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 so that even if an intruder attacks a protocol, its payoff will still be lower than when it does not. This is particularly useful in resisting or discouraging a powerful and *rational* intruder (as present in military applications) who makes many attempts to break a protocol because (1) even if the intruder fails, a denial of service attack is still mounted successfully, 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.

1 Introduction

Ideas from game theory have been used to re-design a number of fair exchange protocols [4, 23] and secret sharing schemes [10, 8, 12, 7, 11] so that parties cannot act on their own interests to bring these schemes to failure [11]. 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 incorporate 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 military and battle fields, 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. Consequently the intruder does no worse (and potentially does better) by intervening and attempting to break many protocol runs because:

- With some probability ϵ : his or her attack on each protocol run succeeds, and this means that the intruder successfully manages to fool trustworthy parties into, for example, believing corrupt data.
- With probability 1ϵ : his or her attack fails, but then at least the intruder has successfully mounted a denial of service attack, and thus has prevented 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.

- In many cases, e.g. password-based protocols, we can put a limit k on the number of failed attempts an intruder can make. Under such a circumstance, a rational intruder only quits or stops attacking the protocol when (1) it succeeds in the t^{th} attempt where $t \leq k$, or (2) it fails in all k attempts.

These features motivate us to use techniques in game theory to redesign authentication protocols to resist this kind of rational intruder.

Our first contribution 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 even a dishonest node or the intruder deviates from a run its payoff will still be lower than in an equilibrium where everyone faithfully follows the protocol. In other words, the intruder does not have any incentive to attack a protocol, which is very similar to the concept of Nash equilibrium in game theory. The main thrust of this paper is to demonstrate how this protocol transformation works and benefits two families of protocols, namely password-based authentication (or key agreement) schemes [1-3] and manual authentication protocols [5, 13-17, 25], though other cryptographic protocols such as distance-bounding schemes [6, 20] also benefit from our work.

In Section 2, we present our protocol transformation and use it to protect pairwise authentication protocols against an intruder who can attack up to a single protocol run. This analysis will be formally extended to deal with multiple-run attacks on both password-based authentication schemes of Section 3 and manual authentication protocols of Section 4. In Annex A, we show how this strategy can be easily adapted to group protocols. The *Machiavelli* adversary model of Syverson et al. [23] will become particularly useful in our analysis of group protocols based on passwords where compromised nodes do not share secret with the intruder.

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 [10] and Fuchsbauer et al. [8] 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. Both of these require extra protocol runs, as in the case in our protocol transformation of Table 1.

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*. The job of a rational intruder is to break a protocol run and maximise his or her payoff. 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 [1–3] or human interactions [15–17, 25, 26]. These changes, which are summarised in Tables 1 and 2, aim to discourage the intruder from attacking a protocol.

From Table 1, although $U_1^+ \ge U_2^+ > U_1^- \ge U_2^-$, these payoffs for the intruder when it attacks a protocol can vary widely relative to one another in practice, e.g. U_1^+ and U_2^+ can be either far apart or roughly the same. We therefore will only tackle the most general case here: regardless of whether 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 Table 2. Consequently, a solution for this most general case 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. Suppose that 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:

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

Protocol transformation

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

- With probability α : the initiator A will authenticate or transmit some useless and random data, even though A still follows every step of a protocol faithfully. When the intruder does not misbehave then a protocol always completes successfully. This means that the other party B, who honestly follows the protocol, will not be aware that the authenticated data are of no use, and so it is a waste of time for B.

When this happens, after a period of time A will initiate another run with B to revoke all useless data agreed in the previous run and to authenticate meaningful data. Since it is in the intruder's interest that A deliberately authenticates useless data, we denote Uthe intruder's payoff in this case. There is no 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 be successful, 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 1ϵ the intruder fails but it will still be given some payoff for stopping A and B agreeing on the same data: $U_1^$ when A is faithful, and U_2^- when A is unfaithful.

In a successful attack and when the initiator A is unfaithful, the intruder can still cause real damage to B before A initiates another runs to revoke the corrupt data, we therefore arrive at $U_1^+ \ge U_2^+ > U$. Also the intruder would prefer B to have a fake protocol run with an unfaithful initiator rather than an unsuccessful attack, and hence $U > U_1^- \ge U_2^-$.

Table 1. Protocol transformation.

Strategy	Strategy	Outcome	Payoff of	Payoff of
of intruder	of initiator A	of protocol	intruder	participants
No attack	Faithful	Succeed	0	V
No attack	Unfaithful	Succeed	U	0
Attack	Faithful	Succeed	U_1^+	0
Attack	Unfaithful	Succeed	U_2^+	0
Attack	Faithful	Fail	U_1^-	0
Attack	Unfaithful	Fail	U_2^-	0
1				

The lower half of this table is the generalised version of the upper half

No attack	Faithful	Succeed	0	V
No attack	Unfaithful	Succeed	U	0
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.

Proof. If the intruder does not misbehave, his or her expected payoff in each run is αU . If the intruder misbehaves, his or her 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:

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

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

So as long as this is true, it is in the intruder's interest not to intervene and manipulate data transmitted in any protocol run. $\hfill\square$

From Tables 1 and 2, we have $U^+ > U > U^-$. But no matter how big U^+ is, $\epsilon(U^+ - U^-)$ can be made extremely small relative to U and U^- , e.g. increasing the length of password or universal hash function in protocols of Sections 3 and 4 will exponentially decrease the value of the successful probability ϵ . However, the value of α is lower bounded by U^-/U . We did not specify the exact range for $U^- = max\{U_1^-, U_2^-\}$ in Tables 1 and 2 because when the intruder benefits from a denial of service attack (and/or further knowledge of a password) then $U^- > 0$. In contrast, U^- would be insignificant or even negative if we also consider honest parties' suspicion that an intruder is active after several failed protocol runs.

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

$$0 < \alpha_1 < \alpha_2 < \dots < \alpha_w < 1$$

Since the main priority is discourage the intruder from attacking a protocol run, according to Theorem 1 node A will only select strategies with $\alpha \in \{\alpha_i, \alpha_{i+1}, \dots, \alpha_w\}$ 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 collective payoff of the protocol participants in each run which is derived from Tables 1 and 2 to be

$$(1-\alpha)V$$

Since $\alpha_i < \alpha_{i+1} < \cdots < \alpha_w < 1$, it is clear from the payoff matrix or Table 3 that node A will pick α_i as his or her optimal strategy.¹

Strategy of	Strategy of the initiator A			
Intruder	α_i	α_{i+1}	•••	$lpha_w$
No attack	$\alpha_i U, (1-\alpha_i)V$	$\alpha_{i+1}U, (1-\alpha_{i+1})V$	•••	$\alpha_w U, (1 - \alpha_w) V$
Attack	P, 0	P, 0	•••	P, 0

Table 3. This matrix shows relative payoffs for the intruder and the protocol participants with each combination of their strategies in each protocol run. Please note that $P = \epsilon U^+ + (1 - \epsilon)U^-$ is the payoff for the intruder when (s)he attacks a protocol.

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

¹ 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.

of Bellovin and Merritt [1, 2]. 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.

Password-base	ed Encrypt	ted Kev	Exchange	Protocol	[1.	2
					- 7	

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 $\{1, \ldots, n\}$, then² $1 \le k \le n$ 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 protocols run is correct even when the initiator A is unfaithful, because except authenticating meaningless data an unfaithful initiator by definition from the protocol transformation of Table 1 still follows every other protocol step properly, including authentication checks that involve the use of the password.

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

– The intruder succeeds in the $t^{\mbox{th}}$ attempt where $t \leq k$ or

- 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, i.e. the initiator is untrustworthy, and the intruder still needs to guess the password correctly.

 $^{^{2}}$ This is true when the password is unchanged throughout a multiple-run attack.

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 accumulative payoff and probability that it is successful or unsuccessful up to k attempts in Table 4.

No. of	Outcome	Probability	Payoff
attempts			of intruder
1	Succeed	$\epsilon = \epsilon_1 = 1/n$	U^+
2	Succeed	$(1-\epsilon_1)\epsilon_2 = 1/n$	$U^{-} + U^{+}$
3	Succeed	$(1-\epsilon_1)(1-\epsilon_2)\epsilon_3 = 1/n$	$2U^{-} + U^{+}$
:	•		
t	Succeed	$\epsilon_t \Pi_{i=1}^{t-1} (1 - \epsilon_i) = 1/n$	$(t-1)U^{-} + U^{+}$
:	:		÷
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 accumulative 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 passwordbased 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 t^{th} attempt where $t \leq k$ or fails in all k attempts as seen in Table 4. Then to discourage the intruder from 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

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

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

$$P = \frac{U^{+}}{n} + \frac{U^{-} + U^{+}}{n} + \dots + \frac{(k-1)U^{-} + U^{+}}{n} + \frac{k(n-k)U^{-}}{n}$$
$$= \frac{kU^{+}}{n} + U^{-} \left[\frac{1}{n} + \frac{2}{n} + \dots + \frac{k-1}{n} + \frac{k(n-k)}{n}\right]$$
$$= \frac{kU^{+}}{n} + \frac{k(2n-k-1)U^{-}}{2n}$$

Since the 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:

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

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

Since $n \ge k \ge 1$, as k increases then so do both Δ and α , and moreover $\epsilon > \Delta \ge 0$. 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 made arbitrarily small by exponentially decreasing the value of ϵ , i.e. increasing the password length.
- If this protocol transformation can discourage a k-run attack, then it can also discourage a t-run attack for any $t \leq k$.

As regards the payoff of protocol participants, we can use the analysis given at the end of Section 2 to select the optimal strategy for A to maximise the collective payoff when α can be chosen from a range of different values, i.e. $\alpha \in \{\alpha_1, \dots, \alpha_w\}$ where $0 < \alpha_1 < \dots < \alpha_w < 1$.

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 [17].

Our analysis here applies to every secure manual authentication protocol, but for clarity we give the pairwise version of the SHCBK protocol of the author [16–18]. 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 \parallel and \oplus denote concatenation and exclusive-or.

Α	pairwise manual authentication protocol [16–18]
1.	$A \longrightarrow B : m_A, hash(k_A)$
2.	$B \longrightarrow A : m_B, k_B$
3.	$A \longrightarrow B : k_A$
4.	$A \Longleftrightarrow B: uhash(k_A \oplus k_B, m_A \parallel m_B)$

To ensure both parties share the same data, the human owners of devices A and B have to compare a short universal hash value 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. [22] An ϵ -almost universal hash function, $uhash : R \times X \to Y$, must satisfy that for every $m, m' \in X$ and $m \neq m'$:

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

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 [25] and Čagalj et al. [5], which do not use a universal hash function.

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 authenticates useless data, though A still follows every other 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 $k \ge 1$, and the intruder quits iff (s)he is successful in the t^{th} attempt where $t \le k$ 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 accumulative payoff and probability of success and failure in Table 5.

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

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

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:

$$N = \epsilon + 2\epsilon(1-\epsilon) + \dots + k\epsilon(1-\epsilon)^{k-1} + k(1-\epsilon)^k$$

= 1 + (1-\epsilon) + (1-\epsilon)^2 + \dots + (1-\epsilon)^{k-1} (1)

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

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

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

$$P = \epsilon U^{+} + \epsilon (1 - \epsilon) (U^{-} + U^{+}) + \dots + \epsilon (1 - \epsilon)^{k-1} ((k-1)U^{-} + U^{+}) + (1 - \epsilon)^{k} k U^{-}$$

= $U^{+} \epsilon \left[1 + (1 - \epsilon) + \dots + (1 - \epsilon)^{k-1} \right] + U^{-} (1 - \epsilon) \left[\epsilon + 2\epsilon (1 - \epsilon) + \dots + (k-1)\epsilon (1 - \epsilon)^{k-2} + k(1 - \epsilon)^{k-1} \right]$
= $U^{+} \epsilon N + U^{-} (1 - \epsilon) N$
= $N \left[U^{+} \epsilon + U^{-} (1 - \epsilon) \right]$

Since the payoff an intruder gets from following this protocol in each run is αU , in order to discourage the intruder from attacking a protocol in multiple runs, we must have:

$$\alpha UN > P$$

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

5 Payoff versus loss maximisation

The notion of maximisation of an attacker's payoff is very much related to maximising the loss of honest players. Recently there have been some work, notably [6] of Dimitrakakis et al., which analyse expected loss in authentication protocols based on a challenge-response exchange lasting multiple rounds between a verifier and a prover. The verifier's loss comes from the possibility of false acceptance (or authenticating a malicious intruder) and false rejection (or rejecting a legitimate user), both of which arise from noise existing in the data layer which exchanges bits during the challenge-response phase of the protocols. The noise therefore necessitates the use of a tolerance threshold, such that a prover is authenticated if the total number of errors of its responses is below the threshold.

We observe that intuitively the same notion of a rational attacker investigated in this paper is applicable in this scenario where the attacker tries to maximise the verifier's expected loss. To discourage such an attacker from disrupting a protocol, we need to adjust the threshold so that even when an attacker impersonates the prover (e.g. 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. It is worth to point out that the noise present in the challenge-response phase causes damage to the verifier, and hence there is no need to introduce irrational behaviour into honest parties' activities as in our protocol transformation of Table 1 to discourage a rational attacker.

The author has made progress regarding this problem as reported in [20], but due to limited space and since this problem is not quite within the scope of this paper, we intend to publish this work in due course.

6 Conclusions and future research

We have introduced the use of ideas from game theory to redesign 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 incorporate to complete a protocol run successfully, since this is in their mutual interest to agree on the same data.

Although we only consider pairwise authentication schemes, our protocol transformation can be easily adapted to group protocols where there are more than two nodes 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 so that all other honest nodes would always authenticate useful data. This is not an issue in a pairwise scenario where two parties share a private key/password or an authentic channel (\Longrightarrow), because it makes no sense for either of them to fool the other. 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. [24] will be interestingly employed to resist a rational intruder in password-based group key-agreement protocols.

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. The first set of questions consists of direct extension of the results presented here. For example, in addition to authentication schemes based on a challenge-response phase as mentioned previously, 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 or long private keys. The kind of rational intruder considered here might not be present in some applications, and hence can one formally define and model a weaker intruder so that the probability that honest nodes need to behave irrationally can be reduced further? Also there is no need to use the protocol transformation of Table 1 in every scenario, instead one can switch it on or off depending on the anticipated level of threat, risk or presence of a powerful intruder in each application.

The second set of questions is more open-ended. As a rational and powerful intruder exists in hostile environment, can one transform other families of cryptographic protocols to make them resilient against this kind of intruder? Since our protocol transformation works best when protocols are immune to (off-line) searching, can it be relaxed or modified to accommodate a wider variety of possible attacks, e.g. substitution attacks that are relevant to other cryptographic primitives (including MACs)? Besides considering reducing the intruder's power in future work, one might consider increasing rationality assumptions for the intruders, as in the *Macchiavellian* adversary [24] who does not share private keys and passwords with its collaborators.

References

- M. Bellare, D. Pointcheval, and P. Rogaway. Authenticated Key Exchange Secure against Dictionary Attacks. Eurocrypt 2000, LNCS 1807.
- 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.
- V. Boyko, P. MacKenzie, and S. Patel. Provably Secure Password-Authenticated Key Exchange Using Diffie-Hellman. Eurocrypt 2000, LNCS 1807.
- 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.
- 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.
- C. Dimitrakakis, A. Mitrokotsa and S. Vaudenay. Expected loss analysis of thresholded authentication protocols in noisy conditions. See http://arxiv.org/pdf/1009.0278
- 7. Y. Dodis and T. Rabin. Cryptography and Game Theory.
- G. Fuchsbauer, J. Katz and D. Naccache. Efficient Rational Secret Sharing in Standard Communication Networks. TCC 2010: 419-436

- C. Gehrmann, C. Mitchell and K. Nyberg. Manual Authentication for Wireless Devices. RSA Cryptobytes, vol. 7, no. 1, pp. 29-37, 2004.
- S.D. Gordon and J. Katz. Rational secret sharing, revisited. In Proceedings of Security and Cryptography for Networks. LNCS vol. 4116, 229-241, 2006.
- 11. 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.
- 12. J. Katz. Bridging Game Theory and Cryptography: Recent Results and Future Directions. In the 5th Theory of Cryptography Conference (TCC) 2008.
- S. Laur and K. Nyberg. Efficient Mutual Data Authentication Using Manually Authenticated Strings. LNCS vol. 4301, pp. 90-107, 2006.
- 14. A.Y. Lindell. Comparison-based key exchange and the security of the numeric comparison mode in Bluetooth v2.1. CT-RSA, LNCS vol. 5473, 2009, pp. 66-83.
- L.H. Nguyen (editor), second edition. ISO/IEC 9798-6 (2010): Information Technology Security Techniques Entity authentication Part 6: Mechanisms using manual data transfer.
- L.H. Nguyen and A.W. Roscoe. Authenticating ad hoc networks by comparison of short digests. Information and Computation 206 (2008), 250-271.
- 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.
- 18. L.H. Nguyen and A.W. Roscoe. *Efficient group authentication protocol based on human interaction*. Proceedings of FCS-ARSPA, pp. 9-31, 2006.
- L.H. Nguyen and A.W. Roscoe. Separating two roles of hashing in one-way message authentication. Proceedings of FCS-ARSPA-WITS, 2008, pp. 195-210.
- 20. L.H. Nguyen. Rational distance-bounding protocols under noisy conditions. See: http://www.comlab.ox.ac.uk/files/3931/loss.pdf
- S. Pasini and S. Vaudenay. SAS-based Authenticated Key Agreement. Public Key Cryptography - PKC 2006. LNCS vol. 3958, pp. 395-409.
- D.R. Stinson. Universal Hashing and Authentication Codes. Advances in Cryptology - Crypto 1991, LNCS vol. 576, pp. 74-85, 1992.
- 23. P. Syverson. Weakly secret bit commitment: Applications to lotteries and fair exchange. The IEEE Computer Security Foundations Workshop, pages 2-13, 1998.
- 24. P. Syverson, C. Meadows, I. Cervesato. *Dolev-Yao is no better than Machiavelli*. First Workshop on Issues in the Theory of Security 2000.
- S. Vaudenay. Secure Communications over Insecure Channels Based on Short Authenticated Strings. Crypto 2005, LNCS vol. 3621, pp. 309-326.
- J. Valkonen, N. Asokan and K. Nyberg. Ad Hoc Security Associations for Groups. In Proceedings of the Third European Workshop on Security and Privacy in Ad hoc and Sensor Networks 2006. LNCS vol. 4357, pp. 150-164.

A Extended protocol transformation for group authentication protocols

Throughout the main text of this paper, we have focused on the application of the protocol transformation of Table 1 on pairwise authentication protocols, and in particular 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 (or authenticates useless data) 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 difficulty arising from the assignment of correct payoffs for the intruder in different combinations of parties' strategies, which will be addressed below.

In a group authentication protocol, there are more than two nodes in a group **G**. 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 nodes present, but we only consider the most general case where the intruder's payoffs is always the same under this condition.
- With probability α^p , all nodes are unfaithful.

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 nodes present, but again we only consider the most general case 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 payoff for the intruder in different scenarios in Table 6.

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

 $U_1 \ge U_2$

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

 Table 6. A summary of the game.

$$U_1^+ \ge U_2^+ \\ U_1^+ \ge U_1^- \\ U_1^- \ge U_2^-$$

What is not clear is how we can compare the payoffs for the intruder when all parties choose to be unfaithful, i.e. U_2 , U_2^+ and U_2^- . It turns out that it very much depends on the type of authentication protocols, and hence in the following subsections we will investigate both manual authentication and password-based authentication schemes in turn.

A.1 Group manual authentication protocols

For clarity, we give the specification of the SHCBK protocol in this Section, though our analysis is applicable to other group manual authentication schemes [17, 26]. In this scheme, all parties As of group **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 ran-

dom 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 $B \in \mathbf{G}$.

SHCBK protocol [16, 18]
$1. \forall A \longrightarrow \forall B : m_A, hash(A, k_A)$
$2. \forall A \longrightarrow \forall B : k_A$
$3. \forall A \Longrightarrow \forall B : uhash(k^*, M)$
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 $A \in \mathbf{G}$

Since there is no shared password or private key that underlies a manual authentication protocol, there is not much benefit for both the intruder and compromised principles when all honest protocol participants are unfaithful even when the intruder in collaboration with compromised nodes succeed in their attack. Consequently, the following inequality appears to be plausible.

$$U_2 \ge U_2^+ \ge U_2^-$$

Under this assumption and the information from Table 7, which summarises the intruder's accumulative payoff and probability of success and failure, we arrive at the following theorem.

No. of	Outcome	Probability	Payoff of
attempts			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	$\epsilon (1-\epsilon)^2$	$(1 - \alpha^p)(2U_1^- + U_1^+) + \alpha^p(2U_2^- + U_2^+)$
:	:	•	
t	Succeed	$\epsilon (1-\epsilon)^{t-1}$	$(1 - \alpha^p)[(t - 1)U_1^- + U_1^+] + \alpha^p[(t - 1)U_2^- + U_2^+]$
÷	:	•	÷
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 7. This tables shows the accumulative 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 4. Suppose that an intruder is allowed to attack a group manual authentication protocol up to k runs for any $k \ge 1$. Moreover there

are p honest protocol participants, and any of whom can be unfaithful with probability α independently. If the following inequality holds then we can discourage the intruder from attacking the protocol.

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

Many details in the following proof are not given since they can be found in the proof of Theorem 3 which concerns a pairwise manual authentication protocol of Section 4.

Proof. When an intruder attacks a protocol up to k runs, from Table 7, the expected number of runs the intruder intervenes in this protocol is:

$$N = \epsilon + 2\epsilon(1 - \epsilon) + \dots + k\epsilon(1 - \epsilon)^{k-1} + k(1 - \epsilon)^k$$

= 1 + (1 - \epsilon) + (1 - \epsilon)^2 + \dots + (1 - \epsilon)^{k-1}

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

$$P = (1 - \alpha^{p})N\left[U_{1}^{+}\epsilon + U_{1}^{-}(1 - \epsilon)\right] + \alpha^{p}N\left[U_{2}^{+}\epsilon + U_{2}^{-}(1 - \epsilon)\right]$$

$$\leq (1 - \alpha^{p})N\left[U_{1}^{+}\epsilon + U_{1}^{-}(1 - \epsilon)\right] + \alpha^{p}NU_{2}$$
(2)

Inequality (2) holds because we have assumed that $U_2 \ge U_2^+ \ge U_2^-$. Since the 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 (2) and dividing both sides by $(1 - \alpha^p)NU_1$, we arrive at

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

	_

We observe that as α increases toward 1, then so is $1 - \frac{(1-\alpha)^p}{1-\alpha^p}$, which therefore makes sense in our security model. Additionally, the value for α is independent of the number of attempts k the intruder is allowed to attack a protocol. As explained in Section 4, this is intuitively correct 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.

A.2 Password-based group authentication protocols

For clarity, we present the group version of the pairwise Diffie-Hellmanbased Encrypted Key Exchange scheme of Bellovin and Merritt [1,2]. 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 Protoco
$1. \forall A \longrightarrow \forall B : A \parallel E_{pw}(g^{x_A})$
$2. \forall B \longrightarrow \forall A : hash(sk_{AB} \parallel 1)$
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. [24], we will allow the presence of compromised protocol participants. But, to make these protocols usable, these 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, i.e. they always exchange meaningful data.
- 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. However as we have observed previously there is always further knowledge of the password the intruder can gain from attacking a protocol run, even when all honest protocol participants are unfaithful. Consequently, it is not possible to compare U_2 against U_2^- , and the only comparisons that appear to be plausible are the followings.

$$U_2^+ \ge U_2^-$$
$$U_2^+ \ge U_2$$

Under this assumption and information from Table 8, which summarises the intruder's accumulative payoff and probability of success and failure, we arrive at the following theorem.

No. of	Outcome	Probability	Payoff of
attempts			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	$(1-\epsilon_1)(1-\epsilon_2)\epsilon_3 = 1/n$	$\left (1 - \alpha^p) \left(2U_1^- + U_1^+ \right) + \alpha^p \left(2U_2^- + U_2^+ \right) \right $
:	:		
t	Succeed	$\epsilon_t \Pi_{i=1}^{t-1} (1-\epsilon_i) = 1/n$	$(1-\alpha^p)\left((t-1)U_1^-+U_1^+\right)+$
			$\alpha^p \left((t-1)U_2^- + U_2^+ \right)$
:	:		
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 8. This tables shows the accumulative 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.

Theorem 5. Suppose that an intruder is allowed to attack a group password-based protocol up to k runs for any $k \ge 1$. There are p honest protocol participants, and any of whom can be unfaithful with probability α independently. If the following inequality holds then we can discourage the intruder from attacking 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 $\delta = \frac{U_2^+ - U_2}{U_1}$.

Many details in the following proof are not given since they can be found in the proof of Theorem 2 which concerns a pairwise password-based protocol of Section 3.

Proof. When an intruder attacks a protocol up to k runs, from Table 8, the expected (average) number of protocol runs the intruder intervenes is

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

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

$$P = (1 - \alpha^p) \left[\frac{kU_1^+}{n} + \frac{k(2n - k - 1)U_1^-}{2n} \right] + \alpha^p \left[\frac{kU_2^+}{n} + \frac{k(2n - k - 1)U_2^-}{2n} \right]$$
$$\leq (1 - \alpha^p) \left[\frac{kU_1^+}{n} + \frac{k(2n - k - 1)U_1^-}{2n} \right] + \alpha^p U_2^+ N \tag{3}$$

Inequality (3) holds because $U_2^+ \ge U_2^-$. Since the 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 (3), we arrive at

$$\left[1 - \alpha^p - (1 - \alpha)^p\right]U_1N + \alpha^p U_2N > (1 - \alpha^p)\left[\frac{kU_1^+}{n} + \frac{k(2n - k - 1)U_1^-}{2n}\right] + \alpha^p U_2^+N$$

Dividing both sides by $(1 - \alpha^p)U_1N$ and rearranging, we have

$$\begin{split} 1 - \frac{(1-\alpha)^p}{1-\alpha^p} - \frac{\alpha^p}{1-\alpha^p} \frac{U_2^+ - U_2}{U_1} &> \frac{kU_1^+}{nU_1N} + \frac{k(2n-k-1)U_1^-}{2nU_1N} \\ 1 - \frac{(1-\alpha)^p + \delta\alpha^p}{1-\alpha^p} &> \frac{\epsilon U_1^+ + (1-\epsilon)U_1^-}{U_1} + \left(\frac{U_1^+ - U_1^-}{U_1}\right)\Delta \\ \end{split}$$
 where $\Delta = \frac{k-1}{n(2n-k+1)}$ and $\delta = \frac{U_2^+ - U_2}{U_1}$.