Integrity Assurance in Resource-Bounded Systems through Stochastic Message Authentication

Aron Laszka Yevgeniy Vorobeychik Xenofon Koutsoukos

Institute for Software Integrated Systems
Department of Electrical Engineering and Computer Science
Vanderbilt University
Nashville, TN

ABSTRACT

Assuring communication integrity is a central problem in security. However, overhead costs associated with cryptographic primitives used towards 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. We propose a formal game-theoretic framework for optimal stochastic message authentication, providing provable integrity guarantees for resource-bounded systems based on an existing MAC scheme. We use our framework to investigate attacker deterrence, as well as optimal design of stochastic message authentication schemes when deterrence is impossible. Finally, we provide experimental results on the computational performance of our framework in practice.

Categories and Subject Descriptors

K.6.5 [**Management Of Computing and Information Systems**]: Security and Protection

General Terms

Security, Economics, Theory

Keywords

message authentication, game theory, economics of security

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, 3, 4, 7]. The severely limited computational power of these devices 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 [3].

Numerous approaches for "lightweight" cryptography have previously been proposed to address this problem [4, 12, 14] (see related work in Section 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.

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. Specifically, our contribution is a stochastic message authentication framework, which authenticates messages randomly in a way

HotSoS'15, April 21 - 22, 2015, Urbana, IL, USA

ACM 978-1-4503-3376-4/15/04...$15.00
http://dx.doi.org/10.1145/2746194.2746195.

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.

Our main contributions are:

- Based on our threat model and objectives (Section 2), we formulate stochastic message authentication as a Stackelberg attacker-defender game (Section 3).
- We study the adversary's best responses (Section 4.1), characterize when the adversary can be deterred from attacking (Section 4.2), and discuss finding an optimal defense when deterrence is impossible (Section 4.3).
- We implement our stochastic authentication scheme on an actual hardware, and demonstrate its viability using experiments (Section 5).

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 Section 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. We model the problem as a two-player, non-deterministic, non-zero-sum, sequential 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.

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. We denote the amount of loss that a modified message of class $c \in \{1, \ldots, C\}$ can potentially cause by $L_c > 0$. Furthermore, 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.

Table 1: List of Symbols

Symbol	Description
C	number of message classes
L_c	amount of loss 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

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 $\boldsymbol{a} = (a_1, \ldots, a_C) \in \mathbb{N}^C$, knowing what strategy the defender has selected.
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. We also follow Kerckhoff's principle by assuming that the attacker knows the defender's algorithms, implementation, etc. and can compute the defender's strategy.

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

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

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

In the analysis, our goal will be to find the adversary's best response and the defender's optimal strategies, which are defined as follows.

Definition 2. An adversarial strategy is a *best response* if it maximizes the adversary's payoff, taking the defense strategy as given.

As is typical in the security literature, we consider a refinement of subgame perfect equilibria, called *strong Stackelberg* equilibria [8]. 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

$$\max_{\boldsymbol{a}^* \in \operatorname{argmax}_{\boldsymbol{a}} \mathcal{U}_A(\boldsymbol{p}, \boldsymbol{a})} \mathcal{U}_D(\boldsymbol{p}, \boldsymbol{a}^*) \ . \tag{6}$$

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

4. ANALYSIS

In this section, we present theoretical results on our message authentication game. First, we discuss the adversary's best-response strategies in Section 4.1. Then, we study the defender's optimal strategies in Sections 4.2 and 4.3. In Section 4.2, we characterize those instances of the message authentication game where the defender's optimal payoff is zero, while in Section 4.3, we study the instances where the optimal payoff is non-zero.

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 the detection probability, the 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 4. 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 an optimal defense. The following lemma provides a necessary condition on 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 will be available in an extended, online version of the paper.

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 non-zero. 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 will be available in an extended, online version of the paper.

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 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 proven the existence of such a best response. □

Unfortunately, this results 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 Figure 2a later.

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*

$$a^* = -\frac{1}{\ln(1-p_1)} - \frac{F}{L_1} \ . \tag{7}$$

The proof of Lemma 3 can be found in Appendix A.1.

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 Equation (7). 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 A.2.

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.

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

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

$$\forall c \in \{1, \ldots, C\} : \; a_c \leq \max\left\{0, \left\lceil -\frac{1}{\ln(1-p_c)} - \frac{F}{L_c} \right\rceil\right\} . \tag{8}$$

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

$$\max\left\{0, \left\lceil -\frac{1}{\ln(1-p_c)} - \frac{F}{L_c} \right\rceil\right\} \tag{9}$$

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 following paragraph and Figure 1). Furthermore, note that this computation should be performed at design time, not by the computationally-limited device during runtime.

Numerical Illustrations.

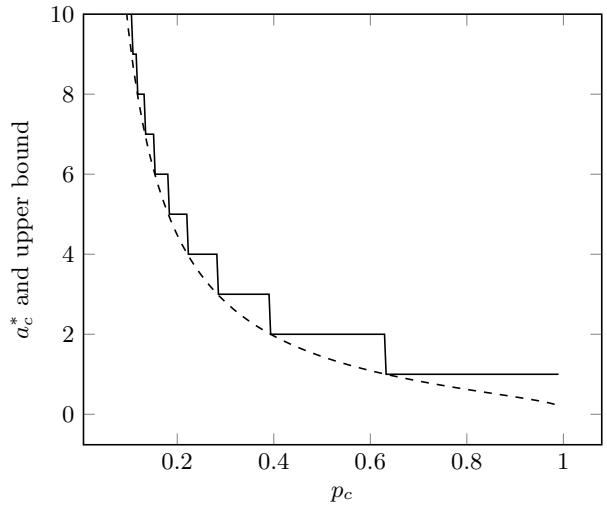
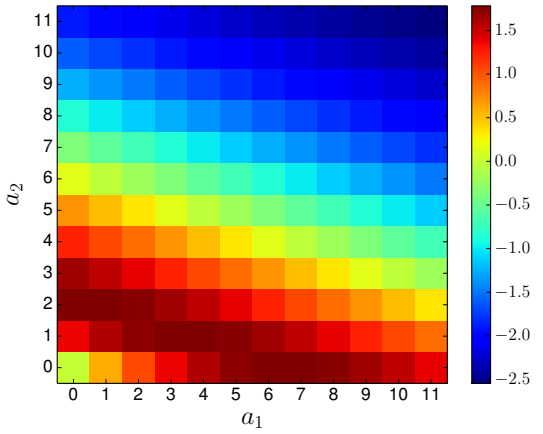
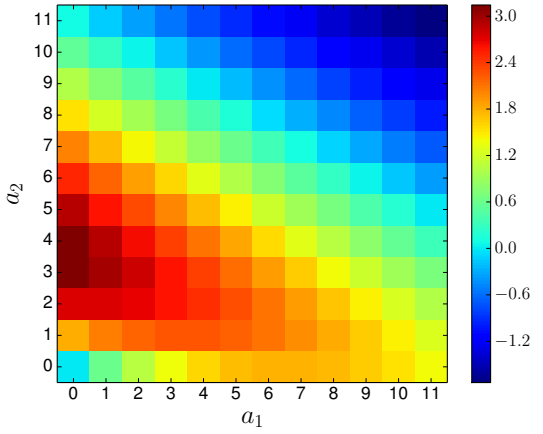


Figure 1: The adversary's single-class best response in the continuous model (dashed line) and the upper bound on her strategy in the original model (solid line) as functions of the verification probability p_c for $F = 0$.

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 $F = 0$. 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 Section 4.1.2 that we can easily compute the adversary's best response for a single message class).



(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 $a_1 = 0$, $a_2 = 3$.

Figure 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 $L_2 = 3$.

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 Figure 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 Figure 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 Deterrence Strategies

Now, we study the 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 *non-zero*.

Definition 5. A defense strategy $\boldsymbol{p}$ is a *deterrence strategy* if not attacking at all (i.e., $\boldsymbol{a} = \mathbf{0}$) is a best response.

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

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

$$\forall c \in \{1, \ldots, C\} : \; p_c \geq \frac{L_c}{L_c + F} \; . \tag{10}$$

PROOF. First, we prove the necessity of Equation (10). For the sake of contradiction, suppose that Equation (10) 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

$$(1 - p_c)L_c - p_c F \geq \frac{F}{L_c + F} L_c - \frac{L_c}{L_c + F} F = 0 \; . \tag{11}$$

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, Equation (10) necessarily holds if not attacking is a best response.

Second, we prove the sufficiency of Equation (10). We show this 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

$$\begin{aligned} F(L_1 + F)^{a_1} =& F\left(F^{a_1} + a_1 F^{a_1 - 1} L_1 + \ldots\right) && (12)\\ \geq& F\left(F^{a_1} + a_1 F^{a_1 - 1} L_1\right) && (13)\\ =& F^{a_1}\left(a_1 L_1 + F\right) \; , && (14) \end{aligned}$$

which implies that

$$\frac{F}{a_1 L_1 + F} \leq \left(\frac{F}{L_1 + F}\right)^{a_1} \; . \tag{15}$$

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

$$\begin{aligned} & (1 - p_1)^{a_1}(a_1 L_1 + F) - F && (16)\\ =& (a_1 L_1 + F)\left((1 - p_1)^{a_1} - \frac{F}{a_1 L_1 + F}\right) && (17)\\ \leq& \underbrace{(a_1 L_1 + F)}_{\geq 0}\left(\underbrace{\left(\frac{F}{L_1 + F}\right)^{a_1} - \frac{F}{a_1 L_1 + F}}_{\leq 0}\right) && (18)\\ \leq& 0 \; . && (19) \end{aligned}$$

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. For the remainder of the proof, let $\hat{L} = \sum_{c=1}^{C-1} a_c L_c$ and $\hat{P} = \prod_{c=1}^{C-1}(1 - p_c)^{a_c}$. 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

$$\begin{aligned} \mathcal{U}_A(\boldsymbol{p}, \boldsymbol{a}) =& \prod_{c=1}^{C}(1 - p_c)^{a_c} \sum_{c=1}^{C} a_c L_c - \left(1 - \prod_{c=1}^{C}(1 - p_c)^{a_c}\right) F \\ =& \; \hat{P}(1 - p_C)^{a_C}(\hat{L} + a_C L_C) \\ & - \left(1 - \hat{P}(1 - p_C)^{a_C}\right) F && (22)\\ =& \; (1 - p_C)^{a_C}\hat{P}\hat{L} + \hat{P}(1 - p_C)^{a_C} a_C L_C \\ & - \left(1 - \hat{P}(1 - p_C)^{a_C}\right) F \; . && (23) \end{aligned}$$

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

$$\begin{aligned} \mathcal{U}_A(\boldsymbol{p}, \boldsymbol{a}) \leq & \; (1 - p_C)^{a_C}(1 - \hat{P})F + \hat{P}(1 - (1 - p_C)^{a_C})F \\ & - \left(1 - \hat{P}(1 - p_C)^{a_C}\right) F && (24)\\ =& F\Big(\; (1 - p_C)^{a_C}(1 - \hat{P}) + \hat{P}(1 - (1 - p_C)^{a_C}) \\ & - 1 + \hat{P}(1 - p_C)^{a_C}\Big) && (25)\\ =& F\left((1 - p_C)^{a_C} + \hat{P} - 1 - \hat{P}(1 - p_C)^{a_C}\right) && (26)\\ \leq& \underbrace{F}_{\geq 0}\left(\underbrace{(\hat{P} - 1)}_{\leq 0}\underbrace{(1 - (1 - p_C)^{a_C})}_{\geq 0}\right) && (27)\\ \leq& 0 \; . && (28) \end{aligned}$$

Hence, no strategy can achieve higher payoff than not attacking (i.e., higher than zero payoff). Therefore, Equation (10) 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.

4.3 Optimal Defense without Deterrence

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

4.3.1 Continuous Relaxation

First, we study the continuous relaxation of the problem (see Definition 4), 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*

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

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 A.3.

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.

4.3.2 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 Figure 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 A.4.

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. □

One of the most important consequences of Lemma 5 is that an optimal defense strategy always uses all of the avail-

able 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 $\sum_i p_i T_i = B$.*

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. □

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 Figure 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 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 Figure 4 and its discussion).

Numerical Illustrations. Note that the number of included figures is limited by the available space, but our observations are consistent throughout the parameter space.

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 $(p_1 = 0.25, p_2 = 0.5)$.

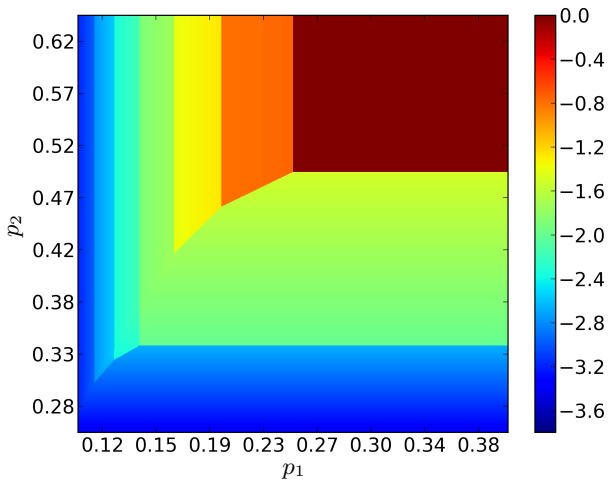


Figure 3: Defender's payoff for various strategies given that the adversary plays her best response. The parameters are $F = 3$, $L_1 = 1$, and $L_2 = 3$.

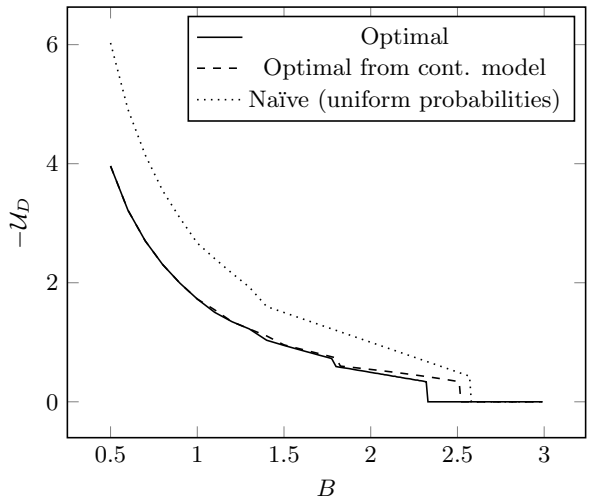


Figure 4: Defender's expected loss for her optimal strategy (solid line) compared to her expected loss for the optimal strategy computed in the continuous model (dashed line) and her expected loss for a naïve strategy using uniform probabilities (dotted line). The parameters are $F = 0.5$, $L_1 = 1$,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

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 (····) 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

[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.

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).

5. IMPLEMENTATION

In this section, we discuss how our theoretical results can be implemented and used in practice, and provide experimental results on the running time of our scheme, which demonstrate its practical viability.

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 practice, these parameters can 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 [2, 9].
- 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.

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

5.2 Implementing 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 nontrivial 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 LFSRs).

5.3 Experimental Results

For the practical evaluation demonstrating the feasibility of our approach, we implemented our stochastic message authentication scheme using SHA-1 HMAC and a linear feedback shift register PRNG on an ATmega328P[4] microcontroller. Using this implementation, we performed experiments measuring the running time of our scheme for various authentication probabilities.

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

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

has to be performed at design time.

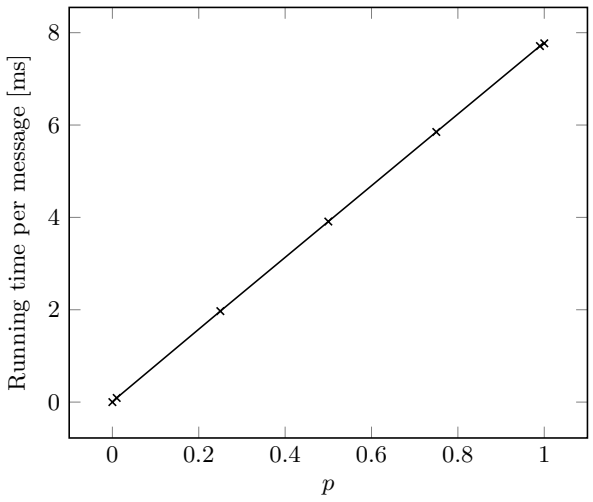


Figure 5: Average running time per message as a function of probability for stochastic MAC verification. Each x marks a measured value. Note that no PRNG was used for $p = 1$.

Figure 5 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.

6. RELATED WORK

To the best of our knowledge, there has been little research on message authentication using game theory. In [16] and [17], the author formulates a game-theoretic model of the contest between the sender, the receiver, and the adversary, to study message authentication on a noisy channel; however, the author does not consider resource bounds. Game theory has been used more generally in security, in attacker-defender games [8, 13]; for example, it can be used to study the optimal interdiction of attack plans [11].

Several research efforts have tried to provide lightweight cryptographic primitives and mechanisms for resource-bounded systems [3, 4, 12, 14, 15]. Note that our approach is complementary to these results, since we build on an 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, in [7], the authors describe a new family of lightweight block ciphers named KLEIN, which are designed to be usable as building blocks for security in resource-constrained devices. Another example is Hummingbird-2 [5], an encryption algorithm targeted for low-end microcontrollers. Besides lightweight primitives, researchers have also proposed mechanisms for resource-constrained systems. For example, the authors of [6] propose a lightweight message authentication scheme for smart grid communications. Finally, in [10], the authors combine lightweight cryptographic primitives for securing ad-hoc networks.

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 studied the adversary's best-response and the defender's optimal 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. 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. Finally, we presented experimental results on the performance of our scheme, 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.

Acknowledgments

This work was supported in part by the National Science Foundation under Award CNS-1238959 and by the Air Force Research Laboratory under Award FA8750-14-2-0180.

8. REFERENCES

[1] J. Akerberg, M. Gidlund, and M. Bjorkman. Future research challenges in wireless sensor and actuator networks targeting industrial automation. In *Proceedings of the 9th IEEE International Conference on Industrial Informatics (INDIN)*, pages 410–415, 2011.

[2] K. Campbell, L. A. Gordon, M. P. Loeb, and L. Zhou. The economic cost of publicly announced information security breaches: empirical evidence from the stock market. *Journal of Computer Security*, 11(3):431–448, 2003.

[3] A. A. Cárdenas, S. Amin, and S. Sastry. Research challenges for the security of control systems. In *Proceedings of the 3rd USENIX Workshop on Hot Topics in Security (HotSec)*, July 2008.

[4] T. Eisenbarth, S. Kumar, C. Paar, A. Poschmann, and L. Uhsadel. A survey of lightweight-cryptography implementations. *IEEE Design & Test of Computers*, 24(6):522–533, 2007.

[5] D. Engels, M.-J. O. Saarinen, P. Schweitzer, and E. M. Smith. The Hummingbird-2 lightweight authenticated encryption algorithm. In *Proceedings 7th International Workshop (RFIDSec), Revised Selected Papers*, pages 19–31, 2011.

[6] M. M. Fouda, Z. M. Fadlullah, N. Kato, R. Lu, and X. Shen. A lightweight message authentication scheme for smart grid communications. *IEEE Transactions on Smart Grid*, 2(4):675–685, 2011.

[7] Z. Gong, S. Nikova, and Y. W. Law. KLEIN: A new family of lightweight block ciphers. In *Proceedings of the 7th Workshop on RFID Security and Privacy (RFIDSec), Revised Selected Papers*, pages 1–18, 2011.

[8] D. Korzhyk, Z. Yin, C. Kiekintveld, V. Conitzer, and M. Tambe. Stackelberg vs. Nash in security games: An extended investigation of interchangeability, equivalence, and uniqueness. *Journal of Artificial Intelligence Research*, 41(2):297–327, 2011.

[9] R. L. Krutz and R. D. Vines. *The CISSP Prep Guide: Mastering the ten domains of Computer Security*. Wiley New York, 2001.

[10] A. Kumar and A. Aggarwal. Lightweight cryptographic primitives for mobile ad hoc networks. In *Proceedings of the 2012 International Conference on Security in Computer Networks and Distributed Systems (SNDS)*, pages 240–251, 2012.

[11] J. Letchford and Y. Vorobeychik. Optimal interdiction of attack plans. In *Proceedings of the 12th International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, pages 199–206, 2013.

[12] D. Maimut and K. Ouafi. Lightweight cryptography for RFID tags. *IEEE Security Privacy*, 10(2):76–79, March 2012.

[13] M. H. Manshaei, Q. Zhu, T. Alpcan, T. Bacşar, and J.-P. Hubaux. Game theory meets network security and privacy. *ACM Computing Surveys (CSUR)*, 45(3):25, 2013.

[14] A. Moradi and A. Poschmann. Lightweight cryptography and DPA countermeasures: A survey. In *Proceedings of the 1st International Workshop on Lightweight Cryptography for Resource-Constrained Devices (WLC)*, pages 68–79, 2010.

[15] D. C. Ranasinghe. Lightweight cryptography for low cost RFID. In *Networked RFID Systems and Lightweight Cryptography*, pages 311–346. Springer Berlin Heidelberg, 2008.

[16] G. J. Simmons. Game theory model of digital message authentication. Technical report, Sandia National Labs., Albuquerque, NM (USA), 1981.

[17] G. J. Simmons. Authentication theory/coding theory. In *Advances in Cryptology*, volume 196, pages 411–431. Springer, 1985.

APPENDIX

A. PROOFS

A.1 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 best response maximizes $\mathcal{U}_A(a) = (1-p)^a(F + aL) - F$.

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

$$\begin{aligned}\frac{d}{da}\mathcal{U}_A(a) &= \ln(1-p)(1-p)^a(F+aL) + (1-p)^a L - 0 \\ &= (1-p)^a(\ln(1-p)(F+aL) + L)\ . \end{aligned} \tag{32}$$

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

$$0 = \underbrace{(1-p)^a}_{>0}(\ln(1-p)(F+aL) + L) \tag{33}$$

$$0 = \ln(1-p)(F+aL) + L \tag{34}$$

$$a\ln(1-p)L = -\ln(1-p)F - L \tag{35}$$

$$a = -\frac{1}{\ln(1-p)} - \frac{F}{L}\ . \tag{36}$$

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). Hence, 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 is a unique best-response strategy. □

A.2 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 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

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

If $\hat{L}$ were zero, then $\hat{\boldsymbol{a}}$ would also be a single-class best response, since its only non-zero element would be $\hat{a}_c$. However, this would contradict our initial assumption that a_c^* is the maximal single-class best response; therefore, we have $\hat{L} > 0$. Then, it follows readily from $\hat{a}_c > a_c^*$ that

$$\frac{\hat{a}_c L_c + F}{a_c^* L_c + F} > \frac{\hat{a}_c L_c + \hat{L} + F}{a_c^* L_c + \hat{L} + F} \tag{37}$$

$$\frac{\hat{a}_c L_c + F}{a_c^* L_c + F}(a_c^* L_c + \hat{L} + F) > \hat{a}_c L_c + \hat{L} + F\ . \tag{38}$$

Since a_c^* is a single-class best response, we have

$$(1-p_c)^{a_c^*}(a_c^* L_c + F) \geq (1-p_c)^{\hat{a}_c}(\hat{a}_c L_c + F) \tag{39}$$

$$(1-p_c)^{a_c^*} \geq (1-p_c)^{\hat{a}_c}\frac{\hat{a}_c L_c + F}{a_c^* L_c + F}\ . \tag{40}$$

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

$$(1-p_c)^{a_c^*}\hat{P}(a_c^* L_c + \hat{L} + F) \tag{41}$$

$$\geq (1-p_c)^{\hat{a}_c}\frac{\hat{a}_c L_c + F}{a_c^* L_c + F}\hat{P}(a_c^* L_c + \hat{L} + F) \tag{42}$$

$$> (1-p_c)^{\hat{a}_c}\hat{P}(\hat{a}_c L_c + \hat{L} + F) \tag{43}$$

$$= \mathcal{U}_A(\boldsymbol{p}, \hat{\boldsymbol{a}})\ . \tag{44}$$

Note that, in the first step, we used Equation (40) and, in the second step, we used Equation (38).

This inequality shows 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. □

A.3 Proof of Theorem 4

Proof. (Sketch.) First, we show that the ratios have to be uniform. 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, we can shown 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 her costs. Hence, the defender can strictly decrease her loss, which contradicts the assumption that the original strategy is optimal; therefore, the claim must hold.

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 $pT = B$ for a single class. Since the adversary will modify $a^* = -\frac{1}{\ln(1-p)} - \frac{F}{L}$ messages, the defender's loss is

$$(1-p)^{-\frac{1}{\ln(1-p)}-\frac{F}{L}}\left(-\frac{1}{\ln(1-p)}-\frac{F}{L}\right)L \tag{45}$$

$$=\frac{L}{e}(1-p)^{-\frac{F}{L}}\left(-\frac{1}{\ln(1-p)}-\frac{F}{L}\right)\ . \tag{46}$$

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

$$\frac{d}{dp}\frac{L}{e}(1-p)^{-\frac{F}{L}}\left(-\frac{1}{\ln(1-p)}-\frac{F}{L}\right) \tag{47}$$

$$=\frac{L}{e}\left(-\frac{F}{L}(1-p)^{-\frac{F}{L}-1}\left(-\frac{1}{\ln(1-p)}-\frac{F}{L}\right)\right.$$
$$\left.-(1-p)^{-\frac{F}{L}}\frac{1}{(1-p)\ln^2(1-p)}\right) \tag{48}$$

$$=\underbrace{\frac{L}{e}(1-p)^{-\frac{F}{L}-1}}_{>0}\left(-\frac{F}{L}\underbrace{\left(-\frac{1}{\ln(1-p)}-\frac{F}{L}\right)}_{=a^*\geq 0}-\underbrace{\frac{1}{\ln^2(1-p)}}_{>0}\right)$$
$$<0\ . \tag{49}$$

Since the derivative is negative, the minimal loss is achieved at the maximal 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 $\boldsymbol{p}$ satisfying $\frac{L_c}{\ln(1-p_c)} = R$ 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 T, there exists a unique defense strategy $\boldsymbol{p}$ satisfying both $\frac{L_1}{\ln(1-p_1)} = \ldots = \frac{L_C}{\ln(1-p_C)}$ and $\sum_c p_c T_c = B$. □

A.4 Proof of Lemma 5

Proof. Let $\boldsymbol{a}^*$ and $\boldsymbol{a}'$ be best responses for $\boldsymbol{p}^*$ and $\boldsymbol{p}'$, respectively. 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, 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

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}^*) \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}') \tag{50}$$

and

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}') \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}^*)\ . \tag{51}$$

Second, observe that $\boldsymbol{p}' > \boldsymbol{p}^*$ follows from the condition of the lemma. If $\boldsymbol{a}' = \boldsymbol{0}$ were true, then the claim of the lemma would obviously hold, since $\boldsymbol{p}^*$ does not deter the adversary while $\boldsymbol{p}'$ does. Hence, we can assume $\boldsymbol{a}' \neq \boldsymbol{0}$ for the remainder of the proof. Then, using the definition of the adversary's payoff, we have

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}') > \tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}')\ . \tag{52}$$

By combining these inequalities, we get

$$\tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}^*) \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}') > \tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}') \geq \tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}^*)\ , \tag{53}$$

which implies that

$$\frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}^*)}{\tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}^*)} \geq \frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}')}{\tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}')}\ . \tag{54}$$

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

$$\frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}^*)}{\tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}^*)} = \prod_i \left(\frac{1-p_i^*}{1-p_i'}\right)^{a_i^*} \tag{55}$$

and

$$\frac{\tilde{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}')}{\tilde{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}')} = \prod_i \left(\frac{1-p_i^*}{1-p_i'}\right)^{a_i'}\ . \tag{56}$$

By substituting these fractions into the previous inequality, we get

$$\prod_i \left(\frac{1-p_i^*}{1-p_i'}\right)^{a_i^*} \geq \prod_i \left(\frac{1-p_i^*}{1-p_i'}\right)^{a_i'} \tag{57}$$

$$\sum_i a_i^* \ln\frac{1-p_c^*}{1-p_c'} \geq \sum_i a_i' \ln\frac{1-p_c^*}{1-p_c'} \tag{58}$$

$$\sum_i a_i^* L_i \geq \sum_i a_i' L_i \tag{59}$$

$$L^* \geq L'\ . \tag{60}$$

Now, for the sake of contradiction, suppose that the claim of the lemma does not hold; that is, suppose $P'L' \geq P^*L^*$. By combining this with Equation (60), we get

$$P' \geq P^* \tag{61}$$

$$P'F \geq P^*F \tag{62}$$

$$P'L' + P'F \geq P^*L^* + P^*F \tag{63}$$

$$P'(L'+F) \geq P^*(L^*+F) \tag{64}$$

$$\hat{\mathcal{U}}_A(\boldsymbol{p}',\boldsymbol{a}') \geq \hat{\mathcal{U}}_A(\boldsymbol{p}^*,\boldsymbol{a}^*)\ . \tag{65}$$

However, this contradicts Equation (53). Therefore, the claim of the lemma has to hold. □