Rational authentication protocols

Long H. Nguyen
Oxford University Computing Laboratory
Email: Long.Nguyen@comlab.ox.ac.uk

Abstract. We use ideas from game theory to transform two families of authentication protocols, namely password-based and manual authentication schemes, so that even if an intruder attempts to cause incorrect authentication, his expected payoff will still be lower than when he does not attack. A *rational* intruder, who always tries to maximise his payoff, therefore has no incentive to disrupt the protocol. Our strategy is useful when a powerful and rational intruder makes many attempts to break a protocol because the potential reward from a successful attack outweighs the cost of many unsuccessful attempts, and moreover (1) even if the intruder fails then a denial of service is still mounted, and (2) in a password-based protocol, the chance of a successful attack increases quite significantly as more and more attempts are launched to guess the password. We provide a case study to assess the performance of our work.

1 Introduction

Ideas from game theory have been used to re-design a number of fair exchange protocols [2, 11] and secret sharing schemes [7, 5, 8] so that parties cannot act on their own interests to bring these schemes to failure [8]. 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 authentication and key-agreement protocols where all honest nodes should cooperate to complete a protocol successfully, because it is in their mutual interest that they agree on the same data.

We instead observe that in a hostile environment, such as battle fields or the financial intrustry, there is always the presence of a powerful intruder who can intercept and modify data transmitted over a communication channel, such as WiFi or the Internet. In addition, the intruder is *rational* in the sense that it always tries to maximise its payoff as in the following scenario. Suppose that a party A has borrowed some money from a lending company, but A intends to delay payment against the wish of the lender who does not trust A. To make a loan, the lender must have previously seen financial proofs from A such as a bank statement, and thus knows that A potentially possesses a very large amount of money in his account. The lender, who can be the mafia or the intruder in disguise, will therefore be highly motivated to break into A's bank account

and steal all of A's money, i.e. this could be achieved by attacking the authentication stage of online transactions carried out between A and the bank. In practice, a delayed payment is usually resolved by A paying an interest, however the disadvantage of this solution is that an interest rate does not take into account the true identity and computational capability of the lender as well as A's bank account balance. The latter can imply that the potential reward that an attacker can get from having access to A's account exceeds the cost of launching many unsuccessful attacks. It is worth to point out that even if the intruder fails, he still manages to mount a denial of service, and thus prevents honest parties from agreeing on the same data. In a password-based protocol, an incorrect guess of a short password means that the chance of correctly guessing the password will increase quite significantly in subsequent runs, which further encourages the intruder to mount another attack. Consequently when the intruder is rational he potentially does better by attacking many protocol runs. The observations therefore motivate us to use techniques in game theory to redesign authentication protocols to resist this kind of rational intruder.

Our first contribution presented in Section 2 is a general transformation in which we introduce *irrational* behaviours an honest node, who is usually the initiator of a protocol, can pursue under some probability, such that an intruder's expected payoff in attempting to cause incorrect authentication does not exceed his expected payoff in not attacking the protocol. The intruder therefore does not have any incentive to cause disruption, which is very similar to the concept of Nash equilibrium in game theory. Since the irrational behaviours of the initiator must be in the interest of the intruder, they vary from one to another applications, and therefore we will largely abstract away from their details which are not immediately important to our analysis. The main thrust of this paper is to formally demonstrate how this protocol transformation works and benefits two families of pairwise authentication protocols, namely password-based authentication (or key agreement) schemes [1] of Section 3 and manual authentication protocols [3, 14] of Section 4, though other cryptographic protocols such as distance-bounding schemes [4] also benefit from our work. To assess the performance of our analysis, in Section 5 we present a case study where we show how a party A who borrowed money from a company can discourage that company from breaking into A's bank account. The study also shreds new light on the usability and economic security of current banking applications regarding the limit of number of failed attempts of entering the correct password. We conclude by drawing similarities between the use of irrational behaviours in our protocol

transformation and the business of insurance and lottery. In Annex A, we show how this strategy can be easily adapted to group protocols.

While we believe that we are the first to study the notion of including irrational behaviours in honest parties' activities tailored specifically for authentication protocols, such idea can be traced back to earlier work in other context of rational secret sharing protocols. To encourage an intruder to give up attacking a protocol, it is probably inevitable that we need to give something, which is less damaging than a successful attack, to the intruder in each normal run. Both rational secret sharing schemes of Gordon and Katz [7] and Fuchsbauer et al. [5] follow this strategy by allowing a trusted dealer to send invalid shares of secret to players at the beginning of some iterations, or forcing nodes to proceed in a sequence of fake runs followed by a single real one.

2 Protocol transformation

For simplicity pairwise authentication schemes are considered, where two trustworthy parties A and B want to authenticate or agree on the same data, though our protocol transformation and analysis can be generalised to group scenarios as presented in Annex A. In the schemes, it is in honest nodes' mutual interest that they follow the protocol. Also among the protocol participants, there is always one party who initiates a protocol by, for example, sending the first message as seen in protocols of Sections 3 and 4, and hence we always denote A the protocol *initiator*. No specific protocol is given until multiple-run attacks are considered in subsequent sections, because for single-run attacks our suggested changes in the behaviour of the initiator A are independent of the type of authentication protocols whether they are based on passwords [1] or human interactions [14]. These changes, which are summarised in Tables 1 and 2, aim to discourage the intruder from causing incorrect authentication.

In our protocol transformation, we give the intruder different payoffs $U, U_1^+, U_2^+, U_1^-, U_2^-$, corresponding to different combination of parties' strategies and protocol outcomes. The payoffs depend on a number of factors, including the cost of launching attacks (computation or energy consumption) and financial reward of a successful attempt, as will be demonstrated in our case study of Section 5. From Table 1, while we should always have $min\{U_1^+, U_2^+\} > max\{U_1^-, U_2^-\}$, the difference among these payoffs for the intruder when it attacks a protocol can vary widely, e.g. U_1^+ and U_2^+ can be either far apart or roughly the same. We therefore will only tackle the worst case scenario here: regardless of whether

Protocol transformation

In an authentication protocol where parties authenticate each other or seek to agree on the same data, we introduce the following behaviour that the protocol initiator A can pursue to resist a rational intruder.

- With probability $\alpha \in [0, 1)$: the initiator A is unfaithful and will pursue some irrational behaviours which must benefit the intruder, even though A still follows every step of a protocol properly. We refer to this event as a fake protocol run. The exact irrational behaviours depend on the intruder's goals in different scenarios. For example, A might exchange some random data of no use, and hence it is a waste of time for the other honest protocol participant B. Alternatively A can transfer a tiny amount of money to a third party who can be the intruder in disguise as seen in our case study of Section 5. We therefore abstract away the details of the irrational behaviours which are not immediately important to our analysis.

  When the intruder does not attack then a fake protocol run always completes successfully, and we denote U the intruder's payoff and respectively V the collective payoff for the protocol participants.

- With probability $1 - \alpha$: party A faithfully follows the protocol, and thus if the intruder does not misbehave then the protocol run will succeed, we denote V^+ the protocol participants' collective payoff. There is no payoff for the intruder in this case.

When the intruder attacks and manipulates data in a protocol run, the protocol participants cannot agree on the same data and are not given any payoff. Regarding of the intruder's payoff, there are two possibilities:

- With probability ϵ the intruder succeeds and is given payoff U_1^+ when A is faithful, and U_2^+ when A is unfaithful.

- With probability $1 - \epsilon$ the intruder fails, but is still given payoff U_1^- when A is faithful, and U_2^- when A is unfaithful. The payoffs can be however negative if the cost of lauching an attack is included.

Since the intruder always benefits from a successful attack regardless of whether A is faithful or not, we therefore arrive at $min\{U_1^+, U_2^+\} > U > 0$. Also the intruder would prefer a fake protocol run to an unsuccessful attack, and hence $U > max\{U_1^-, U_2^-\}$.

Table 1. Protocol transformation.

Strategy of intruder	Strategy of initiator A	Outcome of protocol	Payoff of intruder	Payoff of participants
No attack	Faithful	Succeed	0	V^+
No attack	Unfaithful	Succeed	U	V
Attack	Faithful	Succeed	U_1^+	0
Attack	Unfaithful	Succeed	U_2^+	0
Attack	Faithful	Fail	U_1^-	0
Attack	Unfaithful	Fail	U_2^-	0
The lower half of this table is the worst case scenario of the upper half				
No attack	Faithful	Succeed	0	V^+
No attack	Unfaithful	Succeed	U	V
Attack	Any	Succeed	$U^+ = max\{U_1^+, U_2^+\}$	0
Attack	Any	Fail	$U^- = max\{U_1^-, U_2^-\}$	0

Table 2. A summary of the game.

A is faithful or not the intruder's payoff is $U^+ = max\{U_1^+, U_2^+\}$ when it succeeds, and respectively $U^- = max\{U_1^-, U_2^-\}$ when it fails as seen in the lower half of Table 2. A solution for the worst case scenario applies to every other scenario where $U_1^+ \neq U_2^+$ and/or $U_1^- \neq U_2^-$.

Using the protocol transformation specified in Table 1, we arrive at the following theorem. We note that once parties can discourage an intruder from attacking a protocol, we will turn into maximising the collective payoff of honest protocol participants (from agreeing on the same data).

Theorem 1. If an intruder can only attack up to a single run of an authentication protocol and succeed with probability ϵ, then to discourage the intruder from attacking, this inequality must hold:

$$\alpha > \frac{\epsilon U^+ + (1-\epsilon)U^-}{U}$$

Proof. If the intruder does not misbehave, his expected payoff in each run is αU. If the intruder misbehaves, his expected payoff of a single-run attack is $P = \epsilon U^+ + (1-\epsilon)U^-$. To achieve our aim, we need to find the value of α such that the following inequality holds:

$$\alpha U > \epsilon U^+ + (1-\epsilon)U^-$$

$$\alpha > \frac{\epsilon U^+ + (1-\epsilon)U^-}{U}$$

So as long as this is true, it is in the intruder's interest not to attack any protocol run. □

From Tables 1 and 2, we have $U^+ > U > U^-$. Although U^+ is normally significantly bigger than U, ϵU^+ should be small relative to U when passwords and short authentication strings in authentication protocols considered in this paper are chosen to be of reasonable length.

Suppose that there are a number of strategies regarding different values of α that node A can pursue: $\alpha \in \{\alpha_1, \alpha_2, \cdots, \alpha_w\}$ where

$$0 \leq \alpha_1 < \alpha_2 < \cdots < \alpha_w < 1$$

Since the main priority is to discourage the intruder from attacking a protocol run, according to Theorem 1 node A will only select strategies with $\alpha \in \{\alpha_i, \alpha_{i+1}, \cdots, \alpha_w\}$ such that $\alpha_i U > P$ where $P = \epsilon U^+ + (1-\epsilon)U^-$. Given that this condition is satisfied (i.e. the intruder does not attack), we further want to maximise the expected collective payoff of protocol participants in each run which is derived from Tables 1–2 to be

$$Q = (1-\alpha)V^+ + \alpha V$$

As honest parties prefer a real protocol run to a fake run, we have $V < V^+$. Additionally $\alpha_i < \alpha_{i+1} < \cdots < \alpha_w < 1$, it is clear from the payoff matrix of Table 3 that node A will pick α_i as the optimal strategy. In game theory, it is easy to show that $\alpha = \alpha_i$ and no-attack intruder constitute a strict and also unique Nash equilibrium for the game depicted in Table 3.

Strategy of Intruder	Strategy of the initiator A		
	α_i	$\cdots$	α_w
No attack	$\alpha_i U$, $(1-\alpha_i)V^+ + \alpha_i V$	$\cdots$	$\alpha_w U$, $(1-\alpha_w)V^+ + \alpha_w V$
Attack	P, 0	$\cdots$	P, 0

Table 3. This matrix shows expected payoffs for the intruder and the protocol participants with each combination of their strategies in each protocol run.

The above analysis only takes into account single-run attacks, in practice a rational intruder as defined in Section 1 would attack many protocol runs until (s)he is successful. For this reason, it is desirable that we consider the case of multiple-run attacks on authentication protocols.

3 Multiple-run attacks on password-based protocols

Any secure password-based (authentication or key-agreement) protocol needs to resist off-line searching, i.e. the only way to find out a guess of a password is correct is to interact with the protocol. Our analysis here applies to many secure password-based protocols, but for clarity we give the definition of the Diffie-Hellman-based Encrypted Key Exchange scheme of Bellovin and Merritt [1]. This protocol establishes a shared private key g^{xy}, where g^x and g^y are Diffie-Hellman keys of A and B, from a short password pw using an encryption scheme $E_{pw}()$ and a cryptographic hash function $hash()$. Since passwords are usually very short and unchanged for a period of time, the chance of a successful attack increases quite significantly as more and more attempts are launched to guess the passwords. We stress that this feature of a password-based scheme, which is different from other kinds of authentication protocol, is particularly relevant to our discussion, because it will encourage the intruder to keep guessing the password in many protocol runs until (s)he gets it right.

1. $A \longrightarrow B \ : \ A \parallel E_{pw}(g^x)$
2. $B \longrightarrow A \ : \ E_{pw}(g^y) \parallel hash(sk \parallel 1)$
   where $sk = hash(A \parallel B \parallel g^x \parallel g^y \parallel g^{xy})$
3. $A \longrightarrow B \ : \ hash(sk \parallel 2)$

In practice we usually limit the number of failed attempts an intruder can make, e.g. three wrong guesses and the protocol will stop running, and thus we denote k the limit of number of attacks an intruder can launch on a protocol. If a password is randomly selected from $\{1, \ldots, n\}$, then[1] $1 \leq k \leq n$ and the chance of correctly guessing the password the first time is $\epsilon = \epsilon_1 = 1/n$. If an attacker's first guess is incorrect, then the second guess is successful with probability $\epsilon_2 = 1/(n-1)$. For all $k \in \{1, \ldots, n\}$ we have $\epsilon_k = 1/(n-k+1)$. We emphasise that this increase in the likelihood of successful guess in subsequent protocol runs is correct even when the initiator A is unfaithful, because an unfaithful initiator by definition from the protocol transformation of Table 1 still follows every protocol step properly, including authentication checks that involve the use of the password in Messages 3 and 4 of the above protocol.

In order to be precise in our arguments, we need to be clear about the attacking strategy of the intruder that our protocol transformation of Table 1 seeks to resist. If the intruder decides to attack a protocol up

[1] This is true when the password is unchanged throughout a multiple-run attack.

to k runs, then the intruder only terminates its attack if either of the following two conditions is met:

- The intruder succeeds in the tth attempt where $t \leq k$ or[2]
- The intruder fails in all k attempts.

This strategy includes the scenario when the intruder masquerades as the initiator A to communicate with honest node B, because the intruder still needs to guess the password correctly. Also these k attempts do not need to be consecutive and can be interleaved with any number of protocol runs which are not attacked by the intruder. We summarise the intruder's cumulative payoff and probability that it is successful or unsuccessful up to k attempts in Table 4.

No. of attempts	Outcome	Probability	Payoff of intruder
1	Succeed	$\epsilon = \epsilon_1 = 1/n$	U^+
2	Succeed	$(1-\epsilon_1)\epsilon_2 = 1/n$	$U^- + U^+$
3	Succeed	$(1-\epsilon_1)(1-\epsilon_2)\epsilon_3 = 1/n$	$2U^- + U^+$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	Succeed	$\epsilon_t \Pi_{i=1}^{t-1}(1-\epsilon_i) = 1/n$	$(t-1)U^- + U^+$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	Succeed	$\epsilon_k \Pi_{i=1}^{k-1}(1-\epsilon_i) = 1/n$	$(k-1)U^- + U^+$
k	Fail	$\Pi_{i=1}^{k}(1-\epsilon_i) = (n-k)/n$	kU^-

Table 4. This tables shows the cumulative payoff and probability of the intruder's success and failure when (s)he attacks a password-based protocol up to k runs.

The following theorem shows that as k increases the probability α that the initiator A behaves irrationally also goes up but very slowly.

Theorem 2. Suppose that an intruder is allowed to attack a password-based protocol up to k runs for any $k \in \{1, \ldots, n = 1/\epsilon\}$, and the intruder quits iff (s)he is successful in the tth attempt where $t \leq k$ or fails in all k attempts as seen in Table 4. Then to discourage the intruder from

[2] If A is running an authentication protocol with a bank where A has an account, and the intruder successfully guesses the password in the tth attempt, then the intruder will access the bank account to steal all of A's money. The intruder does not have any incentive to continue its multiple-run attack because the account balance is zero.

attacking the protocol, this inequality must hold:

$$\alpha > \frac{\epsilon U^+ + (1-\epsilon)U^-}{U} + \left(\frac{U^+ - U^-}{U}\right)\frac{k-1}{n(2n-k+1)}$$

Proof. When an intruder attacks a protocol up to k runs, from Table 4, the expected (average) number of protocol runs the intruder intervenes is

$$N = \frac{1}{n} + \frac{2}{n} + \frac{3}{n} + \cdots + \frac{k}{n} + \frac{k(n-k)}{n} = \frac{k(2n-k+1)}{2n}$$

Similarly, the expected cumulative payoff of the intruder's multiple-run attack can be computed as follows

$$\begin{aligned} P &= \frac{U^+}{n} + \frac{U^- + U^+}{n} + \cdots + \frac{(k-1)U^- + U^+}{n} + \frac{k(n-k)U^-}{n} \\ &= \frac{kU^+}{n} + U^-\left[\frac{1}{n} + \frac{2}{n} + \cdots + \frac{k-1}{n} + \frac{k(n-k)}{n}\right] \\ &= \frac{kU^+}{n} + \frac{k(2n-k-1)U^-}{2n} \end{aligned}$$

Since the expected payoff an intruder gets from not attacking a protocol in each run is αU, in order to discourage the intruder from attacking a password-based protocol up to k runs, we must have:

$$\begin{aligned} \alpha UN &> P \\ \alpha &> \frac{kU^+}{nUN} + \frac{k(2n-k-1)U^-}{2nUN} \\ \alpha &> \frac{2U^+}{(2n-k+1)U} + \frac{(2n-k-1)U^-}{(2n-k+1)U} \\ \alpha &> \left(\frac{1}{n} + \frac{k-1}{n(2n-k+1)}\right)\frac{U^+}{U} + \left(1 - \frac{1}{n} - \frac{k-1}{n(2n-k+1)}\right)\frac{U^-}{U} \\ \alpha &> (\epsilon + \Delta)\frac{U^+}{U} + (1-\epsilon-\Delta)\frac{U^-}{U} \\ \alpha &> \frac{\epsilon U^+ + (1-\epsilon)U^-}{U} + \left(\frac{U^+ - U^-}{U}\right)\Delta \end{aligned}$$

where $\Delta = \frac{k-1}{n(2n-k+1)}$. □

Since $n \geq k \geq 1$, as k increases then so do both Δ and α, and moreover $\epsilon > \Delta \geq 0$. This implies that

– The difference between the bounds for α with respect to single-run (see Theorem 1) and n-run attacks is $\left(\frac{U^+-U^-}{U}\right)\Delta < \epsilon\left(\frac{U^+-U^-}{U}\right)$, which can be very small given that the password is random and of a reasonable length. This therefore would affect the current limit for the number of times of entering wrong passwords as usually set in many banking applications. We will discuss this further in Section 5.
– If this protocol transformation can discourage a k-run attack, then it can also discourage a t-run attack for any $t \leq k$.

4 Multiple-run attack on manual authentication protocol

In contrast to password-based schemes, the chance of a successful attack ϵ on a manual authentication protocol run remains unchanged regardless of how many times an attack is launched. This property applies to all secure protocols of this type, whether they provide oneway, pairwise or group authentication.

Our analysis here applies to every secure manual authentication protocol, but for clarity we give a pairwise protocol which is similar to one of Laur and Nyberg [9]. In this scheme, parties A and B want to authenticate their public data $m_{A/B}$ from human interactions to remove the need of passwords, private keys and PKIs. The single arrow ($\longrightarrow$) indicates an unreliable and high-bandwidth link (e.g. WiFi or the Internet) where messages can be maliciously altered, whereas the double arrow ($\Longrightarrow$) represents an authentic and unspoofable channel. The latter is not a private channel (i.e. anyone can overhear it) and it is usually very low-bandwidth since it is implemented by humans, e.g., human conversations, text messages or manual data transfers between devices. $hash()$ and $uhash()$ are cryptographic and universal hash functions. Long random keys $k_{A/B}$ are generated by A/B, and k_A must be kept secret until after k_B is revealed in Message 2. Operators $\|$ and $\oplus$ denote concatenation and exclusive-or.

1. $A \longrightarrow B\ : m_A, hash(k_A)$
2. $B \longrightarrow A\ : m_B, k_B$
3. $A \longrightarrow B\ : k_A$
4. $A \Longleftrightarrow B : uhash(k_A \oplus k_B, m_A \| m_B)$

To ensure both parties share the same data, the human owners of devices A and B have to compare a short universal hash value (or a short authentication string SAS) of 16–32 bits manually. Since the universal hash key $k_A \oplus k_B$ always varies randomly from one to another run, the chance of

a successful attack on each protocol run ϵ equals the collision probability of the universal hash function.[3]

Definition 1. [10] An ϵ-almost universal hash function, $uhash : R \times X \to Y$, must satisfy that for every $m, m' \in X$ and $m \neq m'$:

$$\Pr_{\{k \in R\}}[uhash(k, m) = uhash(k, m')] \leq \epsilon$$

To discourage the intruder from attacking a manual authentication protocol in multiple runs, we use the protocol transformation of Table 1. With probability α the initiator A behaves irrationally, though A still follows every protocol step properly, including an authentication check in Message 4 of the above protocol. Since the chance of a successful single-run attack ϵ is unchanged, intuitively the value of α required to discourage a multiple-run attack is the same as in a single-run attack of Theorem 1. But we will formally state and prove this result in Theorem 3.

Theorem 3. Suppose that an intruder is allowed to attack a manual authentication protocol up to k runs for any $k \geq 1$, and the intruder quits iff (s)he is successful in the tth attempt where $t \leq k$ or fails in all k attempts as seen in Table 5. Then to discourage the intruder from attacking the protocol, this inequality must hold:

$$\alpha > \frac{\epsilon U^+ + (1-\epsilon)U^-}{U}$$

We summarise the intruder's cumulative payoff and probability of success and failure in Table 5.

Proof. When an intruder attacks a protocol up to k runs, from Table 5, the expected number of runs the intruder intervenes in this protocol is:

$$\begin{aligned} N &= \epsilon + 2\epsilon(1-\epsilon) + \cdots + k\epsilon(1-\epsilon)^{k-1} + k(1-\epsilon)^k \\ &= 1 + (1-\epsilon) + (1-\epsilon)^2 + \cdots + (1-\epsilon)^{k-1} \end{aligned} \tag{1}$$

Equality (1) is derived from repeatedly applying the following equality for all $t \in \{1, \cdots, k-1\}$.

$$(1-\epsilon)^t = t(1-\epsilon)^{t-1}\epsilon + (t+1)(1-\epsilon)^t - t(1-\epsilon)^{t-1}$$

[3] We note that our protocol transformation of Table 1 and the analysis of this section also apply to other manual authentication protocols, including schemes of Vaudenay [14] and Čagalj et al. [3], which do not use a universal hash function.

No. of attempts	Outcome	Probability	Payoff of intruder
1	Succeed	ϵ	U^+
2	Succeed	$\epsilon(1-\epsilon)$	$U^- + U^+$
3	Succeed	$\epsilon(1-\epsilon)^2$	$2U^- + U^+$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	Succeed	$\epsilon(1-\epsilon)^{t-1}$	$(t-1)U^- + U^+$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	Succeed	$\epsilon(1-\epsilon)^{k-1}$	$(k-1)U^- + U^+$
k	Fail	$(1-\epsilon)^k$	kU^-

Table 5. This tables shows the cumulative payoff and probability of the intruder's success and failure when (s)he attacks a manual authentication protocol up to k runs.

The expected cumulative payoff of the intruder's multiple-run attack can be computed as follows

$$
\begin{aligned}
P &= \epsilon U^+ + \epsilon(1-\epsilon)(U^- + U^+) + \cdots + \epsilon(1-\epsilon)^{k-1}((k-1)U^- + U^+) + (1-\epsilon)^k kU^- \\
&= U^+\epsilon\left[1 + (1-\epsilon) + \cdots + (1-\epsilon)^{k-1}\right] + \\
&\quad U^-(1-\epsilon)\left[\epsilon + 2\epsilon(1-\epsilon) + \cdots + (k-1)\epsilon(1-\epsilon)^{k-2} + k(1-\epsilon)^{k-1}\right] \\
&= U^+\epsilon N + U^-(1-\epsilon)N \\
&= N\left[U^+\epsilon + U^-(1-\epsilon)\right]
\end{aligned}
$$

Since the expected payoff an intruder gets from not attacking each protocol run is αU, in order to discourage the intruder from attacking a protocol in multiple runs, we must have:

$$
\begin{aligned}
\alpha U N &> P \\
\alpha &> \frac{\epsilon U^+ + (1-\epsilon)U^-}{U}
\end{aligned}
$$

□

5 Case study

We consider a scenario where a party A who borrows money from a lender wants to discourage the lender from breaking into A's account that has a large amount of money in it. For clarity we outline the scenario below.

- Party A has borrowed 100 thousand US dollars from a lending company, who might be the mafia or an intruder in disguise.
- A however delays making the payment because of further investment, and this goes against the interest of the lender who does not trust A.
- To make a loan, the lender must have seen A's financial proofs such as bank statements, and therefore knows that A potentially has a bank account of up to 300 thousand US dollars. With this information, the lender will be extremely tempted to break into A's account and steal all of A's money because the potential reward of a successful attack exceeds the cost of many unsuccessful attempts as explained below.

Whenever A carries out an online transaction, he authenticates himself to the bank by typing in his 20-bit password (or a 5-hexadecimal-digit number) on the bank's website, and hence $\epsilon = 2^{-20}$ or $n = 2^{20}$. Let us suppose that the cost of launching an attack is 0.1 US dollar[4] then upon a successful attack on the authentication stage the payoff for the lender is $U^+ =$ (300,000 - 0.1) US dollars. If the lender fails, he gets a negative payoff $U^- = -0.1$ US dollar due to the running cost.

We therefore have this inequality $\epsilon U^+ + (1-\epsilon)U^- > 0$, which implies that unless our protocol transformation is used the expected payoff for the lender when he launches an attack is greater than when he does not attack. To discourage the lender from breaking into A's account, A will pursue the following strategy or irrational behaviour. Each time A uses his password to authenticate himself to the bank, with probability $\alpha \in [0, 1)$ the borrower will make a small payment of 1 US dollar to the lender. The payoff for the lender in a fake protocol run is therefore $U = 1$ US dollar. In this scenario, not only does A want to keep his account secure but also want to have the freedom to carry out online transactions with other parties without being disrupted by the lender.

In the first graph of Figure 1, we use results from Theorem 2 to plot α against $k \in \{1, \cdots, 2000\}$ where k is the maximum number of attacks the lender can launch. It is very clear that as k increases so does α but very slowly, i.e. α is around 0.186 for any $k \in \{1, \cdots, 2000\}$ corresponding to giving away 0.186 US dollar per transaction. We argue that this can have a significant impact on many banking applications which usually set $k = 3$ and so can be inconvenient to use especially after one comes back from a holiday and there are too many passwords to remember. What this shows is that there is not much economic benefit for the intruder and hence danger to the users when k is significantly greater than three.

[4] According to this website `http://www.csgnetwork.com/elecenergycalcs.html`, the cost of running a home computer system for one hour is 0.08 US dollar.

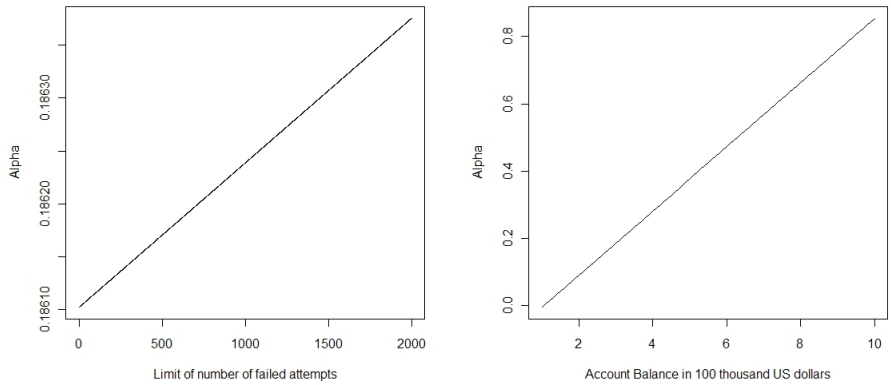

Fig. 1. Experimental results.

Since the account balance U^+ can vary, in the second graph of Figure 1 we use results from Theorem 2 to plot α agaist U^+ ranging from 100 to 1000 thousand US dollars, and k is set at 1000. It is clear that as U^+ increases so does α. To keep α small, one can increase the password bitlength to reduce the likelihood of a successful attack, but longer passwords are harder to remember and require more human interactions. Also the password length is usually fixed for each type of bank account regardless of how much money a customer puts in, and hence in practice it is not simple to change the password length to satisfy different customers' wishes. What this shows is that it is wise to leave only a large enough amount of money in each account to keep α low as in our case study.

6 Conclusions and future research

We have used ideas from game theory to transform two families of authentication protocols, namely password-based authentication and manual authentication protocols, to make them resilient against a powerful and rational intruder. In these protocols, only the intruder and dishonest parties are self-interested and all other trustworthy protocol participants should cooperate to complete a protocol run successfully, since it is in their mutual interest to agree on the same data.

At the heart of our protocol transformation is the introduction of some irrational behaviour protocol participants can pursue to discourage the intruder. Although the irrational behaviours might appear strange,

we argue that similar strategies have been exploited in other businesses including insurance and lottery. What happens with insurance is that customers pay a fixed amount of money regularly to an insurance company, because they will be compensated by the insurance company upon some bad events happening in their lives such as accidents. However, in many cases the compensation is less than the cumulative amount of money clients have paid to the insurance company, but what persuade them to pay insurance is the psychological fear of bad things happening which can be real but as unlikely as the chance of a successful attack.

Although we only consider pairwise schemes, our protocol transformation can be easily adapted to group protocols as presented in Annex A. For group authentication, the same protocol transformation of Table 1 can be used, however the initiator must be trustworthy. For otherwise a compromised node who does not obey our protocol transformation could play the role of the initiator. This is not an issue in a pairwise scenario as in our case study where two parties share a password or an authentic channel ($\Longrightarrow$). To remove the need of the trusted initiator in group authentication, in Annex A we extend our protocol transformation so that every node is assumed to be unfaithful with probability α independently. In particular, the *Macchiavellian* adversary model of Syverson et al. [12] will be interestingly employed to resist a rational intruder in password-based group key-agreement protocols, where the dishonnest (or Macchiavellian) parties are not willing to share passwords with the intruder.

While we have explored the notion of rational and powerful intruder in two types of authentication protocols, our work reported here opens the way to a number of new problems. For example, it would be interesting to investigate how relevant the notion of a rational intruder is to other types of authentication protocols which are based on PKIs, long private keys or a rapid-bit exchange as in distance-bounding protocols. Can one transform other families of cryptographic protocols to make them resilient against this kind of intruder? Also our studies on password-based protocols hopefully would lead to further attempts in improving the usability and economic security of many banking applications based on passwords which are currently quite inconvenient to use under many circumstances.

References

1. S.M. Bellovin and M. Merritt. *Encrypted Key Exchange: Password-Based Protocols Secure Against Dictionary Attacks*. Proceedings of the IEEE Symposium on Research in Security and Privacy (Oakland): 72.

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. M. Čagalj, S. Čapkun and J. Hubaux. *Key agreement in peer-to-peer wireless networks.* IEEE Special Issue on Security and Cryptography, **94**(2), 467-478, 2006.
4. C. Dimitrakakis, A. Mitrokotsa and S. Vaudenay. *Expected loss analysis of thresholded authentication protocols in noisy conditions.* See `http://arxiv.org/pdf/1009.0278`
5. G. Fuchsbauer, J. Katz and D. Naccache. *Efficient Rational Secret Sharing in Standard Communication Networks.* TCC 2010: 419-436
6. C. Gehrmann, C. Mitchell and K. Nyberg. *Manual Authentication for Wireless Devices.* RSA Cryptobytes, vol. 7, no. 1, pp. 29-37, 2004.
7. S.D. Gordon and J. Katz. *Rational secret sharing, revisited.* In Proceedings of Security and Cryptography for Networks. LNCS vol. 4116, 229-241, 2006.
8. 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.
9. Sven Laur and Kaisa Nyberg. Efficient Mutual Data Authentication Using Manually Authenticated Strings. Volume 4301 on *LNSC*, pages 90-107, Dec 2006. Springer.
10. D.R. Stinson. *Universal Hashing and Authentication Codes.* Advances in Cryptology - Crypto 1991, LNCS vol. 576, pp. 74-85, 1992.
11. P. Syverson. *Weakly secret bit commitment: Applications to lotteries and fair exchange.* The IEEE Computer Security Foundations Workshop, pages 2-13, 1998.
12. P. Syverson, C. Meadows, I. Cervesato. *Dolev-Yao is no better than Machiavelli.* First Workshop on Issues in the Theory of Security 2000.
13. J. Valkonen, N. Asokan and K. Nyberg, Ad-hoc security associations for groups, in: *Proceedings of the 3rd European Workshop on Security and Privacy in Ad-hoc and Sensor Networks*, Lecture Notes in Computer Science, Vol. 4357, L. Butty, V.D. Gligor and D. Westhoff, eds., Springer, 2006, pp. 150-164.
14. S. Vaudenay. *Secure Communications over Insecure Channels Based on Short Authenticated Strings.* Crypto 2005, LNCS vol. 3621, pp. 309-326.

A Extended protocol transformation for group authentication protocols

Throughout the main text of this paper, we have focused on the use of the protocol transformation of Table 1 on pairwise authentication protocols where there is a designated role for the protocol initiator who is unfaithful with probability α. Although the same protocol transformation is applicable to group authentication schemes, it might be difficult for multiple protocol participants to agree on who will be the initiator. We therefore would like to remove the need for such an initiator by assuming that every honest node can be unfaithful with probability α independently. In other words, we extend our protocol transformation and apply it to the behaviour of every group protocol participant, and hence

the *extended protocol transformation*. There is however a challenge arising from the assignment of correct payoffs for the intruder in different combinations of parties' strategies and protocol outcomes. We will first address the problem of payoff assignment before going into details of how the extended protocol transformation benefits both password-based and manual authentication protocols.

In a group authentication protocol, there are always two or more nodes in a group $\mathbf{G}$ who want to authenticate or agree on the same data. For the types of considered protocols, which only authenticate public data, not all of the protocol participants have to be honest, i.e. these compromised principles will not obey our protocol transformation. We will discuss further what the compromised nodes might do later. We denote p the number of honest parties out of all protocol participants and without loss of generality $p \geq 2$.

When the intruder does not attack a protocol run, there are three main possibilities that affect the payoff of the intruder:

- With probability $(1-\alpha)^p$, every honest node is faithful and there is no payoff for the intruder.
- With probability $1-(1-\alpha)^p-\alpha^p$, there are at least one faithful node and one unfaithful node. The payoff for the intruder might vary according to the number of faithful or unfaithful protocol participants, but as in pairwise schemes we only consider the worst-case scenario where the intruder's payoffs is always the same under this condition.
- With probability α^p, all nodes are unfaithful. When the irrational behaviour is equivalent to making a tiny payment to a third party who is the intruder in disguise, then the intruder's payoff increases as more nodes are unfaithful. However, if the intruder is only interested in incorrect authentication then unfaithful nodes will exchange or authenticate random data of no use in a fake protocol run. In the latter, the benefits the intruder gets cancel each other out when every node misbehaves because every one is aware of the fake run. That is why we need to deal with this case separately as it depends on the exact irrational behaviour.

When the intruder attacks a protocol run, there are two possibilities that affect the intruder's payoff:

- With probability $1-\alpha^p$, at least one node is faithful. The payoff for the intruder also can vary according to the number of faithful protocol participants, but again we only consider the worst-case scenario where

the intruder's payoff is the same when the intruder's attack succeeds, and also the same when it fails.

– With probability α^p, all nodes are unfaithful.

We summarise the payoffs for the intruder in different scenarios in Table 6.

Strategy of intruder	Strategy of honest parties	Outcome of protocol	Payoff of intruder	Payoff of participants
No attack	All faithful	Succeed	0	V^+
No attack	At least 1 faithful node & 1 unfaithful node	Succeed	U_1	V
No attack	All unfaithful	Succeed	U_2	V^-
Attack	At least 1 faithful node	Succeed Fail	U_1^+ U_1^-	0 0
Attack	All unfaithful	Succeed Fail	U_2^+ U_2^-	0 0

Table 6. A summary of the game.

Based on the damages an intruder might cause to honest parties, it is clear from Table 6 that we always have the followings:

$$U_1^+ > U_1 > U_1^-$$
$$U_2^+ \geq U_2^-$$

A.1 Password-based group authentication protocols

For clarity, we present the group version of the pairwise Diffie-Hellman-based Encrypted Key Exchange scheme of Bellovin and Merritt [1]. This protocol establishes a shared private key $g^{x_A x_B}$ between any two parties A and B, where g^{x_A} and g^{x_B} are public Diffie-Hellman keys of A and B, from a short password pw using an encryption scheme $E_{pw}()$ and a cryptographic hash function $hash()$.

Group password-based Encrypted Key Exchange Protocol
1. $\forall A \longrightarrow \forall B \ : \ A \parallel E_{pw}(g^{x_A})$
2. $\forall B \longrightarrow \forall A \ : \ hash(sk_{AB} \parallel 1)$
where $sk_{AB} = hash(A \parallel B \parallel g^{x_A} \parallel g^{x_B} \parallel g^{x_A x_B})$
3. $\forall A \longrightarrow \forall B \ : \ hash(sk_{AB} \parallel 2)$

In a password-based group protocol, such as the one above, all parties in group **G** share a common and private password pw. Following the *Machiavelli* adversary model introduced by Syverson et al. [12], we will allow the presence of compromised protocol participants. But, to make these protocols usable, those compromised principles are restricted to the following behaviours:

- In an attempt to collaborate with the intruder, compromised protocol participants will not obey our extended protocol transformation.
- However compromised principles will not share the password with the intruder or anyone else outside group **G**.
- Additionally compromised nodes will not use their knowledge of the shared password to fool other honest protocol participants into agreeing on the same and corrupt keys.

The latter two conditions must hold, for otherwise it is impossible to resist an intruder who possesses the password. Thus the intruder will not receive much support from compromised or Machiavellian parties.

No. of attempts	Outcome	Probability	Payoff of intruder
1	Succeed	$\epsilon = \epsilon_1 = 1/n$	$(1-\alpha^p)U_1^+ + \alpha^p U_2^+$
2	Succeed	$(1-\epsilon_1)\epsilon_2 = 1/n$	$(1-\alpha^p)\left(U_1^- + U_1^+\right) + \alpha^p\left(U_2^- + U_2^+\right)$
3	Succeed	$(1-\epsilon_1)(1-\epsilon_2)\epsilon_3 = 1/n$	$(1-\alpha^p)\left(2U_1^- + U_1^+\right) + \alpha^p\left(2U_2^- + U_2^+\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	Succeed	$\epsilon_t \Pi_{i=1}^{t-1}(1-\epsilon_i) = 1/n$	$(1-\alpha^p)\left((t-1)U_1^- + U_1^+\right) + \alpha^p\left((t-1)U_2^- + U_2^+\right)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	Succeed	$\epsilon_k \Pi_{i=1}^{k-1}(1-\epsilon_k) = 1/n$	$(1-\alpha^p)\left((k-1)U_1^- + U_1^+\right) + \alpha^p\left((k-1)U_2^- + U_2^+\right)$
k	Fail	$\Pi_{i=1}^{k}(1-\epsilon_i) = (n-k)/n$	$(1-\alpha^p)kU_1^- + \alpha^p kU_2^-$

Table 7. This tables shows the cumulative payoff and probability of the intruder's success and failure when (s)he attacks a group password-based protocol of p parties up to k runs.

Using information from Table 7, which summarises the intruder's cumulative payoff and probability of success and failure, we arrive at the following theorem.

Theorem 4. Suppose that an intruder is allowed to attack a group password-based protocol up to k runs for any $k \geq 1$. There are p honest protocol participants, and any of whom is unfaithful with probability α independently. If the following inequality holds then the intruder will not have any incentive to attack the protocol.

$$1 - \frac{(1-\alpha)^p + \alpha^p \delta}{1-\alpha^p} > \frac{\epsilon U_1^+ + (1-\epsilon) U_1^-}{U_1} + \left(\frac{U_1^+ - U_1^-}{U_1} \right) \Delta$$

where $\Delta = \frac{k-1}{n(2n-k+1)}$ and $\delta = \frac{U_2^+ - U_2}{U_1}$.

Interpretation of the result: This resembles Theorem 2 that concerns a pairwise password-based protocol of Section 3, especially the right hand side of the above inequality. Regarding the left hand side: as α increases toward 1, then so does $1 - \frac{(1-\alpha)^p}{1-\alpha^p}$, which therefore makes sense in our security model.

Many details in the following proof are not given since they can be found in the proof of Theorem 2.

Proof. When an intruder attacks a protocol up to k runs, from Table 7, the expected (average) number of protocol runs the intruder intervenes is

$$N = \frac{k(2n-k+1)}{2n}$$

Similarly, the expected cumulative payoff of the intruder's multiple-run attack can be computed as follows

$$\begin{aligned} P &= (1-\alpha^p) \left[\frac{kU_1^+}{n} + \frac{k(2n-k-1)U_1^-}{2n} \right] + \alpha^p \left[\frac{kU_2^+}{n} + \frac{k(2n-k-1)U_2^-}{2n} \right] \\ &\leq (1-\alpha^p) \left[\frac{kU_1^+}{n} + \frac{k(2n-k-1)U_1^-}{2n} \right] + \alpha^p U_2^+ N \qquad (2) \end{aligned}$$

Inequality (2) holds because $U_2^+ \geq U_2^-$. Since the expected payoff an intruder gets from not attacking a protocol in each run is

$$[1 - \alpha^p - (1-\alpha)^p] U_1 + \alpha^p U_2$$

To discourage the intruder from attacking a password-based protocol up to k runs, we want to have:

$$[1 - \alpha^p - (1-\alpha)^p] U_1 N + \alpha^p U_2 N > P$$

Using Inequality (2), we arrive at

$$[1-\alpha^p-(1-\alpha)^p]\,U_1N+\alpha^pU_2N>(1-\alpha^p)\left[\frac{kU_1^+}{n}+\frac{k(2n-k-1)U_1^-}{2n}\right]+\alpha^pU_2^+N$$

Dividing both sides by $(1-\alpha^p)U_1N$, using the calculation from the proof of Theorem 2 as well as rearranging, we have

$$1-\frac{(1-\alpha)^p}{1-\alpha^p}-\frac{\alpha^p}{1-\alpha^p}\frac{U_2^+-U_2}{U_1}>\frac{kU_1^+}{nU_1N}+\frac{k(2n-k-1)U_1^-}{2nU_1N}$$
$$1-\frac{(1-\alpha)^p+\delta\alpha^p}{1-\alpha^p}>\frac{\epsilon U_1^++(1-\epsilon)U_1^-}{U_1}+\left(\frac{U_1^+-U_1^-}{U_1}\right)\Delta$$

where $\Delta=\frac{k-1}{n(2n-k+1)}$ and $\delta=\frac{U_2^+-U_2}{U_1}$. □

A.2 Group manual authentication protocols

For clarity, we give the specification of a group manual authentication protocol in this section, though our analysis is applicable to other group manual authentication schemes [13]. In this scheme, all parties As of group $\mathbf{G}$ want to authenticate their public data m_A's from human interactions to remove the need of passwords, private keys and PKIs. The single arrow ($\longrightarrow$) indicates an unreliable and high-bandwidth link (e.g. WiFi or the Internet) where messages can be maliciously altered, whereas the double arrow ($\Longrightarrow$) represents an authentic and unspoofable channel. The latter is not a private channel (i.e. anyone can overhear it) and it is usually very low-bandwidth since it is implemented by humans, e.g., human conversations, text messages or manual data transfers between devices. $hash()$ and $uhash()$ are cryptographic and universal hash functions. Long random key k_A is generated by $A\in\mathbf{G}$, and k_A must be kept secret until after A has received Messages 1 from all other party $B\in\mathbf{G}$.

Manual group authentication protocol
1. $\forall A\longrightarrow\forall B: m_A, hash(A,k_A)$
2. $\forall A\longrightarrow\forall B: k_A$
3. $\forall A\Longrightarrow\forall B: uhash(k^*,M)$
where k^* is the XOR of all k_A's for $A\in\mathbf{G}$
and M is the concatenation of all m_A's for $A\in\mathbf{G}$

Using information from Table 8, which summarises the intruder's cumulative payoff and probability of success and failure, we arrive at the following theorem.

No. of attempts	Outcome	Probability	Payoff of intruder
1	Succeed	ϵ	$(1-\alpha^p)U_1^+ + \alpha^p U_2^+$
2	Succeed	$\epsilon(1-\epsilon)$	$(1-\alpha^p)(U_1^- + U_1^+) + \alpha^p(U_2^- + U_2^+)$
3	Succeed	$\epsilon(1-\epsilon)^2$	$(1-\alpha^p)(2U_1^- + U_1^+) + \alpha^p(2U_2^- + U_2^+)$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
t	Succeed	$\epsilon(1-\epsilon)^{t-1}$	$(1-\alpha^p)[(t-1)U_1^- + U_1^+] + \alpha^p[(t-1)U_2^- + U_2^+]$
$\vdots$	$\vdots$	$\vdots$	$\vdots$
k	Succeed	$\epsilon(1-\epsilon)^{k-1}$	$(1-\alpha^p)[(k-1)U_1^- + U_1^+] + \alpha^p[(k-1)U_2^- + U_2^+]$
k	Fail	$(1-\epsilon)^k$	$(1-\alpha^p)kU_1^- + \alpha^p kU_2^-$

Table 8. This tables shows the cumulative payoff and probability of the intruder's success and failure when (s)he attacks a group manual authentication protocol of p honest parties up to k runs.

Theorem 5. Suppose that an intruder is allowed to attack a group manual authentication protocol up to k runs for any $k \geq 1$. Moreover there are p honest protocol participants, and any of whom is unfaithful with probability α independently. If the following inequality holds then the intruder will not have any incentive to attack the protocol.

$$1 - \frac{(1-\alpha)^p + \delta\alpha^p}{1-\alpha^p} > \frac{\epsilon U_1^+ + (1-\epsilon)U_1^-}{U_1}$$

where $\delta = \frac{U_2^+ - U_2}{U_1}$.

Interpretation of the result: This resembles Theorem 3 which concerns a pairwise manual authentication protocol of Section 4, especially the right hand side of the above inequality. Secondly the value for α is also independent of the number of attempts k the intruder is allowed to attack a protocol. This is true intuitively because the chance of a successful attack on each protocol run of this type remains unchanged regardless of how many times an attack is launched as explained in Section 4.

Many details in the following proof are not given since they can be found in the proof of Theorem 3.

Proof. When an intruder attacks a protocol up to k runs, from Table 8, the expected number of runs the intruder intervenes in this protocol is:

$$\begin{aligned} N &= \epsilon + 2\epsilon(1-\epsilon) + \cdots + k\epsilon(1-\epsilon)^{k-1} + k(1-\epsilon)^k \\ &= 1 + (1-\epsilon) + (1-\epsilon)^2 + \cdots + (1-\epsilon)^{k-1} \end{aligned}$$

The expected cumulative payoff for the intruder's multiple-run attack can be computed as follows

$$\begin{aligned} P &= (1-\alpha^p)N\left[U_1^+\epsilon + U_1^-(1-\epsilon)\right] + \alpha^p N\left[U_2^+\epsilon + U_2^-(1-\epsilon)\right] \\ &\leq (1-\alpha^p)N\left[U_1^+\epsilon + U_1^-(1-\epsilon)\right] + \alpha^p N U_2^+ \end{aligned} \tag{3}$$

Inequality (3) holds because $U_2^+ \geq U_2^-$. Since the expected payoff an intruder gets from following this protocol in each run is

$$(1-\alpha^p-(1-\alpha)^p)\,U_1 + \alpha^p U_2$$

In order to discourage the intruder from attacking a protocol in multiple runs, we want to have:

$$[1-\alpha^p-(1-\alpha)^p]\,U_1 N + \alpha^p U_2 N > P$$

Using Inequality (3) and the calculation from the proof of Theorem 3 as well as dividing both sides by $(1-\alpha^p)NU_1$, we arrive at

$$\begin{aligned} \frac{1-\alpha^p-(1-\alpha)^p}{1-\alpha^p} - \frac{\alpha^p}{1-\alpha^p}\frac{U_2^+-U_2}{U_1} &> \frac{\epsilon U_1^+ + (1-\epsilon)U_1^-}{U_1} \\ 1 - \frac{(1-\alpha)^p + \delta\alpha^p}{1-\alpha^p} &> \frac{\epsilon U_1^+ + (1-\epsilon)U_1^-}{U_1} \end{aligned}$$

□