3 *In the special case of $C = 1$, the adversary's best-response strategies against a given defense strategy $p_1 > 0$ are either $\lfloor a^* \rfloor$, $\lceil a^* \rceil$, or zero, where*

$$a^* = -\frac{1}{\ln(1-p_1)} - \frac{F}{L_1}. \tag{8}$$

The proof of Lemma 3 can be found in "Appendix".

The formula presented in the above lemma can also be used to find a best response in the relaxed model. From Theorem 1, we have that there exists a best-response strategy which modifies messages of only a single class, which has minimal ratio. Hence, we can compute a best-response strategy for the adversary by finding a^* for a class c that has minimal ratio $\frac{L_c}{\ln(1-p_c)}$. Note that, in this case, we obviously do not have to round a^* to the nearest integers.

4.1.3 Original model

Now, we study the adversary's best-response strategies in the general case of the original model (as defined in Definition 1) and discuss how to find a best-response strategy in practice. We have seen that, in the special case of a single message class, we can characterize the adversary's best response using Eq. (8). Unfortunately, we cannot use this characterization directly in the general case, as the adversary's best responses might modify messages from multiple classes. However, we will show that we can use it as an upper bound. First, we have to prove the following lemma.

Lemma 4 *Let $\boldsymbol{p}$ be a defense strategy, and let c be an arbitrary class. If a_c^* were the maximal best-response strategy given that the adversary could modify messages of class c only, then every best response $\hat{\boldsymbol{a}}$ must satisfy $\hat{a}_c \leq a_c^*$.*

The proof of Lemma 4 can be found in "Appendix".

Intuitively, this lemma states that, if the adversary is allowed to modify messages of multiple classes, then for each class, she will modify at most as many messages as she would if she were restricted to that single class. Since we already have a characterization for the case of a single class from Lemma 3, we can use the above lemma to constrain the adversary's best responses. The following theorem establishes class-wise upper bounds on the adversary's best responses.

Theorem 2 *Against a given defense strategy* $\boldsymbol{p} > 0$*, any best-response adversarial strategy* $\boldsymbol{a}$ *must satisfy*

$$\forall c \in \{1, \ldots, C\} : \; a_c \leq \max\left\{0, \left\lceil -\frac{1}{\ln(1-p_c)} - \frac{F}{L_c} \right\rceil\right\}. \tag{9}$$

Proof First, we have from Lemma 3 that, for any class c, the adversary's single-class best responses are either $\lceil a_c^* \rceil$, $\lfloor a_c^* \rfloor$, or zero, where $a_c^* = -\frac{1}{\ln(1-p_c)} - \frac{F}{L_c}$. Hence, the maximal single-class best response is at most

$$\max\left\{0, \left\lceil -\frac{1}{\ln(1-p_c)} - \frac{F}{L_c} \right\rceil\right\} \tag{10}$$

for each class c. Then, it follows readily from Lemma 4 that, for every best-response strategy $\boldsymbol{a}$ and every class c, $a_c \leq \max\left\{0, \left\lceil -\frac{1}{\ln(1-p_c)} - \frac{F}{L_c} \right\rceil\right\}$ has to hold. □

Based on this theorem, we can find the adversary's best response using exhaustive search by enumerating all strategies that satisfy the upper bound constraints. Even though the running time of this approach is exponential in the number of classes, it scales surprisingly well in practice, as the bounds are typically very low (see the following paragraph and Fig. 1). Furthermore, note that this computation should be performed at design time, not by the computationally limited device during runtime.

Numerical illustrations Figure 1 shows the adversary's single-class best response $-\frac{1}{\ln(1-p_c)} - \frac{F}{L_c}$ in the continuous model (dashed line - -) and the upper bound $\left\lceil -\frac{1}{\ln(1-p_c)} - \frac{F}{L_c} \right\rceil$ on her strategy in the original model (solid line —) as functions of the defender's verification probability p_c for $F = 0$.

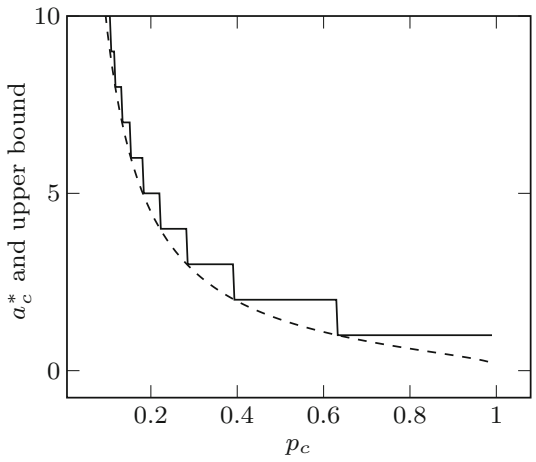

Fig. 1 Adversary's single-class best response in the continuous model (*dashed line*) and the upper bound on her strategy in the original model (*solid line*) as functions of the verification probability p_c for $F = 0$

The figure shows that the bound is low even for very low verification probabilities. For example, at $p_c = 0.2$ the bound is still only 5, which allows us to easily find a best-response strategy in practice, e.g., using an exhaustive search. Note that, for higher values of F, both the continuous best response and the bound are even lower. Since we are primarily interested in finding effective defense strategies, which limit the losses caused by an adversary, the bounds will usually be low. If any of the bounds is high for a given defense strategy, then we can throw away that strategy without finding the adversary's best response, since a single-class attack can be used to show that the given defense strategy is ineffective (recall from Sect. 4.1.2 that we can easily compute the adversary's best response for a single message class).

Figure 2 shows the adversary's payoff for various strategies $\boldsymbol{a} = (a_1, a_2)$ against a given defense strategy $\boldsymbol{p} = (p_1, p_2)$ in the case of two classes (i.e., $C = 2$). First, in Fig. 2a, the ratios $\frac{L_c}{\ln(1-p_c)}$ (i.e., the ratios between the potential losses and the logarithms of the not-verifying probabilities) are the same for the two classes. As expected from Lemma 2, we see that the strategies with the highest payoffs are along a diagonal, and the best response is the strategy ($a_1 = 3$, $a_2 = 1$) that best approximates the optimum of the continuous relaxation. Second, in Fig. 2b, there is a substantial difference between the ratios, and modifying messages of the second class is a better choice for the adversary. Hence, in the best response ($a_1 = 0$, $a_2 = 3$), the adversary modifies messages of the second class only.

4.2 Defender's optimal strategy

Now, we study the problem of finding an optimal strategy for the defender. Recall from Definition 3 that a defense strategy is optimal if it minimizes the defender's loss given that the adversary always plays a best response. With respect to the defender's optimal strategy, we can divide the instances of the message authentication game into two groups: instances where the defender can achieve *zero loss* by *deterring* the adversary from attacking and instances where the defender's optimal *loss* is *nonzero*.

Definition 6 A defense strategy $\boldsymbol{p}$ is a *deterrence strategy* if not attacking at all (i.e., $\boldsymbol{a} = \boldsymbol{0}$) is a best response.

4.2.1 Deterrence strategies

We begin our analysis of the optimal defense strategies by characterizing those instances of the message authentication game where the defender has a deterrence strategy. The following theorem provides a closed-form characterization of deterrence strategies.

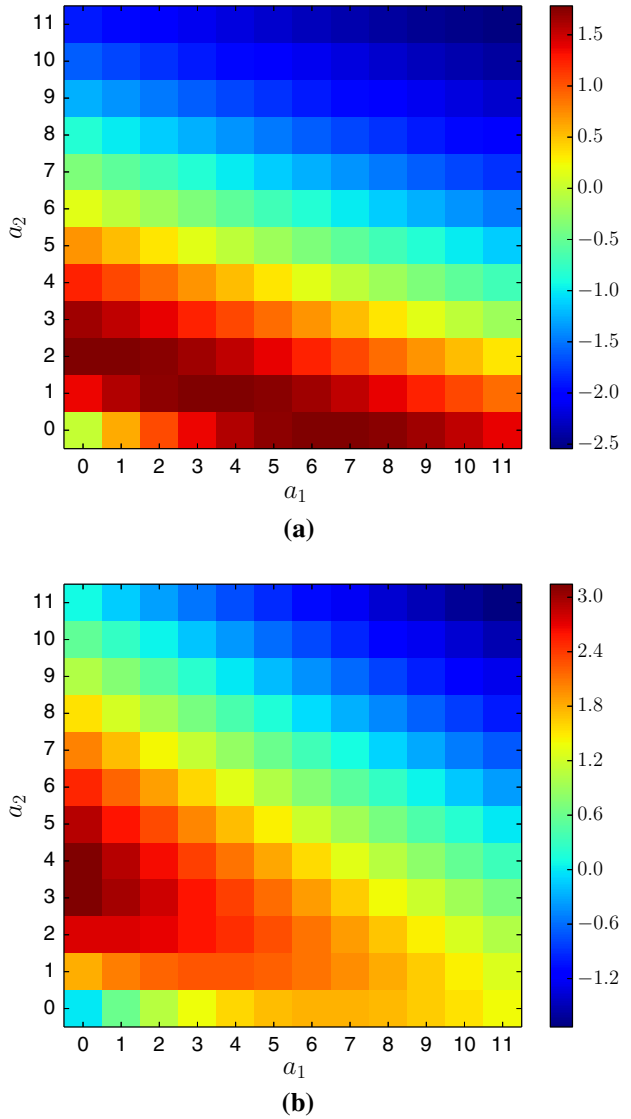

Fig. 2 Adversary's payoff for various strategies against a given defense strategy. The *horizontal axis* shows the number of messages modified from the first class, while the *vertical axis* shows the number for the second class, and the *coloring* shows the adversary's expected payoff (see legend). The parameters are $F = 3$, $L_1 = 1$, and $L_2 = 3$. **a** Case $\frac{L_1}{\ln(1-p_1)} = \frac{L_2}{\ln(1-p_2)}$. The defender's strategy is $p_1 = 0.1$ and $p_2 \approx 0.271$. The best response is $a_1 = 3$, $a_2 = 1$. **b** Case $\frac{L_1}{\ln(1-p_1)} > \frac{L_2}{\ln(1-p_2)}$. The defender's strategy is $p_1 = 0.1$ and $p_2 = 0.2$. The best response is $a_1 = 0$, $a_2 = 3$

Theorem 3 *Given a defense strategy* $\boldsymbol{p}$*, not attacking at all (i.e.,* $\boldsymbol{a} = \boldsymbol{0}$*) is the adversary's best-response strategy if and only if*

$$\forall c \in \{1, \ldots, C\} : \; p_c \geq \frac{L_c}{L_c + F}. \tag{11}$$

Proof First, we prove the necessity of Eq. (11). For the sake of contradiction, suppose that Eq. (11) does not hold for some class c. Then, the adversary's payoff for modifying a single message of class c (i.e., $a_c = 1$) is

$$(1 - p_c)L_c - p_c F \geq \frac{F}{L_c + F} L_c - \frac{L_c}{L_c + F} F = 0. \tag{12}$$

In other words, the adversary's payoff for this strategy is higher than for not attacking (i.e., higher than zero payoff), which implies that not attacking cannot be a best response. Therefore, Eq. (11) necessarily holds if not attacking is a best response.

Second, we prove the sufficiency of Eq. (11). We show sufficiency for any number of classes C using induction. We begin by showing that the condition is sufficient for $C = 1$. For any $a_1 > 0$, we have

$$\begin{aligned} F(L_1 + F)^{a_1} &= F\left(F^{a_1} + a_1 F^{a_1 - 1} L_1 + \cdots\right) && (13)\\ &\geq F\left(F^{a_1} + a_1 F^{a_1 - 1} L_1\right) && (14)\\ &= F^{a_1}\left(a_1 L_1 + F\right), && (15) \end{aligned}$$

which implies that

$$\frac{F}{a_1 L_1 + F} \leq \left(\frac{F}{L_1 + F}\right)^{a_1}. \tag{16}$$

The adversary's payoff for any strategy $a_1 > 0$ is

$$\begin{aligned} &(1 - p_1)^{a_1}(a_1 L_1 + F) - F && (17)\\ &= (a_1 L_1 + F)\left((1 - p_1)^{a_1} - \frac{F}{a_1 L_1 + F}\right) && (18)\\ &\leq \underbrace{(a_1 L_1 + F)}_{\geq 0}\left(\underbrace{\left(\frac{F}{L_1 + F}\right)^{a_1} - \frac{F}{a_1 L_1 + F}}_{\leq 0}\right) \leq 0. && (19) \end{aligned}$$

Hence, no strategy can achieve higher payoff than not attacking (i.e., higher than zero payoff), which proves that not attacking is a best response.

Now, assume that the claim of the theorem holds for $C - 1$ classes. Then, for C classes, we show that the adversary's payoff for any given strategy $\boldsymbol{a}$ is at most zero if the condition of the theorem holds. For the remainder of the proof, let us define

$$\hat{L} = \sum_{c=1}^{C-1} a_c L_c \quad \text{and} \quad \hat{P} = \prod_{c=1}^{C-1} (1 - p_c)^{a_c}.$$

Since the claim holds for $C - 1$ classes, we have

$$\hat{P}\hat{L} \leq (1 - \hat{P})F. \tag{20}$$

Furthermore, we also have from the $C = 1$ case that

$$(1 - p_C)^{a_C} a_C L_C \leq \left(1 - (1 - p_C)^{a_C}\right) F. \tag{21}$$

Using the notations $\hat{L}$ and $\hat{P}$, the adversary's expected payoff for strategy $\boldsymbol{a}$ can be expressed as

$$\mathcal{U}_A(\boldsymbol{p}, \boldsymbol{a}) = \prod_{c=1}^{C}(1-p_c)^{a_c} \sum_{c=1}^{C} a_c L_c - \left(1 - \prod_{c=1}^{C}(1-p_c)^{a_c}\right) F \tag{22}$$

$$= \hat{P}(1-p_C)^{a_C}(\hat{L} + a_C L_C) - \left(1 - \hat{P}(1-p_C)^{a_C}\right) F \tag{23}$$

$$= (1-p_C)^{a_C}\hat{P}\hat{L} + \hat{P}(1-p_C)^{a_C} a_C L_C - \left(1 - \hat{P}(1-p_C)^{a_C}\right) F. \tag{24}$$

Now, we use Eqs. (20) and (21), which give us

$$\mathcal{U}_A(\boldsymbol{p}, \boldsymbol{a}) \leq (1-p_C)^{a_C}(1-\hat{P})F + \hat{P}(1-(1-p_C)^{a_C})F - \left(1 - \hat{P}(1-p_C)^{a_C}\right) F \tag{25}$$

$$= F\Big((1-p_C)^{a_C}(1-\hat{P}) + \hat{P}(1-(1-p_C)^{a_C}) - 1 + \hat{P}(1-p_C)^{a_C}\Big) \tag{26}$$

$$= F\left((1-p_C)^{a_C} + \hat{P} - 1 - \hat{P}(1-p_C)^{a_C}\right) \tag{27}$$

$$\leq \underbrace{F}_{\geq 0}\left(\underbrace{(\hat{P}-1)}_{\leq 0}\underbrace{\left(1-(1-p_C)^{a_C}\right)}_{\geq 0}\right) \leq 0. \tag{28}$$

Hence, no strategy can achieve higher payoff than not attacking (i.e., higher than zero payoff). Therefore, Eq. (11) has to be sufficient for an arbitrary number of classes C, which concludes our proof. □

Based on the above theorem, we can easily characterize those instances of the message authentication game where the defender has a deterrence strategy. Since a defense strategy is a deterrence strategy if and only if every probability is at least as high as some constant value, we only have to test whether the computational budget is high enough to afford all of these probabilities.

Corollary 1 *The defender has a deterrence strategy if and only if*

$$B \geq \sum_c \frac{L_c}{L_c + F} T_c. \tag{29}$$

If the condition of the corollary holds, then the defender can easily construct a deterrence strategy and achieve zero loss.

4.2.2 Optimal defense without deterrence

Next, we consider those instance of the message authentication game where the defender has no deterrence strategy.

Continuous relaxation First, we study the continuous relaxation of the problem (see Definition 5), where the adversary can choose any vector of non-negative real numbers. The following theorem characterizes the defender's optimal strategy.

Theorem 4 *Suppose that the defender has no deterrence strategy. Then, in the continuous model, an optimal defense strategy* $\boldsymbol{p}$ *has to satisfy*

$$\frac{L_1}{\ln(1-p_1)} = \frac{L_2}{\ln(1-p_2)} = \cdots = \frac{L_C}{\ln(1-p_C)} \tag{30}$$

and

$$\sum_c p_c T_c = B. \tag{31}$$

Furthermore, there always exists a unique defense strategy satisfying these conditions.

The proof of Theorem 4 can be found in "Appendix".

Even though we cannot express the optimal defense strategy in closed form, we can compute it easily using the argument presented in the last paragraph of the proof (and some numerical optimization method). Furthermore, observe that the optimal strategy is independent of the value of F; hence, only the relative values of L_c have to be estimated in practice to compute the strategy.

Original model Now, we return to our original, integral model. Compared to the continuous model, the analysis of the integral model is more challenging, since the adversary's payoff is not a continuous function of the defender's strategy, which can lead to many counter-intuitive phenomena. For instance, in the integral model, the defender's payoff can decrease when she increases the verification probability of a single class. More formally, let $\mathcal{U}_D^*(\boldsymbol{p})$ denote the defender's expected payoff for a strategy $\boldsymbol{p}$ given that the adversary always plays her best response. Then, $\mathcal{U}_D^*(\boldsymbol{p})$ is *not necessarily* a non-decreasing function of a variable p_i. For an example, consider the function $\mathcal{U}_D^*(p_1, p_2)$ shown in Fig. 3. Around $p_1 = 0.2$, the value of $\mathcal{U}_D^*(p_1, 0.45)$ clearly decreases when we increase p_1. This is very surprising, since it shows that performing more verifications can sometimes lead to a lower level of security.

However, the following lemma shows that the defender's payoff can only increase if she increases the verification probability of every class, given that she maintains the right ratio between the probabilities.

Lemma 5 *Let $\boldsymbol{p}^*$ be a non-deterrence defense strategy, and let $\boldsymbol{p}'$ be such that* $\ln \frac{1-p_c^*}{1-p_c'} = \varepsilon L_c$ *, where $\varepsilon \in \mathbb{R}_{>0}$. Then, assuming that the adversary always plays a best response, the defender's payoff for $\boldsymbol{p}'$ is higher than for $\boldsymbol{p}^*$.*

The proof of Lemma 5 can be found in "Appendix".

It is interesting to note that, if $\boldsymbol{p}^* = \boldsymbol{0}$ and $\sum_c p_c' T_c = B$ (i.e., if we start with zero verification probabilities and use all of the budget), then $\boldsymbol{p}'$ is actually equal to the optimal defense strategy of the continuous model. This suggests that the continuous model can be used in practice as an approximation to find a reasonably good defense strategy. We will later see that this intuition is indeed right.

Next, we use the above lemma to provide necessary constraints on the optimal defense strategies, which can be used to restrict the search space when searching for an optimal strategy.

Theorem 5 *Suppose that the defender has no deterrence strategy. Then, if $\boldsymbol{p}^*$ is an optimal defense strategy, it must satisfy*

- $p_i^* \leq \frac{L_i}{L_i+F}$ *for every i,*
- *and $p_i^* \geq p_j^*$ for every $L_i > L_j$.*

Proof (Sketch.) We begin with proving the necessity of the first condition. For the sake of contradiction, suppose that the claim does not hold for some optimal strategy $\boldsymbol{p}^*$, and let i be a class for which $p_i^* > \frac{L_i}{L_i+F}$. Then, we can construct a strictly better strategy $\boldsymbol{p}'$ as follows. First, substitute p_i^* with $\frac{p_i^* + \frac{L_i}{L_i+F}}{2}$. This substitution does not change the set of the adversary's best responses or the players' payoffs, since the adversary never attacks a class if its verification probability is higher than $\frac{L_i}{L_i+F}$ (see the proof of Theorem 3). However, this substitution decreases the defender's sum computational cost; hence, $\sum_c p_c^* T_c < B$ holds after the substitution. Second, we show that we can construct a strictly better strategy $\boldsymbol{p}'$ using this saving in computational cost and Lemma 5. Clearly, there exists a strategy $\boldsymbol{p}'$ for every value of ε in Lemma 5; furthermore, every p_c' is a continuous, strictly increasing function of ε. Hence, for every $B < \sum_c T_c$, there exists an ε such that $\sum_c p_c' T_c = B$. Finally, we have from Lemma 5 that this strategy $\boldsymbol{p}'$ is strictly better than $\boldsymbol{p}^*$, which contradicts the initial assumption that $\boldsymbol{p}^*$ is optimal. Therefore, the claim has to hold.

Next, we prove the necessity of the second condition. For the sake of contradiction, suppose that the claim does not hold for some optimal strategy $\boldsymbol{p}^*$, and let i and j be classes for which $p_i^* < p_j^*$ and $L_i > L_j$. Then, attacking class i is strictly superior to attacking class j for the adversary, since messages of class i have both strictly lower probability and strictly higher potential loss. Consequently, no best-response strategy would attack class j, and we can decrease p_j^* without changing the payoffs or the set of best responses. Next, we can construct a strictly better strategy $\boldsymbol{p}'$ using the saving in computational cost and Lemma 5 (see previous paragraph). However, this contradicts our initial assumption that $\boldsymbol{p}^*$ is optimal. Therefore, the claim has to hold. □

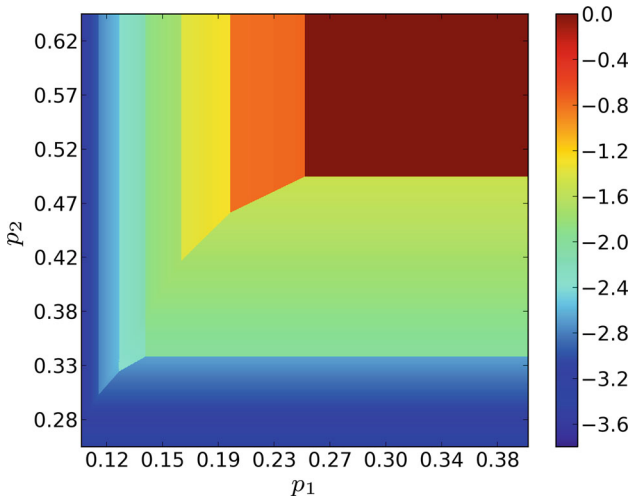

Fig. 3 Defender's payoff for various strategies given that the adversary plays her best response. The parameters are $F = 3$, $L_1 = 1$, and $L_2 = 3$

One of the most important consequences of Lemma 5 is that an optimal defense strategy always uses all of the available computational budget, which allows us to further restrict the search space.

Theorem 6 *Suppose that the defender has no deterrence strategy. Then, if $\boldsymbol{p}^*$ is an optimal defense strategy, it must satisfy $\sum_i p_i T_i = B$.*

The proof of Theorem 6 can be found in "Appendix".

Now, we discuss how to find an optimal defense strategy in practice. First, the defender's payoff changes smoothly over regions where the adversary's best responses are the same (see Fig. 3 for an illustration); hence, once we find the right region, we can easily find the optimal strategy using numerical optimization methods. The challenge lies in the potentially exponential number of regions, whose boundaries can cause large "jumps" in the defender's payoff. However, using the necessary conditions presented in this section, we can restrict the search space greatly. Furthermore, for strategies that are reasonably good, the adversary's possible best responses are very limited (see Theorem 2); hence, the number of regions to actually consider is small.

A very important element of the search is being able to quickly throw inferior strategies away, without computing the adversary's actual best response. Once we have a reasonably good defense strategy with payoff $\mathcal{U}_D^*$, we can do this for any defense strategy by finding an adversarial strategy

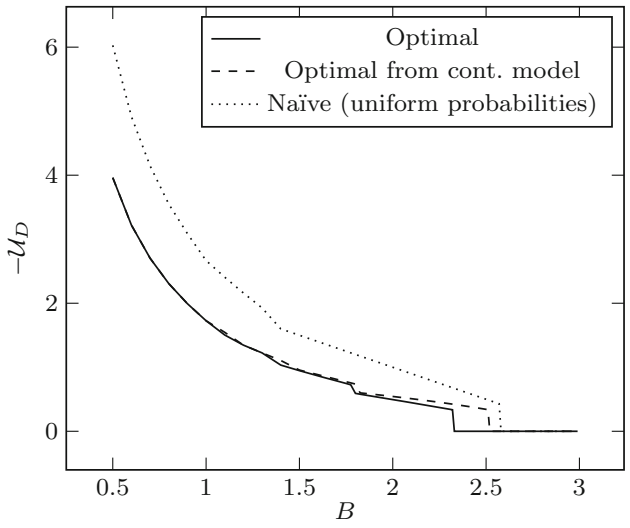

Fig. 4 Defender's expected loss for her optimal strategy (*solid line*) compared to her expected loss for the optimal strategy computed in the continuous model (*dashed line*) and her expected loss for a naïve strategy using uniform probabilities (*dotted line*). The parameters are $F = 0.5$, $L_1 = 1$, $L_2 = 2$, $L_3 = 3$, and $T_1 = 1$, $T_2 = 1$, $T_3 = 1$

that attains at least $-\mathcal{U}_D^*$ payoff for the adversary. Since the defender's loss is always greater than the adversary's payoff, we can safely throw away a defense strategy if we find such an attack against it. For this test, we can use single-class best responses, which can be computed in constant time and perform well against inferior defense strategies. In case a strategy passes the test, we have to determine whether it is better than the current solution by computing the adversary's actual best response. The number of inferior strategies passing the test depends on how far the game is from being zero-sum, that is, their number is high when F is high. However, when F is high, then the problem actually becomes easier, since the adversary's strategy space will be very limited (see Theorem 2). Finally, we can use the optimal defense strategy from the continuous model as an initial solution, as it is generally a good approximation for difficult instances (see Fig. 4 and its discussion).

Numerical illustrations Figure 3 shows the defender's payoff for various strategies $\boldsymbol{p} = (p_1, p_2)$ assuming that the adversary always plays her best response. We can see that the payoff is a non-continuous function of the defense strategy, but it changes smoothly over regions where the adversary's best responses are the same. Furthermore, we can also see that—quite interestingly— the payoff is not always an increasing function of the individual probabilities. Finally, the figure confirms Theorem 3, which predicts the minimal deterrence strategy to be $(p_1 = 0.25, p_2 = 0.5)$.

Figure 4 shows the defender's expected payoff as a function of her budget for various defense strategies: the solid line (—) shows her expected payoff for her optimal strategies, the dashed line (- -) for her optimal strategies computed based on the continuous relaxation of the model[2], and the dotted line $(\cdots)$ for naïve strategies that assign the same verification probability to every class. In every case, we assume that the adversary plays her best-response strategy. The figure shows that, for lower budget values, the solution of the relaxed problem (dashed line) approximates the solution of the original problem (solid line) reasonably well. For higher budget values, the two lines diverge (until the adversary is deterred in both cases); however, for these higher values, solving the original problem is relatively easy.[3] The figure also shows that optimal strategies lead to substantially lower loss for the defender than naïve, non-strategic solutions (dotted line).

4.3 Nash equilibrium

Next, we study the Nash equilibria of the game, in which both players choose best-response strategies. Since we have already analyzed the attacker's best response in Sect. 4.1, all that remains is characterizing the defender's best response.

4.3.1 Defender's best response

We begin our analysis with a necessary condition on the defender's best response.

Theorem 7 *Let $\boldsymbol{p}$ be a best-response strategy against some adversarial strategy $\boldsymbol{a}$. If the defender's expected payoff for $\boldsymbol{p}$ is nonzero, then $\boldsymbol{p}$ has at most one nonzero element.*

The proof of Theorem 7 can be found in "Appendix".

Based on the above theorem, we can characterize the defender's best-response strategy as follows.

Corollary 2 *Against an adversarial strategy $\boldsymbol{a}$, the defender's best response $\boldsymbol{p}$*

- *either satisfies $p_c = 1$ for some $a_c > 0$ and achieves zero loss;*
- *or has a single nonzero element p_c, where c minimizes $\left(1 - \frac{B}{T_c}\right)^{a_c}$ over all the classes, and achieves $\left(1 - \frac{B}{T_c}\right)^{a_c} \sum_i L_i$ loss.*

Note that the first case holds if and only if there is a class c such that $T_c \leq B$ and $a_c > 0$.

[2] Note that we are interested in comparing how different strategies perform in the original, realistic model; hence, we compute an optimal defense strategy in the relaxed model, but evaluate it in the original one.

[3] High budget values allow for high verification probabilities, which mean low upper bounds on the adversary's best responses (see Theorem 2).

4.3.2 Equilibrium characterization

Finally, the following theorem characterizes the existence of Nash equilibria.

Theorem 8 *For $C > 1$, either there exists a deterrence strategy for the defender, or the game has no Nash equilibrium.*

Note that we already have a characterization of the instances where a deterrence strategy exists from Sect. 4.2.1.

Proof (Sketch.) For the sake of contradiction, suppose that the claim of the theorem does not hold, and let $(\boldsymbol{p}, \boldsymbol{a})$ be a Nash equilibrium with nonzero payoffs. Since $\boldsymbol{p}$ is a best response, we have that it has exactly one nonzero element p_n. Combined with $C > 1$, this implies that $\boldsymbol{p}$ has at least one zero element p_z. It is easy to see that the adversary's expected payoff is finite for any strategy where $a_n > 0$, and arbitrarily high for any strategy where $a_n = 0$ and a_z is arbitrarily high. Since $\boldsymbol{a}$ is a best response, $a_n = 0$ and $a_z > 0$ must hold. However, this leads to a contradiction with our initial supposition that $\boldsymbol{p}$ is a best response, as $p_n = 0$ and $p_z > 0$ is obviously a better response to $\boldsymbol{a}$ than $\boldsymbol{p}$. Therefore, the claim of the theorem must hold. □

5 Implementation

In this section, we discuss how our theoretical results can be implemented and used in practice. First, in Sect. 5.2, we consider stochastic verification, a strict implementation of the model presented in Sect. 3, which saves computation only at the receiver. Then, in Sect. 5.3, we consider stochastic generation, which provides the same level of security, but saves computation at the sender as well.

5.1 Mapping the parameters to real-world data

Our model has five parameters: number of classes C, amount of traffic $\boldsymbol{T}$, computational budget B, potential losses $\boldsymbol{L}$, and adversary's punishment F. In case we have no information regarding the messages at design time, we may implement our stochastic message authentication approach with a single message class (i.e., $C = 1$), which provides a baseline level of security. Any additional information (i.e., dividing messages into multiple classes and estimating the constants for those classes) will further increase the provided level of security. This increase in security is shown in Fig, 4, which compares optimal strategies for three classes (solid line —) to uniform values (dotted line · · · ·).

In practice, the parameters of our model may be estimated in the following ways.

- Firstly, messages can be grouped into $C = 2$ classes, "high risk" and "low risk". Based on how detailed our estimations on the remaining parameters can be (see below), the number of classes can be increased, which further reduces the expected amount of losses.
- The traffic values T_c can either be computed from the application and network protocol specifications, or they can be estimated using traffic analysis. For example, one can measure the number of messages of class c in a time unit on a test system (or, if security will be added to a legacy system, even on a real system).
- The computational budget B arises from device resource constraints, which are obviously known at design time. Consequently, this parameter can easily be estimated as, for example, the number of hash computations that can be performed by the target device in a time unit.
- The potential loss values L_c can be quantified as financial damage to the system (e.g., cost of replacing damaged devices) or liability/penalties based on past incidents/settlements, resulting from successful message content manipulation. Note that only the *relative* values of L_c matter, as the results are scale invariant, which makes the setting of these parameters relatively easy for domain experts [4,16].
- Finally, the penalty F was primarily introduced for generality, since we show that the defender's optimal strategy is (essentially) independent of its value. More specifically, the defender's optimal strategy is completely independent of F in the continuous relaxation (see Theorem 4), and it is negligibly affected by F in the original model.

Once the parameter values have been estimated, the probabilities p_c can be computed at design and then loaded into the devices. Note that the p_c values can be stored the same way as the secret key that is used for MAC computation. Furthermore, the values can be stored simply as an array; hence, the computational cost of retrieving the values is negligible.

5.2 Stochastic message verification

We assume that we are given a defense strategy $\boldsymbol{p} \in [0, 1]^C$, an algorithm for determining the class of each received message, and an implementation of MAC verification, whose running time we would like to reduce. Then, stochastic message verification can be implemented easily as follows: for each message, choose a number rnd uniformly at random from $[0, 1]$; if rnd $\leq p_c$, where c is the class of the message, verify the message; otherwise, treat the message as authentic. Clearly, this simple algorithm implements the strategy described by our game-theoretic model.

5.2.1 Random number generation

The only non-trivial part of the implementation is the generation of random numbers. If the amount of true randomness

that is available to the receiver is limited, which is likely the case in most of the envisioned applications, we have to use a pseudorandom number generator (PRNG). This PRNG has to satisfy two requirements: first, its running time has to be less than what we save in computation due to stochastic verification; second, it has to withstand the adversary's attempts to deduce its state using the receiver as an oracle.

However, as the amount of randomness required by our scheme is an order of magnitude smaller than the data processed by a MAC computation, finding a suitable PRNG poses no real challenge. For example, if we generated the random numbers using a cryptographic hash function, the output of a single hash computation could provide enough randomness for hundreds of messages, while each verification would require a separate hash computation in a hash-based MAC scheme. Furthermore, the adversary can gain information regarding the state of the PRNG only when the receiver does not verify a modified message, which can happen with only $1 - p_c$ probability. Since the probability that the adversary remains undetected diminishes exponentially with the amount of information that she can gain, we can use a low-cost PRNG in the implementation (e.g., one based on linear feedback shift registers).

5.3 Stochastic generation by the sender

The stochastic verification scheme, which we have discussed so far, is straightforward to implement, but it requires the sender to compute a correct authentication tag for every outgoing message. Consequently, it can save computation only at the receiver's end of the communication. Now, we discuss a more complex scheme, called stochastic generation, which can save computation at both ends. In this scheme, the sender determines the class c of each outgoing message and decides—based on the probability p_c of the class- –whether to compute the correct authentication tag for the given message.

Implementing stochastic generation poses a number of challenges. First, consider a naïve implementation that sends unauthenticated messages without tags. Since an adversary can easily tell which messages are authenticated from the presence of the tag (or the lack of it), she is able to modify an arbitrary number of messages without taking any risk of being detected. To prevent such attacks, the sender has to attach a *fake tag* to unauthenticated messages, and she has to generate these fake tags in a way such that the adversary will not be able to tell the difference between fake and *correct tags*.

The following theorem establishes the security requirements that fake tags have to satisfy.

Theorem 9 *Assume that we are given a secure MAC scheme for generating and verifying authentication tags, which are indistinguishable from random numbers for the adversary. For each outgoing message of class c, the sender follows the following stochastic generation scheme: with probability p_c, attach a correct authentication tag, and with probability $1 - p_c$, attach a fake tag, where p_c is an optimal probability given by our game-theoretic model. Finally, assume that for messages with fake tags, there exist modifications that cannot be detected by the receiver.*

Then, the expected amount of losses cannot be higher than what is given by our game-theoretic model if and only if the following conditions are satisfied by the fake tags.

1. *The adversary must not be able to distinguish correct tags from fake tags.*
2. *The receiver must be able to distinguish fake tags from incorrect tags.*[4]
3. *The receiver must be able to detect modifications that cause more damage than what is allowed by the original class of the message.*[5]

Note that the first requirement implies that the messages have some varying content; otherwise, the adversary could learn the correct tag values for messages that are sent multiple times. This variability can easily be achieved using sequence numbers or timestamps, which are also used to prevent replay attacks.

Proof (Sketch.) We begin with proving the necessity of the requirements using proof by contradiction. More specifically, for each requirement, we show that if it is not satisfied by the fake tags, then there exists an attack that can achieve higher payoff than what is given by our game-theoretic model.

First, suppose that the adversary can distinguish fake tags from correct tags. Then, she can modify an arbitrary number of messages with fake tags, without modifying any message with a correct tag. Since there exist modifications to messages with fake tags that the receiver cannot detect, the probability of the attack remaining undetected is equal to one, regardless of the number of modified messages. Hence, the adversary can achieve arbitrarily high payoff with zero probability of detection.

Second, suppose that the receiver cannot distinguish fake tags from incorrect tags. Then, the adversary can modify an arbitrary number of messages with zero probability of detection, since the receiver can never tell whether a message was modified or if the sender chose to attach a fake tag to it.

[4] By incorrect, we mean that the message had a correct tag, but it was modified by the adversary.

[5] Note that any modification to a message with a correct tag can be detected; hence, this requirement is actually imposed on the fake tags only.

Third, suppose that the receiver cannot detect modifications that cause more damage than what is allowed by the class (i.e., for a message of class c, modifications that cause more than L_c damage). Recall that the class of a message is determined by how dangerous it is in the worst case, which depends on its contents. By modifying the contents of a message, the adversary can change the class to which it *should* belong. If the receiver cannot detect such modifications, then the adversary is able to modify messages from some class c with a low L_c value and cause more than L_c damage with only p_c probability of being detected.

Now, assume that the fake tags satisfy all three requirements. Then, we have to show that a message causing l damage is detected with probability 1 if $l > \max_c\{L_c\}$ and with probability p_i otherwise, where $i = \operatorname{argmin}_{c\,:\,L_c \geq l} p_c$. Suppose that the adversary modifies a message from some class c. If the damage caused by the modification is greater than L_c, then the receiver can detect it with probability one, since the class to which the message should belong is changed (follows from Requirement 3).

Otherwise, the modification is detected whenever the sender has attached a correct tag to the message, as the receiver can detect incorrect tags (follows from Requirement 2 and the assumption that the MAC scheme is secure). The probability of attaching a correct tag to a message of class c is p_c (independently of any other event), and since the adversary cannot distinguish correct tags from fake tags (follows from Requirement 1), she has no advantage over random guessing. Consequently, the modification is detected with probability p_c. □

To save computation, we have to generate fake tags, which are indistinguishable from correct tags and which authenticate the class to which a message should belong, at a much lower computational cost than correct tags. In the following section, we propose such a scheme.

5.3.1 Partial HMAC using Merkle–Damgård hash functions

Now, we propose an efficient scheme, called *partial HMAC*, for generating and verifying fake tags, as an example of implementing the stochastic generation scheme.

The main idea behind our partial HMAC scheme is that we can generate a fake tag satisfying all three requirements by computing a correct tag for only a portion of the message. It is easy to see that such a scheme can satisfy all three requirements.

- First, if the authenticated portion contains enough variability (e.g., a sequence number or a timestamp), then the adversary will not be able to tell fake tags from correct tags (Requirement 1), as both will be indistinguishable from random numbers.
- Second, if the authenticated portion contains an identifier of the class of the message as well, then the receiver can detect any modification which could cause more damage than what is allowed by the original class of the message (Requirement 3), since she can determine the class of a received message (based on its dangerousness) and compare it with the identifier.
- Finally, if the receiver knows which portion of the message is authenticated, then she can compute both the fake and the correct tags for each message and verify if any of them matches (Requirement 2).

The challenge lies in implementing this verification in an efficient way, which does not perform any unnecessary computation, since simply computing both the fake and the correct tag for each received message would actually increase the computational costs of the receiver.

Now, let us assume that the messages are authenticated using the HMAC construction based on a Merkle–Damgård hash function. The HMAC (keyed-hash message authentication code) is a construction for generating message authentication codes [3], defined as

$$\mathrm{HMAC}(K, \boldsymbol{m}) = H((K \oplus opad) \mid H((K \oplus ipad) \mid \boldsymbol{m})),$$

where K is the secret key, $\boldsymbol{m}$ is the message to be authenticated, H is a cryptographic hash function, $opad = \texttt{0x5c}\ldots\texttt{5c}$ is the outer padding, and $ipad = \texttt{0x36}\ldots\texttt{36}$ is the inner padding. The Merkle–Damgård construction is a method for building a cryptographic hash function from a collision-resistant one-way compression function [22]. Generally, a Merkle–Damgård hash function is computed as follows

$$H(\boldsymbol{m}) = f(f(\ldots f(f(\mathrm{IV}, m_1), m_2), \ldots), \text{length padding}),$$

where IV is an initialization vector (i.e., a constant given by the specification of the hash function), f is a one-way compression function, m_i is the ith block of the message, and length padding is an MD compliant padding.[6]

We can generate fake tags by stopping the HMAC computation after the very first iteration of the hash function, given that the first block of a message contains an identifier of the class and adequate amount of variability. Formally, fake tags can be generated using Algorithm 1.

[6] In some hash functions, the length padding does not take up a complete block, but this is not relevant to our approach.

Algorithm 1 MAC tag generation in partial HMAC

```
1: function GenerateTag(K, m)
2:     rnd ← U(0, 1)
3:     if rnd ≤ p_class(m) then
4:         return HMAC(m)
5:     else
6:         return f(f(IV, K ⊕ ipad), m_1)
7:     end if
8: end function
```

Then, the receiver can verify messages efficiently using Algorithm 2. Observe that this algorithm is optimal in the sense that it computes both fake and correct tags using the minimum amount of computation that is possible, as correct tags are computed by continuing the fake tag computation on demand. Also, note that the case of messages shorter than one block, whose handling requires only a very straightforward extension to the algorithms, is not considered for ease of presentation.

Algorithm 2 MAC tag verification in partial HMAC

```
1: function VerifyTag(K, m, t)
2:     t_f ← f(f(IV, K ⊕ ipad), m_1)
3:     if t = t_f then
4:         return fake
5:     else
6:         t_c ← H((K ⊕ opad) |
               f(f(...f(t_f, m_2), ..., m_n), length padding))
                                  = H(K⊕ipad | m)
7:         if t = t_c then
8:             return correct
9:         else
10:            return incorrect
11:        end if
12:    end if
13: end function
```

5.4 Experimental results

For the practical evaluation demonstrating the feasibility of our approach, we implemented our stochastic message authentication schemes using SHA-1 HMAC and a linear feedback shift register PRNG on an ATmega328P[7] microcontroller. Using this implementation, we performed experiments measuring the running time of our schemes for various authentication probabilities.

For the experiments, we used short messages, which fit into one hash block. Note that, for longer messages, the relative savings in the stochastic generation scheme are even greater. The measured running times generally include both the PRNG and the (partial) HMAC computations. However, to compare the overhead of the PRNG with the savings in computation due to stochastic authentication, we did not run the PRNG for $p = 1$. Finally, the running times obviously do not include any strategy computation, since that has to be performed at design time.

[7] http://www.atmel.com/devices/atmega328p.aspx

5.4.1 Stochastic verification

Figure 5a shows the average running time of stochastic MAC verification as a function of the verification probability. As expected, we see a clear linear relationship between the verification probability and the running time. Although this result seems trivial, it shows that the linear computational cost assumption of our model is valid. Finally, by comparing the data points for $p = 0.99$ and $p = 1$, we can see that the overhead of the PRNG is negligible.

5.4.2 Stochastic generation

Figure 5b and c shows the average running time of the partial HMAC generation (Algorithm 1) and verification (Algorithm 2), respectively, as functions of the authentication probability. The figure clearly shows that the linear computational cost assumption is valid in this scheme as well. Finally, we can see again that the overhead of the PRNG is negligible by comparing the data points for $p = 0.99$ and $p = 1$ in Fig. 5b and c.

By comparing Fig. 5a with b and c, we can see that, at $p = 0$ probability, the running time is zero in the stochastic verification scheme, but it is nonzero in the partial HMAC scheme for both the sender and the receiver. This nonzero running time is due to the computational cost of generating and verifying fake tags, which is non-negligible in the partial HMAC scheme. However, the sum saving (i.e., if we add together the savings in running time at the receiver and the sender) in the partial HMAC scheme is at least as great as in the stochastic verification scheme (and it is strictly greater for longer messages). Consequently, the decision which scheme to use between two nodes depends on both the amount of traffic in each direction and the computational constraints of the two nodes.

6 Related work

To the best of our knowledge, there has been very little research on modeling message authentication using game theory. In [25] and [26], Simmons formulates a game-theoretic model of the contest between the sender, the receiver, and the adversary, in order to study message authentication on a noisy channel; however, the author does not consider any resource bounds. Game theory has been used more generally in security, in attacker–defender games [15,21,28]; for example, it can be used to study the optimal interdiction of attack plans [19].

Several research efforts have tried to provide lightweight cryptographic primitives and mechanisms for resource-bounded systems [2,6,8,20,23,24]. Note that our approach is complementary to these results, since we build on an

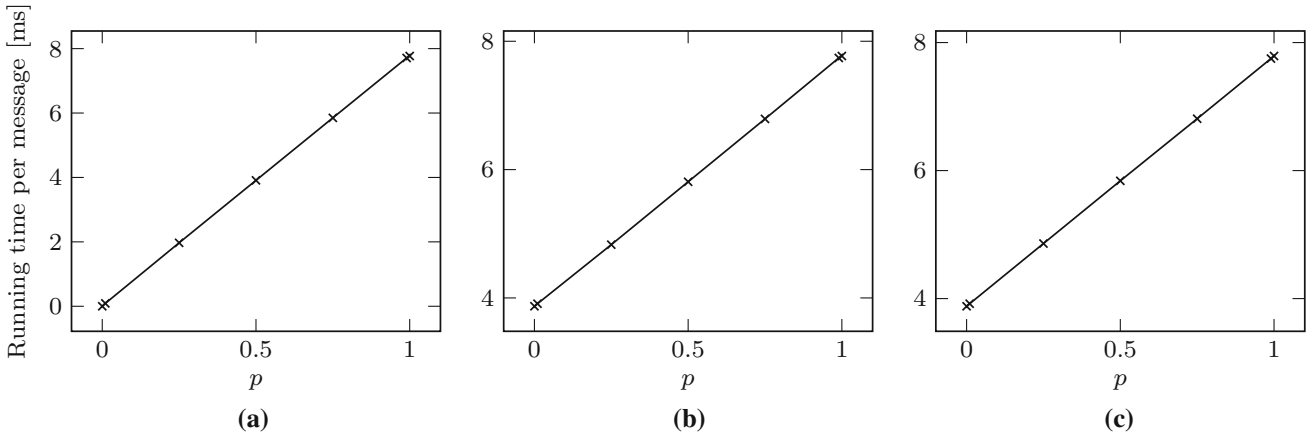

Fig. 5 Average running time per message as a function of authentication probability for **a** stochastic verification, **b** partial HMAC generation (Algorithm 1), and **c** partial HMAC verification (Algorithm 2). Each *cross* marks a measured value. Note that no PRNG was used for $p = 1$

existing MAC scheme to provide optimal authentication for an arbitrary resource bound, while the majority of the literature is concerned with designing new primitives. For example, Gong et al. introduce a new family of lightweight block ciphers named KLEIN, which are designed to be usable as building blocks for security in resource-constrained devices [14]. As another example, Engels et al. propose the Hummingbird and Hummingbird-2 encryption algorithms, which are targeted for low-end microcontrollers [9,10]. Besides lightweight primitives, researchers have also proposed mechanisms for securing various resource-constrained systems. For example, Fouda et al. propose a lightweight message authentication scheme for smart grid communications [12]. In their proposed scheme, the smart meters that are distributed at different hierarchical networks first achieve mutual authentication and establish a shared session key with the Diffie–Hellman protocol and then authenticate the subsequent messages in a lightweight way using hash-based authentication codes. As another example, Kumar and Aggarwal combine lightweight cryptographic primitives for securing ad hoc networks [17]. Finally, Tsang and Smith circumvent the problem of resource-bounded systems by deploying additional hardware modules into the communication link. More specifically, they construct a bump-in-the-wire (BITW) solution that retrofits security into time-critical communications over bandwidth-limited serial links between devices in supervisory control and data acquisition (SCADA) systems [29]. Note that this is complementary to our approach since they propose a specific scheme, while we propose a general purpose approach that can be used with any specific message authentication scheme, including that of Tsang and Smith.

Beyond message authentication, our idea and theoretical results for stochastic authentication could also be to applied to other authentication problems in resource-bounded systems. For example, in hierarchical wireless sensor networks, strict constraints on computational capabilities and energy consumption may pose challenges to the design of secure remote user authentication schemes for real-time data access [30]. A number of lightweight user authentication schemes have been proposed for such systems; however, some of these have already been shown to suffer from severe weaknesses [30]. As an alternative to novel lightweight authentication schemes, our stochastic authentication approach could be adapted for user authentication in order to provide some provable level of security based on existing non-lightweight schemes. Further, our approach could also be combined with novel user authentication schemes (e.g., [31]) to enhance them with the capability of smooth trade-off between security and computational cost.

7 Conclusion

In this paper, we proposed the stochastic authentication of messages in order to save computation, while maintaining a level of integrity and authenticity protection for the messages. We formulated the problem as a game-theoretic model, and we studied the adversary's best-response and the defender's Stackelberg and Nash equilibrium strategies. We showed that optimal authentication strategies can substantially outperform naïve strategies. We also showed that a continuous relaxation of the problem can be used to find authentication strategies for computationally challenging instances. Finally, we characterized the defender's best-response strategies and the existence of Nash equilibria.

Then, we studied the problem of implementing stochastic message authentication in practice, given that we have a solution (i.e., a vector of probabilities) from our theoretical model. First, we considered a simpler scheme, called stochastic verification, which implements our model in a straightforward way and which decreases the computational

cost of only the receiver. Second, we proposed a more complex scheme, called stochastic generation, which decreases the computational cost of both the sender and the receiver. Finally, we presented experimental results on the performance of our schemes, which showed that our approach is feasible in practice.

Our approach has two important advantages. Firstly, it provides a smooth trade-off between security and reduction in computational costs. Thus, we can apply it to an arbitrary resource-bounded device and attain the maximum level of security that is feasible for a given scheme. Secondly, our approach can be based on standardized and trusted cryptographic primitives. This is advantageous because we do not have to place trust in a novel cryptographic primitive, which has not been thoroughly field tested.

7.1 Future work

In this paper, we assumed potential loss to be a linear function of the set of modified messages. In future work, we plan to extend our model to consider other, nonlinear functions as well. Naturally, all of the implementation results presented in Sect. 5 hold regardless of the loss function. More generally, the principle of stochastic message authentication applies regardless of the loss function. As for our game-theoretic analysis, the validity of our results depends on the properties of the loss function. While the analysis depends on the loss function, we can generalize some of our results in a fairly straightforward manner for broad classes of functions. For example, many of our theoretical results hold for submodular loss functions (e.g., our results on deterrence strategies).

Acknowledgements This work is supported in part by the National Science Foundation (CNS-1238959), the Air Force Research Laboratory (FA 8750-14-2-0180), and by NIST (70NANB15H263).

Appendix: Proofs

Proof of Lemma 1

Proof Recall that the adversary's best response maximizes the payoff function $\mathcal{U}_A(\boldsymbol{a}) = \prod_{c=1}^{C}(1-p_c)^{a_c}\left(F+\sum_{c=1}^{C}a_cL_c\right)-F$.

First, suppose that $p_i = 1$ holds for some class i. Then, by definition, the adversary's payoff is $-F$ for every strategy $\boldsymbol{a}$ where $a_i > 0$, as the probability of detection is 1. Since the adversary can always choose not to attack, which achieves zero payoff, $a_i = 0$ has to hold obviously for any best-response strategy $\boldsymbol{a}$. Consequently, for the remainder of this proof, we can disregard classes i with $p_i = 1$ and assume that $\boldsymbol{p} < \mathbf{1}$.

Then, the first-order partial derivative of the adversary's payoff function $\mathcal{U}_A$ with respect to some a_i is

$$\frac{\partial}{\partial a_i}\mathcal{U}_A(\boldsymbol{a}) \tag{32}$$

$$= \frac{\partial}{\partial a_i}(1-p_i)^{a_i}\prod_{c\neq i}(1-p_c)^{a_c}\left(F+a_iL_i+\sum_{c\neq i}a_cL_c\right) \tag{33}$$

$$= \ln(1-p_i)(1-p_i)^{a_i}\prod_{c\neq i}(1-p_c)^{a_c}\left(F+a_iL_i+\sum_{c\neq i}a_cL_c\right) + (1-p_i)^{a_i}\prod_{c\neq i}(1-p_c)^{a_c}L_i \tag{34}$$

$$= \prod_{c}(1-p_c)^{a_c}\left[\ln(1-p_i)\left(F+\sum_{c=1}^{C}a_cL_c\right)+L_i\right]. \tag{35}$$

To find the maximum of the payoff function (i.e., the adversary's best response), we set the first derivative (with respect to a_i) equal to zero and solve the resulting equality as follows

$$0 = \underbrace{\prod_{c}(1-p_c)^{a_c}}_{>0}\left[\ln(1-p_i)\left(F+\sum_{c=1}^{C}a_cL_c\right)+L_i\right] \tag{36}$$

$$0 = \ln(1-p_i)\left(F+\sum_{c=1}^{C}a_cL_c\right)+L_i \tag{37}$$

$$0 = -\frac{L_i}{\ln(1-p_i)} - F - \sum_{c=1}^{C}a_cL_c. \tag{38}$$

When the payoff function $\mathcal{U}_A$ attains its maximum, then for every variable a_i, either the first-order partial derivative with respect to variable a_i must be zero (i.e., the above equation must hold for i) or the variable a_i must be at an endpoint. Since the only constraint on the adversary's strategy is $\boldsymbol{a} \geq 0$, the only endpoint for variable a_i is 0. Hence, in a best-response strategy, for every class i, either Eq. (38) or $a_i = 0$ must hold. □

Proof of Lemma 2

Proof (Sketch.) Note that $\ln(1-p_i)$ and $\ln(1-p_j)$ are both negative. Hence, we have $\frac{L_i}{L_j} \leq \frac{\ln(1-p_i)}{\ln(1-p_j)}$.

First, we show that the sum $\sum_{c=1}^{C}a_cL_c$ remains the same after we decrease a_i by Δ and increase a_j by $\Delta\frac{L_i}{L_j}$. To see this, consider the sum of the terms belonging to i and j in the modified strategy, which is

$$(a_i-\Delta)L_i + \left(a_j+\Delta\frac{L_i}{L_j}\right)L_j \tag{39}$$

$$= a_i L_i - \Delta L_i + a_j L_j + \Delta L_i \tag{40}$$
$$= a_i L_i + a_j L_j. \tag{41}$$

Since the remaining terms are not changed, the sum $\sum_{c=1}^{C} a_c L_c$ also has to remain the same.

Second, we show that the product $\prod_{c=1}^{C}(1-p_c)^{a_c}$ does not increase after we decrease a_i by Δ and increase a_j by $\Delta\frac{L_i}{L_j}$. To see this, consider the product of the factors belonging to i and j in the modified strategy, which is

$$(1-p_i)^{a_i-\Delta}(1-p_j)^{a_j+\Delta\frac{L_i}{L_j}} \tag{42}$$
$$= (1-p_i)^{a_i}(1-p_i)^{-\Delta}(1-p_j)^{a_j}(1-p_j)^{\Delta\frac{L_i}{L_j}} \tag{43}$$
$$= (1-p_i)^{a_i}e^{-\ln(1-p_i)\Delta}(1-p_j)^{a_j}e^{\ln(1-p_j)\Delta\frac{L_i}{L_j}} \tag{44}$$
$$\leq (1-p_i)^{a_i}e^{-\ln(1-p_i)\Delta}(1-p_j)^{a_j}e^{\ln(1-p_j)\Delta\frac{\ln(1-p_i)}{\ln(1-p_j)}} \tag{45}$$
$$= (1-p_i)^{a_i}e^{-\ln(1-p_i)\Delta}(1-p_j)^{a_j}e^{\ln(1-p_i)\Delta} \tag{46}$$
$$= (1-p_i)^{a_i}(1-p_j)^{a_j}. \tag{47}$$

Since the remaining factors are not changed, the product $\prod_{c=1}^{C}(1-p_c)^{a_c}$ does not increase either. By combining the above equality and inequality with the definition of the adversary's payoff, we see that the payoff cannot decrease when we decrease a_i by Δ and increase a_j by $\Delta\frac{L_i}{L_j}$.

Finally, observe that (45) is a strict inequality if and only if the inequality between the ratios is strict. Therefore, the product $\prod_{c=1}^{C}(1-p_c)^{a_c}$, and consequently, the adversary's payoff strictly increases if and only if the inequality between the ratios is strict. □

Proof of Lemma 3

Proof For the ease of presentation, we let L denote L_1, a denote a_1, and p denote p_1 in this proof. Using this notation, in the special case of $C = 1$, the adversary's payoff function can be expressed simply as $\mathcal{U}_A(a) = (1-p)^a(F+aL) - F$.

The first derivative of the payoff function $\mathcal{U}_A(a)$ with respect to a is

$$\frac{\mathrm{d}}{\mathrm{d}a}\mathcal{U}_A(a) = \ln(1-p)(1-p)^a(F+aL) + (1-p)^a L - 0 \tag{48}$$
$$= (1-p)^a(\ln(1-p)(F+aL) + L). \tag{49}$$

To find the maximum of the payoff function $\mathcal{U}_A(a)$, we set the first derivative equal to zero and solve for a:

$$0 = \underbrace{(1-p)^a}_{>0}(\ln(1-p)(F+aL) + L) \tag{50}$$
$$0 = \ln(1-p)(F+aL) + L \tag{51}$$
$$a\ln(1-p)L = -\ln(1-p)F - L \tag{52}$$
$$a = -\frac{1}{\ln(1-p)} - \frac{F}{L}. \tag{53}$$

If the adversary's strategy a were continuous, then the maximum of the objective function would be attained at either the endpoint (i.e., $a = 0$) or where the first derivative is zero (i.e., the unique solution of the above equation). Consequently, if the solution of the above equation, denoted by a^*, is positive, then the best integer response is either $\lfloor a^* \rfloor$ or $\lceil a^* \rceil$ (or both). Otherwise, zero (i.e., not attacking) is the unique best-response strategy. □

Proof of Lemma 4

Proof For the sake of contradiction, suppose that the claim of the lemma does not hold, that is, suppose that there exist a_c^* and $\hat{\boldsymbol{a}}$ such that a_c^* is the maximal single-class best response for some class c and $\hat{\boldsymbol{a}}$ is a best response, but $\hat{a}_c > a_c^*$. For the remainder of the proof, let

$$\hat{L} = \sum_{i \neq c} \hat{a}_i L_i$$

and

$$\hat{P} = \prod_{i \neq c}(1-p_i)^{\hat{a}_i}.$$

First, if $\hat{L}$ were zero, then $\hat{\boldsymbol{a}}$ would also be a single-class best response, since its only nonzero element would be $\hat{a}_c$. However, this would contradict our initial supposition that a_c^* is the maximal single-class best response for class c. Consequently, $\hat{L} > 0$ has to hold. Then, it follows readily from $\hat{a}_c > a_c^*$ that

$$\frac{\hat{a}_c L_c + F}{a_c^* L_c + F} > \frac{\hat{a}_c L_c + F + \hat{L}}{a_c^* L_c + F + \hat{L}} \tag{54}$$
$$\frac{\hat{a}_c L_c + F}{a_c^* L_c + F}(a_c^* L_c + \hat{L} + F) > \hat{a}_c L_c + \hat{L} + F. \tag{55}$$

Since a_c^* is a single-class best response, the adversary's payoff for modifying a_c^* messages from class c must be higher than for modifying $\hat{a}_c$ messages (given that it does not modify messages from other classes) by definition. Hence, we have

$$(1-p_c)^{a_c^*}(a_c^* L_c + F) \geq (1-p_c)^{\hat{a}_c}(\hat{a}_c L_c + F) \tag{56}$$
$$(1-p_c)^{a_c^*} \geq (1-p_c)^{\hat{a}_c}\frac{\hat{a}_c L_c + F}{a_c^* L_c + F}. \tag{57}$$

Now, consider the strategy which modifies $\hat{a}_i$ messages for classes $i \neq c$, and a_c^* messages of class c. The adversary's payoff for this strategy is

$$
\begin{aligned}
&(1-p_c)^{a_c^*} \prod_{i \neq c} (1-p_i)^{\hat{a}_i} \left(a_c^* L_c + \sum_{i \neq c} \hat{a}_i L_i + F \right) \\
&= (1-p_c)^{a_c^*} \hat{P}(a_c^* L_c + \hat{L} + F) && (58) \\
&\geq (1-p_c)^{\hat{a}_c} \frac{\hat{a}_c L_c + F}{a_c^* L_c + F} \hat{P}(a_c^* L_c + \hat{L} + F) && (59) \\
&> (1-p_c)^{\hat{a}_c} \hat{P}(\hat{a}_c L_c + \hat{L} + F) && (60) \\
&= \mathcal{U}_A(\boldsymbol{p}, \hat{\boldsymbol{a}}). && (61)
\end{aligned}
$$

Note that, for the inequality, we used Eq. (57), and for the strict inequality, we used Eq. (55).

These inequalities show that the adversary's payoff for the strategy constructed above is strictly higher than for strategy $\hat{\boldsymbol{a}}$. However, this contradicts our initial assumption that $\hat{\boldsymbol{a}}$ is a best-response strategy; therefore, the claim of the lemma has to hold. □

Proof of Theorem 4

Proof (Sketch.) First, we show that the ratios have to be uniform. For the sake of contradiction, suppose that the claim does not hold for some optimal defense strategy. Then, from Theorem 1, we have that the adversary will attack only the classes with minimal ratios. Furthermore, it is easy to see that that the defender can increase the probabilities of the classes with minimal ratios and decrease the probabilities of the classes with maximal ratios, without changing the set of adversarial best responses or the sum computational cost. Hence, the defender can strictly decrease her loss, which contradicts the supposition that the original strategy is optimal. Therefore, the original claim must hold (i.e., the ratios have to uniform).

Second, we show that an optimal strategy uses all of the budget. Since we already have that the ratios are uniform, we have that all the classes are "payoff-equivalent" (see the adversary's best response in the relaxed model). Consequently, it suffices to show that $pT = B$ is optimal for the case of a single class. Since the adversary always plays a best response, it will modify $a^* = -\frac{1}{\ln(1-p)} - \frac{F}{L}$ messages, and we can compute the defender's loss for any strategy p as

$$
\begin{aligned}
&(1-p)^{-\frac{1}{\ln(1-p)} - \frac{F}{L}} \left(-\frac{1}{\ln(1-p)} - \frac{F}{L} \right) L && (62) \\
&= \frac{L}{e} (1-p)^{-\frac{F}{L}} \left(-\frac{1}{\ln(1-p)} - \frac{F}{L} \right). && (63)
\end{aligned}
$$

The first derivative of the defender's loss with respect to p is

$$
\begin{aligned}
&\frac{d}{dp} \frac{L}{e} (1-p)^{-\frac{F}{L}} \left(-\frac{1}{\ln(1-p)} - \frac{F}{L} \right) && (64) \\
&= \frac{L}{e} \left(-\frac{F}{L} (1-p)^{-\frac{F}{L}-1} \left(-\frac{1}{\ln(1-p)} - \frac{F}{L} \right) \right. \\
&\quad \left. - (1-p)^{-\frac{F}{L}} \frac{1}{(1-p)\ln^2(1-p)} \right) && (65) \\
&= \underbrace{\frac{L}{e} (1-p)^{-\frac{F}{L}-1}}_{>0} \left(-\frac{F}{L} \underbrace{\left(-\frac{1}{\ln(1-p)} - \frac{F}{L} \right)}_{=a^* \geq 0} - \underbrace{\frac{1}{\ln^2(1-p)}}_{>0} \right) && (66) \\
&< 0. && (67)
\end{aligned}
$$

Since the first derivative is negative, the minimal loss is attained at the maximal feasible probability (i.e., at the budget limit).

It remains to show that a unique strategy satisfying both conditions exists. First, observe that each ratio $\frac{L_c}{\ln(1-p_c)}$ is a strictly monotonic continuous function of the corresponding probability p_c. Consequently, for any $R \in \mathbb{R}_{<0}$, there always exists a unique vector of probabilities $\boldsymbol{p}$ that satisfies $\frac{L_c}{\ln(1-p_c)} = R$ for every class c. Furthermore, the weighted sum $\sum_c p_c T_c$ of these probabilities is also a strictly monotonic continuous function of R. Consequently, for every budget B, there has to exist a unique defense strategy $\boldsymbol{p}$ that satisfies both $\frac{L_1}{\ln(1-p_1)} = \cdots = \frac{L_C}{\ln(1-p_C)}$ and $\sum_c p_c T_c = B$. □

Proof of Lemma 5

Proof Let $\boldsymbol{a}^*$ and $\boldsymbol{a}'$ be best responses against $\boldsymbol{p}^*$ and $\boldsymbol{p}'$, respectively. If $\boldsymbol{a}' = \boldsymbol{0}$ is true, then the claim of the lemma obviously holds, since $\boldsymbol{p}^*$ does not deter the adversary while $\boldsymbol{p}'$ does. Hence, we only have to show that the claim of the lemma holds for the case where $\boldsymbol{a}' \neq \boldsymbol{0}$ (i.e., for the remainder of the proof, we can assume $\boldsymbol{a}' \neq \boldsymbol{0}$).

For the remainder of the proof, let P^* denote $\prod_i (1-p_i^*)^{a_i^*}$, let P' denote $\prod_i (1-p_i')^{a_i'}$, let L^* denote $\sum_i a_i^* L_i$, and let L' denote $\sum_i a_i' L_i$. Furthermore, for any $\boldsymbol{p}$ and $\boldsymbol{a}$, let $\tilde{\mathcal{U}}_A(\boldsymbol{p}, \boldsymbol{a})$ denote $\mathcal{U}_A(\boldsymbol{p}, \boldsymbol{a}) + F = \prod_c (1-p_c)^{a_c} \left(\sum_c a_c L_c + F \right)$.

First, since both $\boldsymbol{a}^*$ and $\boldsymbol{a}'$ are best responses, we have

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}^*) \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}') \quad (68)$$

and

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}') \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}^*). \quad (69)$$

Second, observe that $\boldsymbol{p}' > \boldsymbol{p}^*$ follows from the condition of the lemma. Then, using the definition of the adversary's

payoff, we have

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}') > \tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}'). \tag{70}$$

By combining these inequalities, we get

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}^*) \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}') > \tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}') \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}^*)\,, \tag{71}$$

which implies that

$$\frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}^*)}{\tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}^*)} \geq \frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}')}{\tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}')}. \tag{72}$$

Using the definition of the adversary's payoff, we can express these fractions as

$$\frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}^*)}{\tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}^*)} = \prod_i \left(\frac{1 - p_i^*}{1 - p_i'}\right)^{a_i^*} \tag{73}$$

and

$$\frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}')}{\tilde{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}')} = \prod_i \left(\frac{1 - p_i^*}{1 - p_i'}\right)^{a_i'}. \tag{74}$$

By substituting these fractions into the previous inequality, we get

$$\prod_i \left(\frac{1 - p_i^*}{1 - p_i'}\right)^{a_i^*} \geq \prod_i \left(\frac{1 - p_i^*}{1 - p_i'}\right)^{a_i'} \tag{75}$$

$$\sum_i a_i^* \ln \frac{1 - p_c^*}{1 - p_c'} \geq \sum_i a_i' \ln \frac{1 - p_c^*}{1 - p_c'} \tag{76}$$

$$\sum_i a_i^* L_i \geq \sum_i a_i' L_i \tag{77}$$

$$L^* \geq L'. \tag{78}$$

Now, for the sake of contradiction, suppose that the claim of the lemma does not hold, that is, suppose that $P'L' \geq P^*L^*$. By combining this with Eq. (78), we get

$$P' \geq P^* \tag{79}$$

$$P'F \geq P^*F \tag{80}$$

$$P'L' + P'F \geq P^*L^* + P^*F \tag{81}$$

$$P'(L' + F) \geq P^*(L^* + F) \tag{82}$$

$$\hat{\mathcal{U}}_A(\boldsymbol{p}', \boldsymbol{a}') \geq \hat{\mathcal{U}}_A(\boldsymbol{p}^*, \boldsymbol{a}^*). \tag{83}$$

However, this contradicts Eq. (71). Therefore, the claim of the lemma has to hold. □

Proof of Theorem 6

Proof (Sketch.) For the sake of contradiction, suppose that the claim of the theorem does not hold for some $\boldsymbol{p}^*$. Then, we can construct a strictly better strategy $\boldsymbol{p}'$ using the excess budget and Lemma 5 the same way as in the proof of Theorem 5. However, this contradicts the assumption that $\boldsymbol{p}^*$ is optimal; hence, the claim of the theorem has to hold. □

Proof of Theorem 7

Proof Clearly, if the defender's payoff is nonzero, then there must be at least one class c with $a_c > 0$. Furthermore, as the probability of detection cannot be 1, we also have that $p_c < 1$ for every class c with $a_c > 0$. Consequently, the verification probability has to be zero for every c with $a_c = 0$; otherwise, the defender could increase the detection probability (and, hence, her payoff) by decreasing the verification probability of some class i having $a_i = 0$ and increasing the verification probability of some other class j having $a_j > 0$. Since every class c with $a_c = 0$ must have zero verification probability (i.e., $p_c = 0$), we will disregard them for the remainder of the proof and focus only on the remaining classes.

Now, suppose that all the elements of the defender's strategy are given except for two classes. Without loss of generality, assume that the two variable elements are p_1 and p_2. To prove the claim, we have to show that either $p_1 = 0$ or $p_2 = 0$ in a best response. Let the budget remaining for these two elements be denoted by B^* (i.e., $B^* = B - \boldsymbol{p}'\boldsymbol{T}$). Then, $p_1 \in [0, \frac{B^*}{T_1}]$ and $p_2 = \frac{B^* - T_1 p_1}{T_2}$.

If $\boldsymbol{p}$ is a best response, then it minimizes the probability of not detecting an attack. Hence, p_1 has to minimize

$$(1 - p_1)^{a_1}(1 - p_2)^{a_2} = (1 - p_1)^{a_1}\left(1 - \frac{B^*}{T_2} + \frac{T_1}{T_2}p_1\right)^{a_2}. \tag{84}$$

The minimum is attained at either $p_1 = 0$, $p_1 = \frac{B^*}{T_1}$, or where the first derivative with respect to p_1 is zero. We show that only $p_1 = 0$ or $p_1 = \frac{B^*}{T_1}$ can actually be a minimum.

First, assume that $a_1, a_2 > 1$. Then, the first derivative of the above expression is

$$\begin{aligned} &- a_1(1 - p_1)^{a_1 - 1}\left(1 - \frac{B^*}{T_2} + \frac{T_1}{T_2}p_1\right)^{a_2} \\ &+ (1 - p_1)^{a_1} a_2 \frac{T_1}{T_2}\left(1 - \frac{B^*}{T_2} + \frac{T_1}{T_2}p_1\right)^{a_2 - 1}. \end{aligned} \tag{85}$$

Where the first derivative is equal to zero, we have

$$a_1\left(1 - \frac{B^*}{T_2} + \frac{T_1}{T_2}p_1\right) = a_2\frac{T_1}{T_2}(1 - p_1). \tag{86}$$

The solution for p_1 is a local maximum only if the second derivative in that point is positive. The second derivative is

$$\begin{aligned} & a_1(a_1-1)(1-p_1)^{a_1-2}\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2} \\ & \quad - a_1(1-p_1)^{a_1-1}a_2\frac{T_1}{T_2}\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2-1} \\ & \quad - a_1(1-p_1)^{a_1-1}a_2\frac{T_1}{T_2}\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2-1} \\ & \quad + (1-p_1)^{a_1}a_2(a_2-1)\frac{T_1^2}{T_2^2}\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2-2} \end{aligned} \tag{87}$$

$$\begin{aligned} & = a_1\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right) \\ & \quad \times\left[(a_1-1)\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)-a_2\frac{T_1}{T_2}(1-p_1)\right] \\ & \quad + a_2\frac{T_1}{T_2}(1-p_1) \\ & \quad \times\left[(a_2-1)\frac{T_1}{T_2}(1-p_1)-a_1\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)\right]. \end{aligned} \tag{88}$$

By substituting the solution of the first derivative into the above equation, we get

$$\begin{aligned} & \underbrace{a_1\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)}_{>0}\left[(-1)\underbrace{\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)}_{>0}\right] \\ & \quad + \underbrace{a_2\frac{T_1}{T_2}(1-p_1)}_{>0}\left[(-1)\underbrace{\frac{T_1}{T_2}(1-p_1)}_{>0}\right] < 0. \end{aligned} \tag{89}$$

Hence, any extremum between 0 and $\frac{B^*}{T_1}$ has to be a local maximum.

It remains to show the same for $a_1 = 1$ or $a_2 = 1$. For $a_1 = a_2 = 1$, the probability to be minimized is

$$(1-p_1)\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right) \tag{90}$$

$$= 1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1-p_1+p_1\frac{B^*}{T_2}-\frac{T_1}{T_2}p_1^2\,, \tag{91}$$

its first derivative is

$$\frac{T_1}{T_2}-1+\frac{B^*}{T_2}-2\frac{T_1}{T_2}p_1\,, \tag{92}$$

and the second derivative is

$$-2\frac{T_1}{T_2} < 0. \tag{93}$$

Hence, any extremum between 0 and $\frac{B^*}{T_1}$ has to be a local maximum.

For $a_1 = 1$ and $a_2 > 1$, the probability to be minimized is

$$(1-p_1)\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2}, \tag{94}$$

its first derivative is

$$\begin{aligned} & -1\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2} \\ & + (1-p_1)a_2\frac{T_1}{T_2}\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2-1}, \end{aligned} \tag{95}$$

and its second derivative is

$$\begin{aligned} & \underbrace{a_2\frac{T_1}{T_2}\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)^{a_2-2}}_{>0} \\ & \quad \times\left[\underbrace{-2\left(1-\frac{B^*}{T_2}+\frac{T_1}{T_2}p_1\right)+(1-p_1)(a_2-1)\frac{T_1}{T_2}}_{<\text{ first derivative }=0}\right] < 0. \end{aligned} \tag{96}$$

Finally, for the same reasons, the minimum can be attained only at the endpoints of the interval in the case of $a_1 > 1$ and $a_2 = 1$ as well.

Since the above holds for any pair of elements p_i and p_j, the claim of the theorem has to hold. □

References

1. Akerberg, J., Gidlund, M., Bjorkman, M.: Future research challenges in wireless sensor and actuator networks targeting industrial automation. In: Proceedings of the 9th IEEE International Conference on Industrial Informatics (INDIN), pp. 410–415 (2011)
2. Andreeva, E., Bilgin, B., Bogdanov, A., Luykx, A., Mennink, B., Mouha, N., Yasuda, K.: APE: Authenticated permutation-based encryption for lightweight cryptography. In: Proceedings of the 21st International Workshop on Fast Software Encryption (FSE), pp. 168–186. Springer (2014)
3. Bellare, M., Canetti, R., Krawczyk, H.: Keying hash functions for message authentication. In: Proceedings of the 16th Annual Crypto Conference (CRYPTO), pp. 1–15 (1996)
4. Campbell, K., Gordon, L.A., Loeb, M.P., Zhou, L.: The economic cost of publicly announced information security breaches: empirical evidence from the stock market. J. Comput. Secur. **11**(3), 431–448 (2003)

5. Campbell, R.J.: The smart grid and cybersecurity: regulatory policy and issues. Congressional Research Service Report for Congress. http://fas.org/sgp/crs/misc/R41886.pdf (2011). Accessed 01 May 2015
6. Cárdenas, A.A., Amin, S., Sastry, S.: Research challenges for the security of control systems. In: Proceedings of the 3rd USENIX Workshop on Hot Topics in Security (HotSec) (2008)
7. Cisco Systems, Inc.: Securing the smart grid. White Paper. http://www.cisco.com/web/strategy/docs/energy/SmartGridSecurity_wp.pdf (2009). Accessed 01 May 2015
8. Eisenbarth, T., Kumar, S., Paar, C., Poschmann, A., Uhsadel, L.: A survey of lightweight-cryptography implementations. IEEE Des. Test Comput. **24**(6), 522–533 (2007). doi:10.1109/MDT.2007.178
9. Engels, D., Fan, X., Gong, G., Hu, H., Smith, E.M.: Hummingbird: ultra-lightweight cryptography for resource-constrained devices. In: Proceedings of the 14th International Conference on Financial Cryptography and Data Security (FC), pp. 3–18. Springer (2010)
10. Engels, D., Saarinen, M.J.O., Schweitzer, P., Smith, E.M.: The Hummingbird-2 lightweight authenticated encryption algorithm. In: Proceedings of the 7th International Workshop, RFIDSec, Revised selected papers, pp. 19–31 (2011)
11. Fang, X., Misra, S., Xue, G., Yang, D.: Smart gridthe new and improved power grid: a survey. IEEE Commun. Surv. Tutor. **14**(4), 944–980 (2012)
12. Fouda, M.M., Fadlullah, Z.M., Kato, N., Lu, R., Shen, X.: A lightweight message authentication scheme for smart grid communications. IEEE Trans. Smart Grid **2**(4), 675–685 (2011)
13. Ghena, B., Beyer, W., Hillaker, A., Pevarnek, J., Halderman, J.A.: Green lights forever: analyzing the security of traffic infrastructure. In: Proceedings of the 8th USENIX Workshop on Offensive Technologies (WOOT'14). USENIX Association (2014)
14. Gong, Z., Nikova, S., Law, Y.W.: KLEIN: A new family of lightweight block ciphers. In: Proceedings of the 7th Workshop on RFID Security and Privacy (RFIDSec), Revised selected papers, pp. 1–18 (2011)
15. Korzhyk, D., Yin, Z., Kiekintveld, C., Conitzer, V., Tambe, M.: Stackelberg vs. Nash in security games: an extended investigation of interchangeability, equivalence, and uniqueness. J. Artifi. Intell. Res. **41**(2), 297–327 (2011)
16. Krutz, R.L., Vines, R.D.: The CISSP Prep Guide: Mastering the ten domains of Computer Security. Wiley, New York (2001)
17. Kumar, A., Aggarwal, A.: Lightweight cryptographic primitives for mobile ad hoc networks. In: Proceedings of the 2012 International Conference on Security in Computer Networks and Distributed Systems (SNDS), pp. 240–251 (2012)
18. Laszka, A., Vorobeychik, Y., Koutsoukos, X.D.: Integrity assurance in resource-bounded systems through stochastic message authentication. In: Proceedings of the 2nd Symposium and Bootcamp on the Science of Security, (HotSoS), pp. 1–12 (2015)
19. Letchford, J., Vorobeychik, Y.: Optimal interdiction of attack plans. In: Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS), pp. 199–206 (2013)
20. Maimut, D., Ouafi, K.: Lightweight cryptography for RFID tags. IEEE Secur. Priv. **10**(2), 76–79 (2012)
21. Manshaei, M.H., Zhu, Q., Alpcan, T., Bacşar, T., Hubaux, J.P.: Game theory meets network security and privacy. ACM Comput. Surv. (CSUR) **45**(3), 25 (2013)
22. Merkle, R.C.: Secrecy, authentication, and public key systems. Ph.D. thesis, Stanford University, Stanford (1979)
23. Moradi, A., Poschmann, A.: Lightweight cryptography and DPA countermeasures: a survey. In: Proceedings of the 1st International Workshop on Lightweight Cryptography for Resource-Constrained Devices (WLC), pp. 68–79 (2010)
24. Ranasinghe, D.C.: Lightweight cryptography for low cost RFID. In: Networked RFID Systems and Lightweight Cryptography, pp. 311–346. Springer, Berlin (2008)
25. Simmons, G.J.: Game theory model of digital message authentication. Tech rep., Sandia National Labs, Albuquerque (1981)
26. Simmons, G.J.: Authentication theory/coding theory. In: Blakley, G.R., Chaum, D. (eds.) Advances in Cryptology. CRYPTO 1984. Lecture Notes in Computer Science, vol. 196, pp. 411–431. Springer, Berlin, Heidelberg (1985)
27. Sridhar, S., Hahn, A., Govindarasu, M.: Cyber-physical system security for the electric power grid. Proc. IEEE **100**(1), 210–224 (2012)
28. Tambe, M.: Security and Game Theory: Algorithms, Deployed Systems, Lessons Learned. Cambridge University Press, Cambridge (2011)
29. Tsang, P.P., Smith, S.W.: YASIR: A low-latency, high-integrity security retrofit for legacy SCADA systems. In: Proceeding of the IFIP TC 11 23rd International Information Security Conference (IFIP SEC), pp. 445–459. Springer (2008)
30. Wang, D., Wang, P.: Understanding security failures of two-factor authentication schemes for real-time applications in hierarchical wireless sensor networks. Ad Hoc Netw. **20**, 1–15 (2014)
31. Wang, D., Wang, P.: Two birds with one stone: Two-factor authentication with security beyond conventional bound. In: IEEE Transactions on Dependable and Secure Computing (2016)