Rational Distance-Bounding Protocols over Noisy Channel

Long Hoang Nguyen
Oxford University Department of Computer Science
Parks Road, Wolfson Building, OX1 3QD, UK
Long.Nguyen@cs.ox.ac.uk

ABSTRACT

We use ideas from game theory to define a new notion for an optimal threshold for the number of erroneous responses that occur during the rapid-bit exchange over noisy channels in a distance-bounding protocol. The optimal threshold will ensure that even if an intruder attacks the protocol, the expected loss the verifier suffers will still be lower than when the intruder does not attack. Any rational intruder, who always tries to maximise the verifier's loss, will not therefore have any incentive to attack the protocol. We then demonstrate how statistical analysis and binary search are used to locate the unique and optimal threshold accurately.

Categories and Subject Descriptors: H.m [Information System]: Miscellaneous

General Terms: Security.

1. INTRODUCTION

Ideas from game theory have been used to re-design a number of fair exchange protocols [2, 15] and secret sharing schemes [4, 5] so that parties cannot act on their own interests to bring these schemes to failure [5]. As an example, in a fair exchange, a party accepts to deliver an item iff it receives another item in return, and hence even unmalicious but self-interested parties will be tempted to deviate from a protocol to gain advantage. This notion of players' rationality or self-interest is however not applicable to distance-bounding protocols because in those protocols it is in the mutual interest of both the legitimate verifier and prover that they should cooperate to complete a protocol successfully to authenticate each other.

We instead observe that in many scenarios the intruder is *rational* in the sense that it always tries to maximise the expected loss or cost it can cause to the verifier (or tag reader) in a distance-bounding protocol. In the scheme, the loss a verifier suffers mainly comes from either false rejection (rejecting a legitimate tag or user) or false acceptance (authenticating a malicious tag or attacker). Both of these arise from noise existing in the data layer which exchanges bits during the rapid-bit exchange or the challenge-response phase lasting multiple rounds of the protocols. The noise therefore necessitates the use of a tolerance threshold, such that a prover is authenticated if the total number of its erroneous responses is below the threshold. These features motivate us to use techniques in game theory to define and locate an optimal threshold to resist this kind of rational intruder: even if an attacker impersonates a user (e.g. in an attempt to forge a false acceptance), the expected loss the verifier suffers will still be lower than when the verifier communicates with the legitimate user. The attacker therefore does not have any incentive to disrupt a protocol.

The use of a rapid-bit exchange is to compute the upper bound on the distance of the prover as in a RFID distance-bounding protocol. Since only bits are rapidly exchanged over a noisy channel, the information can be corrupted.[1] To counter noise, Hancke and Kuhn [6] introduce the use of a threshold value on the number of erroneous responses. This work is subsequently extended by Kim et al. [7]. The contributions however do not attempt to define and locate an optimal threshold value given that the number of rounds and the noise level are fixed. They simply point out that the likelihood of authenticating a legitimate user will become much higher than an attacker as the number of rounds increases without taking into consideration the cost of running these protocols and the losses arising from false acceptance and false rejection. This therefore motivates Dimitrakakis et al. [3] to introduce a general framework enabling an expected loss analysis for all protocols of this type. In this framework, different costs are assigned to a variety of possible events or protocol outcomes taking into account transmission energy, computation, and time overhead. They then show how to find a nearly optimal value for the threshold so that the worst-case expected loss is minimised, we will however demonstrate that this is not the same as discouraging an intruder from attacking a protocol.

Our first contribution presented in Section 3 is to show that there exists a unique value for the threshold on the number of erroneous responses in a distance-bounding protocol over noisy channel such that not only does a rational intruder not have any incentive to attack but also the verifier's loss due to false rejection is minmimised as much as possible. We then demonstrate how such a threshold can be exactly calculated by using statistical analysis in combina-

[1]Existing distance-bounding protocols usually separate noise analysis from cryptographic analysis, i.e. the rapid-bit exchange is assumed to take place in noiseless channel and so a separate error-correction protocol will be added later.

SIN'11, November 14–19, 2011, Sydney, Australia.

tion with binary search. This new approach will be used in this paper to resist two different and well-studied attacking strategies on any protocols of this type, namely distance-fraud attackers in Section 4 and mafia-fraud attackers in Section 5. In Section 7, we point out that our approach can be easily adapted to locate exactly a different kind of optimal threshold due to Dimitrakakis et al. [3] that minimises the worst-case expected loss. This improves on the finite-sample loss bounds for a nearly optimal threshold of Dimitrakakis et al. that provides a good approximation but for a limited range of the probability of erroneous transmission.

To assess the performance of this new approach in resisting rational attackers, we have attempted to quantify the costs a verifier suffers due to false rejection and false acceptance and then carried out experiments and report our results in Section 6. The experimental results clearly demonstrate that the optimal threshold value appears to grow linearly with respect to the round number while sublinearly with respect to the level of noise.

2. NOTATIONS

Capital letters V, P, U and A denote the legitimate verifier, the prover, the legitimate user and the attacker in a distance-bounding protocol, where V always seeks to authenticate the legitimate user U who shares a private key with V. The role of the prover P can be played by either U or A. $\mathbb{P}(X)$ denotes the probability of event X, while $\mathbb{E}(L|X)$ denotes the conditional expectation of the verifier's loss when X is true.

We consider distance-bounding protocols whose rapid-bit exchange lasts n rounds under noisy conditions.[2] This implies that whenever a symbol (or a bit $x \in \{0,1\}$) is sent during the rapid-bit exchange between the verifier V and the prover P, it can be altered due to noise in the physical medium. Let us denote $w = \mathbb{P}(x \neq x')$ the probability of erroneous transmission in the data layer, where x' is the received symbol at the other end of the communication. Without loss of generality $w \in [0, 0.5]$, because when $w > 0.5$ and w is known prior to a rapid-bit exchange,[3] parties can adopt the following strategy to turn this into the case where $w' = 1 - w \leq 0.5$: if a node intends to commmunicate a bit b, then it will transmit $\overline{b}$ and the probability the receiver gets $\overline{b}$ is $1 - w$.

During the n-round rapid-bit exchange, V sends one-bit challenges $c_1, \cdots, c_n$ to P, who then responds by transmitting one-bit responses $r_1, \cdots, r_n$. The verifier can calculate the correct response $r_i = R_i(c_i)$ for any $i \in \{1, \cdots, n\}$, and hence can check the accuracy of P's response in each round.[4] These responses coming from P can however be corrupted due to noise, and so this necessitates the use of a tolerance threshold $\tau \in \{1, \cdots, n\}$, such that P is authenticated if the total number of erroneous responses is below τ.

[2]A distance-bounding protocol usually consists of three phases: initialisation, rapid-bit exchange and termination. Since rapid-bit exchange is most sensitive to noise, and not the other two, we focus on the rapid-bit exchange here.

[3]The noise level can be estimated through communication during the intialisation and termination phases of a distance-bounding protocol [3].

[4] The function $R_i()$ depends on a private key shared between V and U though $R_i()$ appears to be random and uniformly distributed, i.e. for any $i \in \{1, \cdots, n\}$ and given that $c_i \neq c_i'$ we have $\mathbb{P}(R_i(c_i) \neq R_i(c_i')) = \mathbb{P}(R_i(c_i) = R_i(c_i')) = 1/2$.

Also due to noise, there are two possibilities that can go wrong in any protocol of this type:[5]

- False rejection: this happens when V rejects a legitimate user ($P = U$) because the number of erroneous responses coming from the user ϵ_U is greater than or equal to τ. The verifier V therefore suffers a cost l_U. We denote $\mathbb{P}(\epsilon \geq \tau | P = U)$ or $\mathbb{P}(\epsilon_U \geq \tau)$ the probability of false rejection.
- False acceptance: this happens when V authenticates a malicious attacker ($P = A$) because the number of erroneous responses coming from the attacker ϵ_A is smaller than τ. The verifier V therefore suffers a cost l_A and $\mathbb{P}(\epsilon < \tau | P = A)$ or $\mathbb{P}(\epsilon_A < \tau)$ denotes the probability of false acceptance.

In addition to costs l_A and l_U associated with false acceptance and false rejection, a cost l_B is assigned to each round of the rapid-bit exchange as suggested by Dimitrakakis et al. [3]. We note that even though the costs can vary widely in different scenarios in practice, if we can restrict the range of the intruder's interests, such as incorrect authentication and denial of service, then it is possible to accurately estimate and compute the costs as demonstrated in Section 6.

Since there are n rounds in a rapid-bit exchange, as in [3], the losses V suffers in different protocol outcomes are:

$$L = \begin{cases} nl_B + l_U & \text{if } \epsilon \geq \tau \text{ and } P = U, \\ nl_B + l_A & \text{if } \epsilon < \tau \text{ and } P = A, \text{ and} \\ nl_B & \text{for otherwise.} \end{cases}$$

From this definition, the expected loss the verifier suffers when the prover is either U or A can be derived as follows

$$\mathbb{E}(L|U,\tau) = nl_B + \mathbb{P}(\epsilon_U < \tau) \cdot 0 + \mathbb{P}(\epsilon_U \geq \tau) \cdot l_U \quad (1)$$

$$\mathbb{E}(L|A,\tau) = nl_B + \mathbb{P}(\epsilon_A < \tau) \cdot l_A + \mathbb{P}(\epsilon_A \geq \tau) \cdot 0 \quad (2)$$

In the sections to come, we will demonstrate how noisy communication can be, perhaps interestingly, used to discourage a rational intruder from attacking protocols of this type.

2.1 A distance-bounding protocol

Although our work applies to a wide range of distance-bounding protocols under noisy condition, for clarity we give the specification of the distance-bounding protocol of Reid et al. [14] below. In this protocol, the verifier V (or the tag reader) and the legitimate prover $P = U$ (or the tag) share a common secret $x \in \{0,1\}^n$. The messages exchanged are:

1. V chooses a random number $y_V \in \{0,1\}^m$ and sends it to P.
2. P also chooses a random number $y_P \in \{0,1\}^m$ and sends it to V.
3. Both entities now use a key derivation function $f : \{0,1\}^n \times \{0,1\}^* \to \{0,1\}^n$ to derive a key $k = f(x, V \parallel P \parallel y_V \parallel y_P)$. This is used to split the common secret key into two shares, k and $d = k \oplus x$.
4. V and P start a rapid-bit exchange which is subject to noise. The following steps are repeated for n rounds. At each round $i \in \{1, \cdots, n\}$:

[5]We assume that P is within a short distance from V, and thus there should not be any delay in transmission during the rapid-bit exchange.

(a) V chooses a random bit c_i, transmits it to P, and starts a clock.

(b) Upon receiving c_i, P replies $r_i = R_i(c_i)$, with $R_i(c_i) = d_i$ if $c_i = 0$, and $R_i(c_i) = k_i$ if $c_i = 1$.

(c) After the reception of r_i, V stops the clock and stores the delay time Δt_i and checks r_i.

It is easy to see that for any $i \in \{1, \cdots, n\}$ and given that $c_i \neq c_i'$ we have

$$\begin{aligned} \mathbb{P}(R_i(c_i) \neq R_i(c_i')) &= \mathbb{P}(d_i \neq k_i) = \mathbb{P}(k_i \oplus x_i \neq k_i) \\ &= \mathbb{P}(x_i = 1) = 1/2 \end{aligned}$$

The same property applies to other protocols of this type, such as protocols of Hancke and Kuhn [6], and Kim et al. [7].

3. LOSS ANALYSIS

Using ideas from game theory we observe that for a *rational* attacker, who always tries to maximise the verifier's loss, we want to choose the threshold so that even if the attacker impersonates the prover in an attempt to forge a false acceptance, the expected loss it causes to the verifier is still lower than when the verifier communicates with a legitimate user. This is captured by the following inequality

$$\begin{aligned} \mathbb{E}(L|P = U, \tau) &> \mathbb{E}(L|P = A, \tau) \qquad (3) \\ \mathbb{P}(\epsilon_U \geq \tau) \cdot l_U &> \mathbb{P}(\epsilon_A < \tau) \cdot l_A \\ \frac{\mathbb{P}(\epsilon_U \geq \tau)}{\mathbb{P}(\epsilon_A < \tau)} &> \frac{l_A}{l_U} \end{aligned}$$

Additionally we want to locate τ so that $\mathbb{E}(L|P = U)$ is as small as possible while Inequality (3) is still satisfied. From Equations (1) and (2), when we increase the threshold value τ, then $\mathbb{E}(L|P = U)$ decreases while $\mathbb{E}(L|P = A)$ increases, and hence there always exit a unique value for τ satisfying both Inequality (3) and the second condition captured by the following inequality.

$$\begin{aligned} \mathbb{E}(L|P = U, \tau + 1) &\leq \mathbb{E}(L|P = A, \tau + 1) \qquad (4) \\ \frac{\mathbb{P}(\epsilon_U \geq (\tau + 1))}{\mathbb{P}(\epsilon_A < (\tau + 1))} &\leq \frac{l_A}{l_U} \end{aligned}$$

Given that the number of rounds n and the probability of erroneous transmission w are fixed, we will show in the subsequent sections that it is possible to calculate the unique integer threshold τ exactly such that Inequalities (3) and (4) hold at the same time.

Our strategy can be simply explained as follows. We first use statistical analysis to calculate the probabilities of false rejection $\mathbb{P}(\epsilon_U \geq \tau)$ and false acceptance $\mathbb{P}(\epsilon_A < \tau)$ with respect to any common value of the threshold τ and any level of noise over the communication channel. Since τ is an integer chosen from $\{1, \cdots, n\}$, intuitively it is feasible to carry out a binary search to locate such an optimal value.

In order to understand the distribution of the number of erroneous responses coming from an attacker, we need to consider a variety of attacking strategies. To our knowledge, there are three main sources of attacking strategies in the literature that an intruder can pursue to cause incorrect authentication to distance-bounding protocols: (1) distance fraud; (2) mafia fraud; and (3) terrorist fraud. The last of these requires a legitimate user to be dishonest: although the user does not reveal the private key shared between itself and the legitimate verifier to the attacker, it cooperates with the terrorist attacker, i.e. giving the attacker the correct response to any challenge chosen by the attacker. Since it is impossible to resist a terrorist fraud attacker when there exist such a dishonest user, we will only investigate how to deal with the first two of these attackers in this paper.

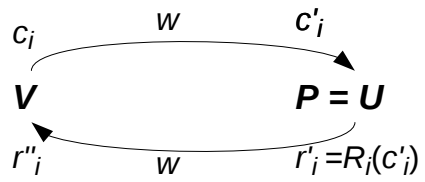

Figure 1: The verifier V is communicating with a legitimate user U with whom V shares a private key that underlies the function $R_i()$ for any $i \in \{1, \cdots, n\}$. w denotes channel noise.

As regards a legitimate user, there is only one type of legitimate user who shares a private key with the verifier and the user always follows the protocol to be authenticated to the verifier. In the next subsection we will first calculate the likelihood of false rejection as required in Inequalities (3) and (4) given the values for τ, w and n.

3.1 False rejection

In this scenario the prover is the legitimate user or $P = U$ and a protocol is run without the interference from any attacker as seen in Figure 1. Since we are interested in the probability of false rejection $\mathbb{P}(\epsilon_U \geq \tau)$ which depends on the number of erroneous responses ϵ_U that occur out of n rounds of a rapid-bit exchange, we first need to compute the probability of erroneous response in each round.

Theorem 1. *A legitimate user U wants to authenticate itself to the legitimate verifier V. If w denotes the probability of erroneous transmission then the probability of erroneous response from U in each round is:*

$$\mathbb{P}(Error|P = U) = w(3/2 - w)$$

Proof. We need to deal with noise existing in both the challenge and response phases of any round $i \in \{1, \cdots, n\}$:

- As seen in Figure 1, a challenge c_i is sent from V to U, but because of noise what U actually receives is c_i' such that $\mathbb{P}(c_i \neq c_i') = w$.
- The response $r_i' = R_i(c_i')$ is sent from U back to V, and also because of noise what V receives is r_i'' such that $\mathbb{P}(r_i'' \neq r_i') = w$.

From the above analysis, we can derive the probability of erroneous response from U in each round as follows:

$$\begin{aligned} \mathbb{P}(\text{Error}|P = U) &= \mathbb{P}(r_i'' \neq R_i(c_i)|P = U) \\ &= \mathbb{P}(c_i = c_i' \wedge r_i'' \neq r_i') + \\ &\quad \mathbb{P}(c_i \neq c_i' \wedge R_i(c_i) = R_i(c_i') \wedge r_i'' \neq r_i') + \\ &\quad \mathbb{P}(c_i \neq c_i' \wedge R_i(c_i) \neq R_i(c_i') \wedge r_i'' = r_i') \\ &= (1 - w)w + w^2/2 + w(1 - w)/2 \\ &= w(3/2 - w) \end{aligned}$$

The third equality follows because given that $c_i \neq c_i'$ we always have $\mathbb{P}(R_i(c_i) = R_i(c_i')) = \mathbb{P}(R_i(c_i) \neq R_i(c_i')) = 1/2$ as pointed out at the end of Section 2.1 and in Footnote 4. □

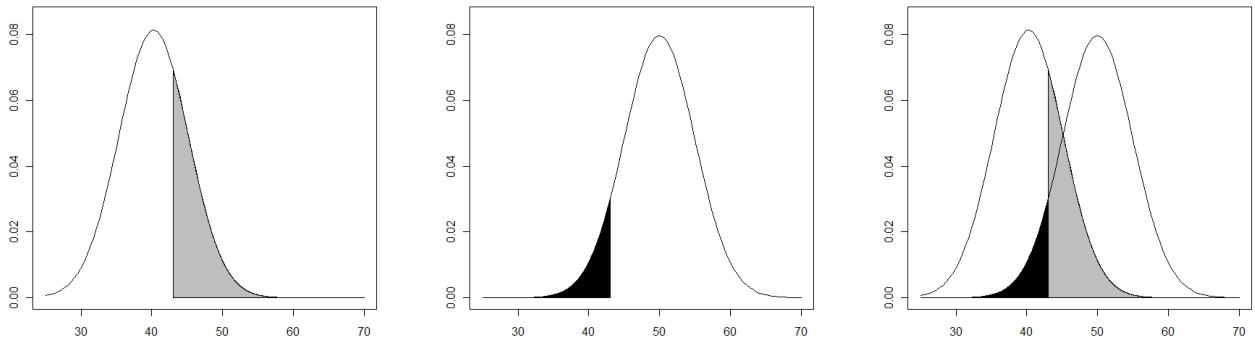

Figure 2: The first graph represents the binomial distribution $Binom(n, w(3/2 - w))$ of ϵ_U. The second graph represents the binomial distribution $Binom(n, 0.5)$ of ϵ_A. Third graph combines of the first two graphs.

The number of erroneous responses ϵ_U out of n rounds between V and U is represented by a binomial distribution

$$\epsilon_U \sim Binom(n, w(3/2 - w))$$

Given n, τ and w, the likelihood of false rejection $\mathbb{P}(\epsilon_U \geq \tau) = 1 - \mathbb{P}(\epsilon_U < \tau)$, which corresponds to the grey area of the first graph of Figure 2, can be easily calculated to a high level of accuracy by using tabulations of the probability mass functions of the normal or binomial distribution [1].

4. DISTANCE FRAUD ATTACKER

In this attack there are two active parties: a legitimate verifier V and a *distance fraud* attacker A_{DF} who acts as a legitimate U to convince V of being nearer to U than it really is. It is crucial to emphasise that A_{DF} does not interact with U who shares a private key with V.

Since we are interested in the probability of false acceptance $\mathbb{P}(\epsilon_A < \tau)$ which depends on the number of erroneous responses ϵ_A that occur out of n rounds of a rapid-bit exchange, we first need to compute the probability of erroneous response in each round.

Theorem 2. *An attacker A_{DF} only interacts with the legitimate verifier V but not the legitimate user U as seen in Figure 3. Regardless of the level of noise, if we assume that A_{DF} does not have any knowledge about the private key shared between V and U then the probability of erroneous response from A_{DF} in each round is:*

$$\mathbb{P}(Error|P = A_{DF}) = 1/2$$

Proof. Since the attacker does not have any knowledge about the private key shared between the verifier and the legitimate user, upon receiving any challenge c_i the best response the attacker can come up with is a random bit $r'_i \in_R \{0, 1\}$, and that means $\mathbb{P}(r'_i \neq R_i(c_i)) = 1/2$. Since the actual value of the challenge is of no importance to the attacker, there is no need to consider noise in the challenge phase of the exchange. We however need to take into account noise in the response phase, i.e. the response V receives is r''_i such that $\mathbb{P}(r''_i \neq r'_i) = w$.

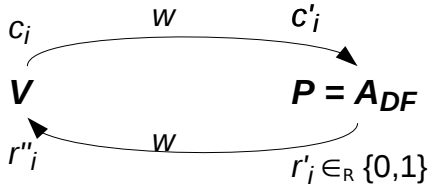

Figure 3: A distance fraud attacker A_{DF} posing as a prover communicates with a verifier V.

We therefore arrive at:

$$\begin{aligned}\mathbb{P}(\text{Error}|P = A_{DF}) &= \mathbb{P}(r''_i \neq R_i(c_i)|P = A_{DF}) \\ &= \mathbb{P}(r'_i \neq R_i(c_i) \wedge r''_i = r'_i) + \\ &\quad \mathbb{P}(r'_i = R_i(c_i) \wedge r''_i \neq r'_i) \\ &= (1 - w)/2 + w/2 = 1/2\end{aligned}$$

□

It is important to point out that when $w \in [0, 0.5]$

$$1/2 = \mathbb{P}(\text{Error}|P = A_{DF}) \geq \mathbb{P}(\text{Error}|P = U)$$

The number of erroneous responses ϵ_A out of n rounds between V and A_{DF} can also be represented by the following binomial distribution

$$\epsilon_A \sim Binom(n, 1/2)$$

Given n, τ and w, the likelihood of false acceptance $\mathbb{P}(\epsilon_A < \tau)$ that corresponds to the black area of the second graph of Figure 2 can also be easily calculated to a high level of accuracy by using tabulations of the probability mass functions of the normal distribution [1].

4.1 Optimal threshold value

Since the threshold value τ is the same regardless of whether $P = A_{DF}$ or $P = U$, we can combine the first and second graphs of Figure 2 into a single graph as seen in Figure 2.

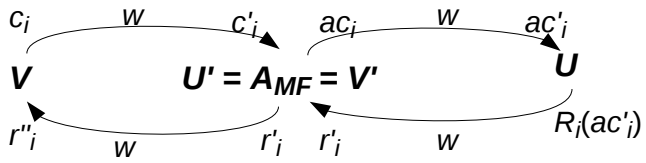

Figure 4: A mafia fraud attacker A_{MF} interacts with both a verifier V and a legitimate user U.

From the combined graph, it becomes very clear that iteratively we can find an integer value of $\tau \in \{1, \cdots, n\}$ such that Inequalities (3) and (4) of Section 3 hold.

Since we assumed that $w \leq 0.5$, and so $w(3/2 - w) \leq 0.5$. Starting from the interval $\tau \in [1, n/2]$ and by using binary search, we are guaranteed to locate an optimal integer value for τ after at most $(\log_2 n - 1)$ iterations such that the following two conditions are satisfied.

$$\frac{\mathbb{P}(\epsilon_U \geq \tau)}{\mathbb{P}(\epsilon_A < \tau)} > \frac{l_A}{l_U} \quad \text{and} \quad \frac{\mathbb{P}(\epsilon_U \geq (\tau+1))}{\mathbb{P}(\epsilon_A < (\tau+1))} \leq \frac{l_A}{l_U}$$

We will give our experimental results in Section 6 and our analysis on the behaviour of the optimal threshold as a function of either round number n or channel noise w.

5. MAFIA FRAUD ATTACKER

When an attacker can interact with both the legitimate user U and the verifier V who share a common private key, there is another source of attack called *mafia fraud*, and so $P = A_{MF}$. In a mafia fraud attack, the attacker plays the role of the man-in-the-middle: A_{MF} poses as U to interact with V and poses as V to interact with U. The attacker tries to convince V that it is communicating with the legitimate user U while in reality V communicates with the attacker.

Under noiseless conditions, the attacker could transmit an anticipated challenge ac_i to the legitimate user U before the verifier V sends its challenge c_i. Half of the time, $ac_i = c_i$, so the attacker can correctly reply $R_i(c_i)$ to the verifier. Otherwise, the attacker can make a random guess, being correct half of the time. However, we also need to consider the case when there exist noise in the communication channel. Without loss of generality, we assume that the probability of erroneous transmission w is the same over channels between A_{MF} and U, and between A_{MF} and V.[6]

Theorem 3. *An attacker A_{MF} interacts with the legitimate user U and the verifier V as seen in Figure 4. If w denotes the probability of erroneous transmission then the chance of erroneous response from A_{MF} in each round is:*

$$\mathbb{P}(Error|A_{MF}) = \frac{1}{4} + \frac{3w}{2} - \frac{7w^2}{2} + 4w^3 - 2w^4$$

Interpretation of the result: it is clear that when there is no noise or $w = 0$ we have $\mathbb{P}(\text{Error}|A_{MF}) = 1/4$ which is the same as previously reported in [14]. For $w = 1/2$ we have $\mathbb{P}(\text{Error}|A_{MF}) = 1/2$, which implies that a mafia fraud attacker does no better than a distance fraud attacker.

[6]Although a noise analysis of mafia fraud attack was provided in [8], the author is very grateful to Christos Dimitrakakis for confirming an error in the analysis.

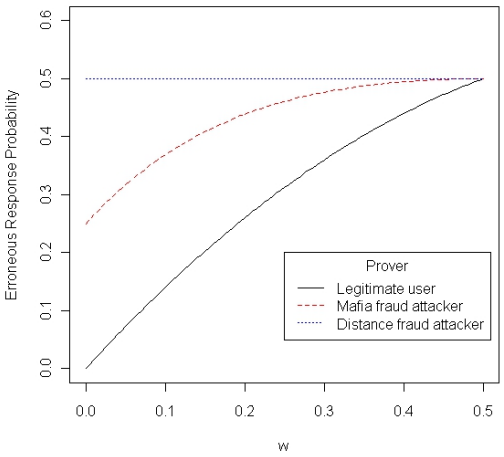

Figure 5: This is the erroneous response probability when the prover is (1) a user, (2) a mafia fraud attacker, or (3) a distance fraud attacker.

Moreover as we increases w from 0 to 1/2, using differentiation technique it is not hard to prove that $\mathbb{P}(\text{Error}|A_{MF})$ also increases from 1/4 to 1/2, and more importantly

$$\mathbb{P}(\text{Error}|P = A_{DF}) \geq \mathbb{P}(\text{Error}|P = A_{MF}) \geq \mathbb{P}(\text{Error}|P = U)$$

This can be graphically illustrated by Figure 5. A mafia fraud attacker is therefore more powerful than a distance fraud attacker, which is indeed the case. Although it is more complicated to analyse the performance of a mafia fraud attacker, we can use the same analysis of Section 4 for a distance fraud attacker to tackle this problem. In the following we present a sketch of the proof, a formal (and detailed) proof is given in full version of this paper [9].

PROOF. (*Sketch*) Following the strategy of a mafia fraud attacker over noisy channel described above and in Figure 4, in any round $i \in \{1, \cdots, n\}$ the legitimate verifier V sends a challenge c_i which is intercepted by the attacker A_{MF}. The challenge c'_i the attacker receives however might not equal c_i due to noise existing over the communication channel between V and A_{MF}, and hence we have two possibilities:

- With probability $1 - w$ we have $c_i = c'_i$, this leads to two further cases:
  - With probability 1/2: $c'_i \neq ac_i$ where ac_i is the anticipated challenge previously sent from A_{MF} posing as V to the legitimate user U. The best A_{MF} can do is to make a random guess $r'_i \in_R \{0, 1\}$ as a response to the verifier V.
  - With probability 1/2: $c'_i = ac_i$. A_{MF} therefore can forward the response r'_i, which A_{MF} previously received from U, to the verifier. Since there is noise over the communication channel between A_{MF} and U, we can use Theorem 1 to compute $\mathbb{P}(r'_i \neq R_i(c'_i))$.

  The response r'_i is then sent from A_{MF} to V, but again it can be corrupted due to noise. r''_i denotes what V

receives at the other end of the communication. Using calculation detailed in [9], we arrive at:

$$\mathbb{P}(r_i'' \neq R_i(c_i)|c_i = c_i') = \frac{1}{4} + \frac{5w}{4} - 2w^2 + w^3$$

- With probability w: $c_i \neq c_i'$, but A_{MF} is not aware of this fact and therefore still follows his or her strategy by first comparing c_i' with the anticipated challenge ac_i. Using calculation from [9], we arrive at

$$\mathbb{P}(r_i'' \neq R_i(c_i)|c_i \neq c_i') = \frac{1}{2} - \frac{w}{4} + w^2 - w^3$$

where r_i'' denotes the final response V receives.

Using the above analysis, we can derive the probability of erroneous response from A_{MF} in each round as follows:

$$\begin{aligned}\mathbb{P}(\text{Error}|P = A_{MF}) &= \mathbb{P}(c_i \neq c_i')\mathbb{P}(r_i'' \neq R_i(c_i)|c_i \neq c_i') + \\ &\quad \mathbb{P}(c_i = c_i')\mathbb{P}(r_i'' \neq R_i(c_i)|c_i = c_i') \\ &= w\left(\frac{1}{2} - \frac{w}{4} + w^2 - w^3\right) + \\ &\quad (1-w)\left(\frac{1}{4} + \frac{5w}{4} - 2w^2 + w^3\right) \\ &= \frac{1}{4} + \frac{3w}{2} - \frac{7w^2}{2} + 4w^3 - 2w^4\end{aligned}$$

□

The number of erroneous responses ϵ_A out of n rounds between V and A_{MF} can be represented by the following binomial distribution

$$\epsilon_A \sim Binom\left(n, \frac{1}{4} + \frac{3w}{2} - \frac{7w^2}{2} + 4w^3 - 2w^4\right)$$

We note that for any values of w and τ, both $\mathbb{P}(\epsilon_A < \tau)$ and $\mathbb{P}(\epsilon_U \geq \tau)$ can be easily calculated to a high level of accuracy by using tabulations of the probability mass functions of the normal distribution [1]. And hence an optimal threshold satisfying Inequalities (3) and (4) can also be found by binary search as described in Section 4.1.

6. EXPERIMENTS

Even though Dimitrakakis et al. [3] introduced a general framework enabling an expected loss analysis for distance-bounding protocols, they did not explain how the costs due to false rejection and false acceptance can be derived. In this section we will first attempt to estimate the costs the verifier suffers corresponding to different protocol outcomes. We then use this information to carry out experiments to assess the performance of our approach in locating exactly the optimal threshold that can resist a rational attacker.

Let us suppose that the intruder is most interested in causing incorrect authentication which is either false rejection or false acceptance, and completely ignores what (s)he might further benefit from the protocol outcomes.[7] Consequently the costs are closely related to the duration of time between successive protocol runs under different circumstances, because these durations determine how long incorrect authentication is maintained following either a false acceptance or a false rejection.

[7]We do not consider what the intruder might further benefit from incorrect authentication, because the corresponding losses vary widely in practice, and hence a careful analysis for each scenario will be required to determine the losses.

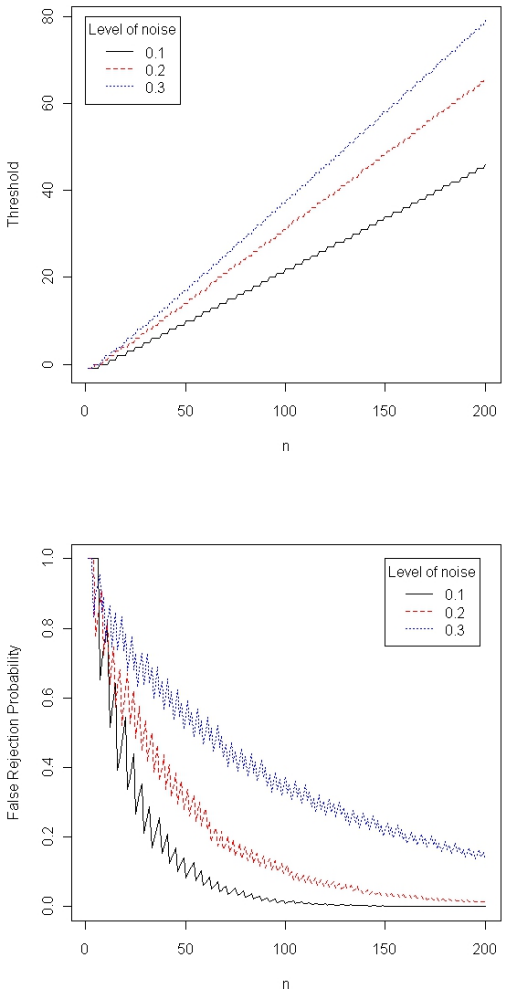

Figure 6: Under the presence of a mafia fraud attacker: these graphs depict the optimal threshold τ and the corresponding probability of false rejection $\mathbb{P}(\epsilon_U \geq \tau)$ versus the number of rounds n.

As an example, we can further extend any distance-bounding protocols to take into account the duration between successive runs as follows:

- False rejection: If a verifier fails to authenticate a legitimate user U then another session will be initiated by U after a short period of time, say 1 hour.
- False acceptance: If a verifier authenticates a malicious party then it will take a longer time, say up to 10 hours, before the protocol is run again because neither the verifier nor the legitimate user is aware of the attack.

Using this information, we have carried out two experiments to assess our analysis regarding distance and mafia fraud attackers, where the ratio $l_A/l_U = 10$ is chosen due to the durations of time between successive runs as specified above. The exact values for l_U, l_A and l_B do not play any role in

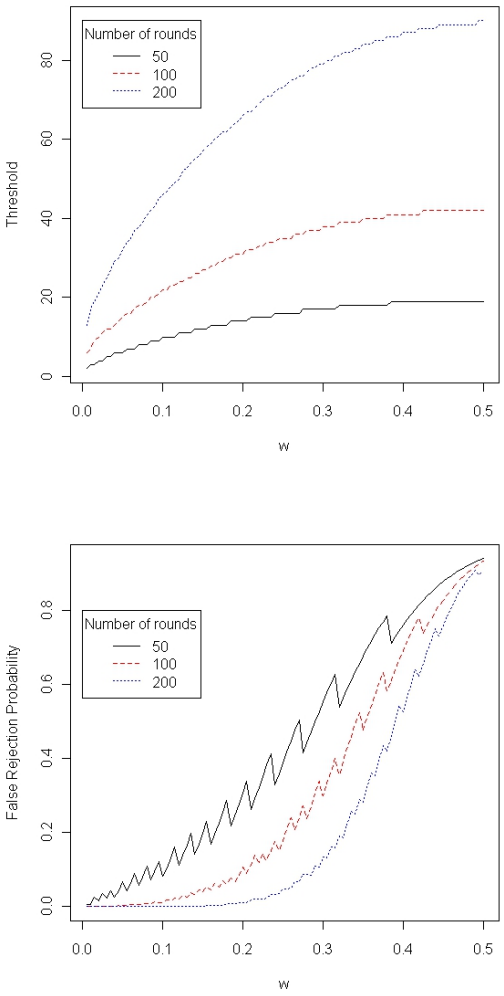

Figure 7: Under the presence of a mafia fraud attacker: these graphs depict the optimal threshold τ and the corresponding probability of false rejection $\mathbb{P}(\epsilon_U \geq \tau)$ versus noise.

the computation of the optimal threshold, and hence are ignored here. Although our analysis is applicable to different levels of noise over communication channels between the different pairs of entities, for simplicity we will assume that the probabilities of erroneous transmission over all channels are equal to one another in our experiments. Since a mafia fraud attacker is more powerful than a distance fraud one, our experiments reported here only look at the optimal threshold that resists a mafia fraud attacker.[8]

In the first experiment, we fix the level of noise w and let the number of rounds n vary from 1 to 200. This experiment is repeated three times with different values for w: 0.1, 0.2 and 0.3 under the presence of a mafia fraud attacker. The first graph of Figure 6 depicts the optimal threshold τ corresponding to each value of $n \in [1, 200]$. With each value of n, and its corresponding optimal threshold τ, we can calculate the probability of false rejection $\mathbb{P}(\epsilon_U \geq \tau)$ which is plotted in the second graph of Figure 6. It is encouraging to see that as the number of rounds increases, the likelihood of false rejection $\mathbb{P}(\epsilon_U \geq \tau)$ falls sharply. Moreover $\mathbb{P}(\epsilon_U \geq \tau)$ decreases faster with less noisy channel.

In the second experiment, we fix the number of rounds n and let the level of noise w vary from 0 to 0.5. This experiment is also repeated three times with different values for n: 50, 100 and 200 again under the presence of a mafia fraud attacker. Figure 7 depicts the optimal threshold τ and the probability of false rejection $\mathbb{P}(\epsilon_U \geq \tau)$ corresponding to each value of $w \in [0, 05]$. It is clear that a higher level of noise leads to bigger threshold as well as the likelihood of false rejection. Moreover, $\mathbb{P}(\epsilon_U \geq \tau)$ increases faster with a smaller round number.

7. COMPARATIVE ANALYSIS

Dimitrakakis et al. [3] introduced a different notion for an optimal threshold (or *minimax* threshold) that minimises the worst-case expected loss $\mathbb{E}(L)$ defined as follows:

$$\mathbb{E}(L) \quad = \quad max\{\mathbb{E}(L|P = U), \mathbb{E}(L|P = A)\} \tag{5}$$

As previously observed, when the threshold τ increases, $\mathbb{E}(L|P = A, \tau)$ also increases while $\mathbb{E}(L|P = U, \tau)$ decreases, and thus to minimise $\mathbb{E}(L)$ we need to locate a threshold such that

$$\begin{aligned} \mathbb{E}(L|P = U, \tau) &= \mathbb{E}(L|P = A, \tau) \\ \frac{\mathbb{P}(\epsilon_U \geq \tau)}{\mathbb{P}(\epsilon_A < \tau)} &= \frac{l_A}{l_U} \end{aligned} \tag{6}$$

In [3], Dimitrakakis et al. give the following formula to compute the nearly-optimal value for this kind of threshold.

$$\tau^* \quad = \quad \frac{n(p_A + p_U)}{2} - \frac{\log(l_A/l_U)}{4(p_A - p_U)} \tag{7}$$

where $p_A = \mathbb{P}(\text{Error}|P = A)$ and $p_U = \mathbb{P}(\text{Error}|P = U)$ are the probabilities of erroneous response from the attacker and respectively the legitimate user in each round.

Minimising the worst-case expected loss is useful, there are however two further points we want to mention here.

- As pointed out in [3] the nearly optimal threshold τ^* is constrained to be between np_A and np_U, for otherwise τ^* would go to negative infinity when $p_A = p_U$. Given that $l_A \geq l_U$ and $p_A \geq p_U$, we always have $\tau^* \leq np_A$. The other condition $\tau^* \geq np_U$ is equivalent to

$$p_A - p_U \quad \geq \quad \sqrt{\frac{\log(l_A/l_U)}{2n}} \tag{8}$$

  From Theorems 1–3 and Figure 4, when the noise level w approaches $1/2$, $p_A - p_U$ decreases toward zero. This means that τ^* is not applicable for w that is near to $1/2$. However as the number of round n increases, τ^* works with bigger range of noise as seen in Inequality 8. In contrast our method introduced here can locate the minimax threshold exactly for any value of w.

  In Figure 8, we plot the actual minimax threshold computed by our strategy as a function of noise level $w \in [0, 0.5]$ under the presence of a mafia fraud attacker. The number of round n is fixed at 100. We

[8]We have also carried out experiments under the presence of a distance fraud attacker whose results are very similar to those reported here.

also plot the nearly optimal threshold τ^* on the same graph but only for $w \in [0, 0.32]$ because Inequality 8 does not hold when $w > 0.32$. This clearly shows that the nearly-optimal threshold τ^* gives good approximation for a limited range of noise.

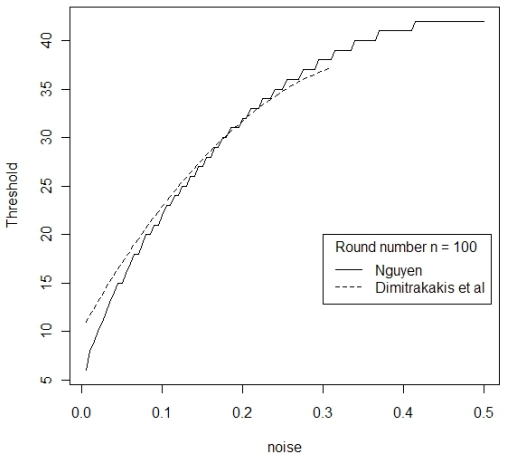

Figure 8: Under the presence of a mafia fraud attacker: the graph depicts the optimal and nearly-optimal minimax threshold τ versus noise.

- The minimisation of $\mathbb{E}(L)$ does not give any conclusive indication about the relative difference between $\mathbb{E}(L|P = U)$ and $\mathbb{E}(L|P = A)$, which is crucial in our goal of resisting a rational intruder as explained in Section 3. In other words, when an integer threshold τ satisfying Equality (6) does not exist, the minimisation of $\mathbb{E}(L)$ leads to either $\mathbb{E}(L|P = U, \tau) > \mathbb{E}(L|P = A, \tau)$ or $\mathbb{E}(L|P = U, \tau) < \mathbb{E}(L|P = A, \tau)$, where the latter will even encourage a rational intruder to attack.

  This implies that even though the optimal threshold defined by Dimitrakakis et al. is similar to ours, they are not the same.

8. CONCLUSIONS

We have introduced the use of ideas from game theory to define a new notion for an optimal threshold for the number of erroneous responses during the rapid-bit exchange of distance-bounding protocols to make the protocols resilient against a rational intruder. The advantage of our approach based on statistical analysis and binary search over existing results in this area is that given any level of noise and round number we can locate exactly the optimal threshold value. Experimental results are provided to demonstrate the accuracy of our analysis when being used to discourage a mafia fraud attacker from attacking a protocol of this type.

In addition to distance-bounding protocols, we have also investigated the relevance of the notion of a rational intruder to other types of authentication protocols which are based on passwords or human interactions [11, 12, 13]. Readers who are interested in this area can find out more at [10].

Acknowledgements: The author is financially supported by the US Office of Naval Research. The idea developed here arose from fruitful discussions with Dr. Aikaterini Mitrokotsa while the author was visiting Prof. Serge Vaudenay at the Security and Cryptography Laboratory at EPFL from 7th Febraury to 7th March 2011. The author is indebted to Prof. Serge Vaudenay for giving him the opportunity to visit EPFL and would like to thank Drs Aikaterini Mitrokotsa and Christos Dimitrakakis for their many helpful comments.

9. REFERENCES

[1] M. Abramowitz and I.A. Stegun. *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables.* ISBN 0-486-61272-4.

[2] L. Buttyán, Jean-Pierre Hubaux, and S. Čapkun. *A Formal Analysis of Syverson's Rational Exchange Protocol.* Proceedings of the 15th IEEE workshop on Computer Security Foundations, 2002.

[3] C. Dimitrakakis, A. Mitrokotsa and S. Vaudenay. *Expected loss analysis of thresholded authentication protocols in noisy conditions.* See `http://arxiv.org/pdf/1009.0278`

[4] S.D. Gordon and J. Katz. *Rational secret sharing, revisited.* In Proceedings of Security and Cryptography for Networks. LNCS vol. 4116, 229-241, 2006.

[5] J. Halpern and V. Teague. *Rational Secret Sharing and Multiparty Computation.* In Proceedings of the thirty-sixth annual ACM symposium on Theory of computing STOC '04. Pages 623-632, 2004.

[6] G. Hancke and M. Kuhn. *An RFID distance-bounding protocol.* Proceedings SecureComm 2005.

[7] C.H. Kim, G. Avoine, F. Koeune, F.X. Standaert, and O. Pereira. *The Swiss-knife RFID distance bounding protocol.* In Proceedings of ICISC 2008.

[8] A. Mitrokotsa, C. Dimitrakakis, P. Peris-Lopez, J.C. Hernandez-Castro. *Reid et al.'s Distance Bounding Protocol and Mafia Fraud Attacks over Noisy Channels.* IEEE Communications Letters, Feb 2010, Vol. 14, No. 2, pp. 121-123.

[9] L.H. Nguyen. *Rational distance-bounding protocols over noisy channels.* With formal proof of Theorem 3. See `http://eprint.iacr.org/2011/492`

[10] L.H. Nguyen. *Rational authentication protocols.* See `http://eprint.iacr.org/2011/070.pdf`

[11] L.H. Nguyen (editor), second edition. ISO/IEC 9798-6 (2010): *Information Technology – Security Techniques – Entity authentication – Part 6: Mechanisms using manual data transfer.*

[12] L.H. Nguyen and A.W. Roscoe. *Authenticating ad hoc networks by comparison of short digests.* Information and Computation 206 (2008), 250-271.

[13] L.H. Nguyen and A.W. Roscoe. *Authentication protocols based on low-bandwidth unspoofable channels: a comparative survey.* Journal of Computer Security, Vol. 9, No. 1 (2011), 139-201.

[14] J. Reid, J.M. Gonzalez Nieto, T. Tang, and B. Senadji. *Detecting relay attacks with timing-based protocols.* In Proc. ASIACCS 2007.

[15] P. Syverson. *Weakly secret bit commitment: Applications to lotteries and fair exchange.* The IEEE Computer Security Foundations Workshop, 2-13, 1998.

APPENDIX

A. DETAILED PROOF OF THEOREM 3

Theorem 3. *An attacker A_{MF} can interact with both the legitimate user U and the verifier V. If w denote the probabilities of erroneous transmission then the probability of erroneous response from A_{MF} in each round is:*

$$\mathbb{P}(\text{Error}|A_{MF}) = \frac{1}{4} + \frac{3w}{2} - \frac{7w^2}{2} + 4w^3 - 2w^4$$

This proof should be read along side Figure 4 which clearly depicts not only notations but also the strategy of a mafia fraud attacker. The order of reasoning and calculation of this proof is the same as in the sketch of this proof given in Section 5.

Proof. We first have the following

$$\begin{aligned}\mathbb{P}(\text{Error}|A_{MF}) &= \mathbb{P}(c_i \neq c_i')\mathbb{P}(r_i'' \neq R_i(c_i)|c_i \neq c_i') + \\ &\quad \mathbb{P}(c_i = c_i')\mathbb{P}(r_i'' \neq R_i(c_i)|c_i = c_i') \\ &= w \cdot \mathbb{P}(r_i'' \neq R_i(c_i)|c_i \neq c_i') + \\ &\quad (1-w) \cdot \mathbb{P}(r_i'' \neq R_i(c_i)|c_i = c_i') \quad (9)\end{aligned}$$

We now attempt to calculate $\mathbb{P}(r_i'' \neq R_i(c_i)|c_i = c_i')$ and $\mathbb{P}(r_i'' \neq R_i(c_i)|c_i \neq c_i')$ respectively.

Using information from Figure 4 and the sketch of this proof given in Section 5, we have

$$\begin{aligned}\mathbb{P}(r_i'' \neq R_i(c_i)|c_i = c_i') &= \mathbb{P}(c_i' \neq ac_i \wedge r_i' = R_i(c_i) \wedge r_i'' \neq r_i'|c_i = c_i') + \\ &\quad \mathbb{P}(c_i' \neq ac_i \wedge r_i' \neq R_i(c_i) \wedge r_i'' = r_i'|c_i = c_i') + \\ &\quad \mathbb{P}(c_i' = ac_i \wedge r_i' = R_i(c_i) \wedge r_i'' \neq r_i'|c_i = c_i') + \\ &\quad \mathbb{P}(c_i' = ac_i \wedge r_i' \neq R_i(c_i) \wedge r_i'' = r_i'|c_i = c_i') \\ &= \frac{1}{2}\frac{1}{2}w + \frac{1}{2}\frac{1}{2}(1-w) + \frac{1}{2}[1 - w(3/2-w)]w + \frac{1}{2}w(3/2-w)(1-w) \\ &= \frac{1}{4} + \frac{5w}{4} - 2w^2 + w^3 \quad (10)\end{aligned}$$

The second equality holds due to Theorem 1 which gives us

$$\begin{aligned}\mathbb{P}(r_i' \neq R_i(c_i)|c_i = c_i' = ac_i) &= w(3/2-w) \\ \mathbb{P}(r_i' = R_i(c_i)|c_i = c_i' = ac_i) &= 1 - w(3/2-w)\end{aligned}$$

Also using information from Figure 4 and the sketch of this proof given in Section 5, we have

$$\begin{aligned}\mathbb{P}(r_i'' \neq R_i(c_i)|c_i \neq c_i') &= \mathbb{P}(r_i'' = r_i')\mathbb{P}(r_i' \neq R_i(c_i)|c_i \neq c_i') + \\ &\quad \mathbb{P}(r_i'' \neq r_i')\mathbb{P}(r_i' = R_i(c_i)|c_i \neq c_i') \\ &= (1-w) \cdot \mathbb{P}(r_i' \neq R_i(c_i)|c_i \neq c_i') + \\ &\quad w \cdot \big[1 - \mathbb{P}(r_i' = R_i(c_i)|c_i \neq c_i')\big] \quad (11)\end{aligned}$$

This probability $\mathbb{P}(r_i' \neq R_i(c_i)|c_i \neq c_i')$ can be computed as follows

$$\begin{aligned}\mathbb{P}(r_i' \neq R_i(c_i)|c_i \neq c_i') &= \mathbb{P}(c_i' = ac_i \wedge ac_i = ac_i' \wedge r_i' = R_i(ac_i') \wedge R_i(c_i) \neq R_i(c_i')|c_i \neq c_i') + \\ &\quad \mathbb{P}(c_i' = ac_i \wedge ac_i = ac_i' \wedge r_i' \neq R_i(ac_i') \wedge R_i(c_i) = R_i(c_i')|c_i \neq c_i') + \\ &\quad \mathbb{P}(c_i' = ac_i \wedge ac_i \neq ac_i' \wedge r_i' \neq R_i(ac_i')|c_i \neq c_i') + \\ &\quad \mathbb{P}(c_i' \neq ac_i \wedge ac_i \neq ac_i'|c_i \neq c_i') \\ &= \frac{1}{2}(1-w)(1-w)\frac{1}{2} + \frac{1}{2}(1-w)w\frac{1}{2} + \frac{1}{2}w^2 + \frac{1}{2}\frac{1}{2} \\ &= \frac{1}{2} - \frac{w}{4} + \frac{w^2}{2} \quad (12)\end{aligned}$$

The second equality follows because we always have

$$\mathbb{P}(R_i(c_i) = R_i(c_i')|c_i \neq c_i') = \mathbb{P}(R_i(c_i) \neq R_i(c_i')|c_i \neq c_i') = 1/2$$

as pointed out at the end of Section 2.1 and in Footnote 4.

Substituting Equality (12) into Formula (11) arrives at

$$\mathbb{P}(r_i'' \neq R_i(c_i)|c_i \neq c_i') = \frac{1}{2} - \frac{w}{4} + w^2 - w^3 \quad (13)$$

Substituting Equalities (10) and (13) into Formula (9) gives us the proof. □