# Multichannel protocols to prevent relay attacks*

Frank Stajano[1], Ford-Long Wong[2] and Bruce Christianson[3]**

[1] University of Cambridge Computer Laboratory, Cambridge, United Kingdom
[2] DSO National Laboratories, Singapore
[3] University of Hertfordshire, School of Computer Science, Hatfield, United Kingdom.

**Abstract.** A number of security systems, from Chip-and-PIN payment cards to contactless subway and train tokens, as well as secure localization systems, are vulnerable to *relay attacks*.

Encrypting the communication between the honest endpoints does not protect against such attacks. The main solution that has been offered to date is distance bounding, in which a tightly timed exchange of challenges and responses persuades the verifier that the prover cannot be further away than a certain distance. This solution, however, still won't say whether the specific endpoint the verifier is talking to is the intended one or not—it will only tell the verifier whether the real prover is "nearby".

Are there any alternatives? We propose a more general paradigm based on multichannel protocols. Our class of protocols, of which distance bounding can be modelled as a special case, allows a precise answer to be given to the question of whether the unknown device in front of the potential victim is a relaying attacker or the device with which the victim intended to communicate.

We discuss several instantiations of our solution and point out the extent to which all these countermeasures rely, often implicitly, on the alertness of a honest human taking part in the protocol.

## 1 Introduction

In a relay attack, the victims are two honest parties acting respectively as a prover (e.g. a door-opening token) and a verifier (e.g. a door-mounted token reader). In normal operation, when the prover (token) is authenticated by the verifier (door), the verifier grants some privilege (the door opens).

During a relay attack[4], a pair of communicating attackers splice themselves in the communication channel between the two victims. One of the attackers acts as a fake verifier to the victim prover and the other acts as a fake prover to the victim verifier. When the victim verifier issues a challenge, the attackers relay it unchanged to the victim prover; and when the prover issues its response to the original challenge, the attackers relay that too, unchanged, to the true

* Revision 39 of 2010-02-27 22:23:18 +0100 (Sat, 27 Feb 2010).

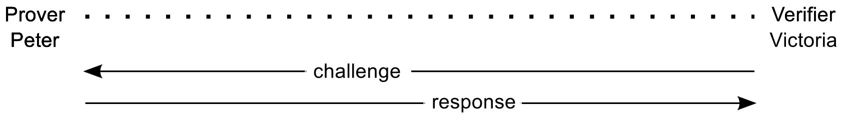
** On sabbatical at the University of Cambridge Computer Laboratory while the core of this research was carried out.

[4] Sometimes also called a *wormhole* attack, especially in secure localization contexts.

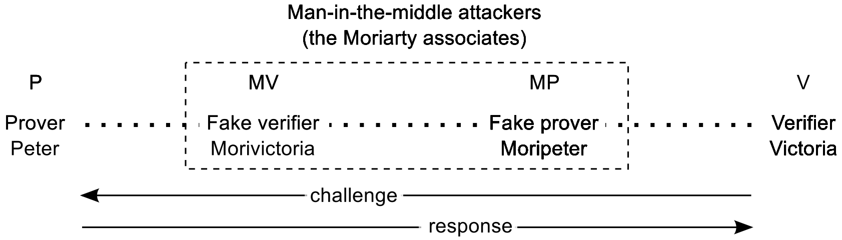
(Authors' final version as submitted to Springer)
In R. Sion (ed): *Proceedings of Financial Cryptography 2010*, LNCS 6052, pp. 4–19.
© IFCA/Springer-Verlag Berlin Heidelberg 2010.

verifier. The outcome is that the victim verifier grants the privilege to the fake prover, who was accepted thanks to the credentials unknowingly provided by the victim prover.

The honest participants[5]:



When a relay attack is taking place:



Even if the victim prover and verifier share a secret unknown to the attackers, they are still vulnerable: since their messages are relayed unchanged, the attackers succeed in fooling the verifier regardless of whether they can decrypt the messages they relay.

This problem has been known for several decades: Conway [6] described the "chess grandmaster problem", in which an unskilled player defeats (or at least draws with) a chess grandmaster by simultaneously challenging *two* grandmasters at postal chess, one as white and one as black, and countering the moves of one grandmaster with those of the other. Beth and Desmedt [2] revisited the problem, noting that it matched the scenarios of the "mafia fraud"[6] and "terrorist fraud"[7], both previously described by Desmedt et al. [8], and they introduced

[5] We only show the essential core of the protocol here: clearly, in a more realistic situation, one would expect the protocol to be initiated by a preliminary request from Peter "hey, please challenge me so I can prove I'm worthy of getting the benefits". We omit this and other non-essential messages for brevity and clarity.

[6] In the mafia fraud, $P$ is a customer who is electronically paying his restaurant bill to $MV$. Restaurant owner $MV$ is a member of a mafia gang who alerts his accomplice $MP$ to go and buy a diamond from jeweller $V$. Jeweller $V$ challenges $MP$ for his credentials, but $MP$ and $MV$ relay $P$'s credentials to $V$. So $P$ thinks he's paying for a meal, whereas he is buying the mafiosi a diamond.

[7] In the terrorist fraud, the verifier $V$ is an immigration officer of country $\alpha$ and the fake prover $MP$ is a terrorist who wants to enter the country. The fake prover $MP$ is helped by $P$, a sympathetic citizen of $\alpha$ who supplies the correct answers to the

the defensive technique of measuring the round-trip time, relying on the fact that the speed of light is finite to detect whether the actual prover is further away than expected. Brands and Chaum [3] refined that technique into a specific and more secure low-level protocol, with precomputation of single-bit challenges and responses that are then exchanged as quickly as the channel allows. More recently, Hancke and Kuhn [12] developed a distance-bounding protocol optimized for the resource-constrained RFID environment and, with colleagues [5], studied a variety of attacks on the timing measurements. Drimer and Murdoch [9] built electronic circuitry to demonstrate the relay attack[8] against modern Chip-and-PIN bank cards and implemented the Hancke-Kuhn protocol to demonstrate its viability as a practical countermeasure. Hancke's doctoral dissertation [11] contains a good survey of the distance-bounding protocols in the literature.

The purpose of any distance-bounding protocol in such a context is to convince the honest verifier that the honest prover she is ultimately interacting with (the one who *can* respond to the challenges, whereas the attackers can't because they don't know the shared secret) is, with high probability, the prover currently in front of her. By construction, the distance bounding protocol can only give a verdict of the form "the owner of the shared secret just proved that he is no further away than $d$ metres". If the verifier is interacting with a prover (whether genuine or fake) that is less than 1 metre away, but the distance bounding protocol says that he was unable to prove that he is within 10 metres, then the verifier should suspect that she is interacting with a relaying attacker.

Still, the distance-bounding solution does not really identify a specific principal but only its approximate location[9]. At least theoretically, depending on the spatial resolution of the distance-bounding protocol, it is still possible for attackers to go undetected if they stay within the bounds of the error margin, as in the scenario of multiple adjacent cash machines of which one is fake and performs a relay attack on another.

In this paper we propose a new paradigm for detecting and preventing relay attacks that is more general than distance bounding. Our strategy is to use a multichannel protocol [20,15,4,18,16] in which the traditional challenge-response between verifier and prover on the regular channel is augmented with an additional verification on a special channel whose main property is that it cannot be relayed.

Our multichannel approach includes the distance-bounding solution as a special case[10]. More importantly, our family of solutions includes ones that give a

questions of the immigration officer $V$. The main difference between this case and the mafia fraud is that the prover $P$ is not a victim of the scam but an accomplice: he cooperates with the fake prover $MP$ against the verifier $V$ and therefore there is no need for a fake verifier $MV$.

[8] With explicit reference to the "mafia fraud" scenario.

[9] Within a sphere, or within the intersection of several spheres in the substantially more complicated case where one repeats the protocol from several reference points.

[10] Insofar as you cannot relay beyond a certain distance the special channel implicitly defined by the distance-bounding procedure without being noticed by the victim endpoints.

clear and definite "yes / no" answer to the question "is the principal in front of me really the one with whom I share this secret key, or is it just a middleperson attacker?", which the distance-bounding protocols can only answer with a less stringent assurance such as "it probably is, provided there are no other principals within $d$ metres of Victoria".

Our approach also models the anti-relay alternative proposed by Damgård et al. [7] of somehow limiting the bandwidth with which the prover can communicate to the outside world to a value lower than the one needed in order to conduct the protocol—their arrangement implicitly relies on unrelayable channels because, by construction, at least one of the channels used in the protocol cannot be relayed to third parties outside.

We also highlight the extent to which all these anti-relay protocols, including both our new ones and the traditional ones based on distance bounding, implicitly rely on the presence of an honest human. We discuss whether they are still secure when the human takes part in the protocol without actively cheating but without thoroughly investigating all possible suspicious clues.

## 2 The core idea

Our core idea is that, although the man-in-the-middle attackers are usually able to relay the information between the two honest endpoints over whatever channels are normally used for the transaction, we might be able to augment the system with an additional special channel that the attackers won't be able to relay. Over *that* channel, the two endpoints can verify whether they are talking directly to each other or not.

Traditionally, the authentication problem[11] can be framed in the following terms: "I know I am talking to you; now, prove to me that you know our shared secret". Here, instead, we examine the dual problem: "I know I am talking to someone who knows my shared secret; now, prove to me that you, the principal in front of me, are that someone".

The intuition behind the multichannel approach is that the verifier asking that question should use the special channel to sample some physical aspect of the prover which the men in the middle are not able to relay, and then ask the prover (assumed to be honest and cooperative) to say, even over the regular channel subject to relay, what the correct value should be. Since prover and verifier already share a secret, they can use standard cryptographic techniques to protect the integrity (and confidentiality, though generally less relevant here) of the regular relay-vulnerable channel, thereby preventing the fake prover from replacing the true prover's "model answer" with one matching the fake prover's own physical aspect.

[11] According to our definition the authentication phase, which takes place repeatedly, is distinct from the preliminary "enrollment" or "pairing" phase, performed only once and under more controlled circumstances, in which the two principals establish a common secret.

Since the fake prover can't reproduce the true prover's physical aspect (by hypothesis of unrelayability of the special channel) and can't substitute the prover's description with his own (because the regular channel is integrity-protected by the secret shared between the honest prover and verifier), the verifier can justifiably deduce that the principal in front of it is the genuine prover if and only if the value sampled directly over the special channel is consistent with the one received over the integrity-protected channel. That's the core idea in a nutshell.

Looking at the problem in greater detail, the first issue is to define more precisely the "unrelayability" property, and the second is to clarify the subtle interactions between humans and their digital representatives in the course of the verification process: how much of the verification protocol can run unattended and how much of it does instead implicitly rely on human vigilance? We wish to make everything explicit.

Readers should note that using a multichannel protocol (such as acquiring a 2D barcode from a screen with a cellphone camera, as in the classic "Seeing Is Believing" protocol [15]) does *not*, by itself, prevent relay attacks. Without elaborate precautions, the auxiliary channel could itself be relayed[12], which would totally negate its purpose. What we need is a multichannel protocol where one of the channels is by design *unrelayable*.

## 3 Unrelayable channels and protocols that use them

Our investigation of unrelayable channels brings to mind the work by Pappu et al. on unclonable "physical one-way functions" [17]:

> These physical one-way functions are inexpensive to fabricate, prohibitively difficult to duplicate, admit no compact mathematical representation, and are intrinsically tamper-resistant.

To implement an unrelayable channel we require similar properties. In the context of a unidirectional channel in which a detector (sink) acquires information by sampling some physical aspect of an emitter (source), we need:

**weak unclonability:** it must be prohibitively difficult to produce a copy of a given source[13];

[12] For example, the on-screen barcode that Peter acquires with his cellphone could have been generated by Morivictoria by replicating the one acquired by Moripeter's cellphone from Victoria's screen.

[13] Some will claim that this property is redundant because it is implied by each of the next two. But it is conceptually different and therefore we mention it as distinct to clarify the issues involved. By analogy, think of the source as a walnut. Weak unclonability means the attacker can't produce another identical walnut. Strong unclonability means it's infeasible for the attacker to produce any two walnuts that are indistinguishable. Unsimulability means the attacker can't fool you by just showing you a photograph of your walnut.

**strong unclonability:** it must be prohibitively difficult to manufacture two indistinguishable sources[14];

**unsimulability:** it must be prohibitively difficult to fool the sink by simulating the response of the genuine source using some other device[15];

**untransportability:** it must be prohibitively difficult to manufacture a "data pipe" device capable of transporting to another location $L$ the output of the source with sufficient fidelity that a sink at location $L$ would not be able to distinguish whether it is sampling the genuine source or the output of the data pipe.

The unsimulability and untransportability requirements highlight the necessity of looking at the whole system, not just the source and sink endpoint devices, and of including the whole verification process in the evaluation. We must in particular clarify whether we are implicitly relying on the presence of a human verifier (e.g. to check that what is being sampled is the genuine artifact rather than, say, a box of electronics that simulates it, or a set of mirrors and prisms that reproduce its appearance) and the extent to which the overall unrelayability property depends on the care with which the human helper supervises the verification.

To help the reader follow the discussion, we shall now present several examples of unrelayable channels and associated protocols. They are not meant to be adopted as they are: take them as illustrations whose purpose is to help us think about the required properties of an acceptable solution.

To simplify matters, we deal with unidirectional authentication, with one prover and one verifier[16]. Prover and verifier are connected by a regular bidirectional channel, subject to relay attacks, and by a special unrelayable channel, which is unidirectional and goes from prover to verifier. The two principals have previously performed the pairing phase and therefore share a secret with which, using well-known cryptographic techniques, they can make the regular channel confidential and integrity-protected. Notation-wise, in the rest of this paper we shall say "lock $X$ with $K$", written as $L_K(X)$, to mean "cryptographically protect both the integrity and the confidentiality of $X$ using $K$ as the key", for example with encrypt-then-MAC.

With reference to our earlier figures, prover Peter must prove to verifier Victoria that the principal to whom Victoria is talking (and of whom Victoria can

[14] This would be analogous to a cryptographic "collision". As with collision resistance, this clonability resistance property is *stronger* than the previous one, which it implies: if an attacker can't make two identical sources of his own choice then a fortiori he can't make a copy of a designated target source.

[15] This property, too, implies the first one: if the attacker can't simulate the designated source using another device then a fortiori he can't make a clone of it.

[16] We believe that what we really want in most practical applications is *mutual* authentication. For the moment, ignore possible optimizations and assume you can achieve mutual authentication by running the unidirectional protocol twice, once in each direction. Note however that this glosses over some subtle issues about the incentives of the two parties. We shall discuss them at the end of section 3.1.

physically observe/measure/probe some physical aspect over the special channel) is Peter, i.e. the same principal that shares the secret with her. The attacker model is still that man-in-the-middle Moriarty has recruited two accomplices, Moripeter who looks like Peter and will try to fool Victoria, and Morivictoria who looks like Victoria and will try to fool Peter. Victoria wins if she can distinguish whether the principal to whom she is directly talking is Peter (who shares a secret with her) or Moripeter (who doesn't). Conversely Moriarty wins if, after placing Moripeter next to Victoria, and Morivictoria next to Peter, he persuades Victoria that she is talking directly to Peter, even though she really isn't.

Normally, Victoria would run some kind of challenge-response protocol; she could for example ask Peter (or Moripeter, since she can't tell the difference yet) a question such as: "Here is a random nonce $N$. What do you obtain if you lock it with our shared secret $K_{PV}$?". But, with a relay attack, Moripeter would relay the question to Morivictoria, who would ask the same question to Peter, who would provide the correct answer; then Moripeter would get the correct answer from Morivictoria and repeat it to Victoria, who would then be fooled into thinking that Moripeter knew the secret $K_{PV}$, whereas he didn't (and still doesn't).

### 3.1 Example: banknote

In this first example, Peter's unrelayable physical characteristic is a banknote. The banknote is, by design, prohibitively difficult to duplicate (yielding weak and strong unclonability), and there are well-established methods for verifying that it is not a forgery.

Victoria now says, to the principal in front of her (Moripeter if they are under attack, or Peter under normal circumstances): "Give me a banknote."[17] She checks that it's not a forgery (thereby reassuring herself that it is unclonable and that no duplicates of it exist) and then reads its serial number $S$ and burns the banknote, making sure that that particular serial number will never be used again in any other run of this protocol[18]. Then she asks: "What do you obtain when you lock $S$, the serial number of the banknote you gave me, with our shared secret $K_{PV}$?"

How can Moripeter answer that question? He could tell the serial number $S$ to Morivictoria if it helped, but Morivictoria must run with Peter the same protocol as Victoria did with Moripeter (otherwise Peter would not respond), so she must ask for a banknote of that type from Peter, which will have a different serial number, say $S_2$. Peter will lock that $S_2$ with the shared secret and there is no way that Morivictoria can persuade him to lock $S$ instead, since

[17] The banknote must be of a well-specified currency, issue and denomination, to avoid substitution attacks. To minimize the cost of each run of the protocol, it is OK for the banknote to be almost worthless—e.g. one from a country with runaway inflation—provided it is still unclonable. Alternatively, one might use the same technology as banknote printing to create low-value tickets with similar unclonability properties, as is sometimes done for concert or public transport tickets.

[18] Burning the banknote at each protocol run makes $S$ a nonce.

- the banknote is chosen by Peter; and, anyway,
- no other banknote exists with $S$ on it: the only one that did was burnt.

So Moripeter will not be able to answer correctly and Victoria will be able to tell that she received the banknote from someone who didn't know the secret.

**Attack: reverse pickpocketing** Now here is an attack: Moripeter and Morivictoria take a genuine banknote and make a counterfeit copy of it. The forgery is as good as it gets, but it is (by hypothesis of weak unclonability) detectable by someone who runs the proper checks. But, crucial point: Peter is the prover, not the verifier, so why should he be running any serious checks (UV light, colour-changing marker etc etc)? Do you do that on the banknotes you get from your cash machine, or as change from the supermarket? So the scam is for the Moriarty associates to "give" the forged banknote to Peter (as change in a transaction, or by letting him "find" it on the floor, or by reverse pickpocketing him, or whatever) and ensure that he will use it in the subsequent protocol run (no guarantee, but still non-negligible probability). The full run then goes as follows.

Victoria asks Moripeter for a banknote. He gives her the genuine banknote, with serial number $S$. She asks him to lock $S$ with the shared secret $K_{PV}$, which Moripeter doesn't know. Morivictoria asks Peter for a banknote. With some probability, she gets back the forged banknote that has the same serial number $S$: Peter didn't check very carefully and never realized he had a forged banknote[19] so he thinks he is handing over a genuine one. Morivictoria asks Peter to lock with $K_{PV}$ the serial number of the banknote he just handed over; he obliges, and Morivictoria obtains $L_{K_{PV}}(S)$ which she relays to Moripeter who can then correctly answer Victoria's challenge and pass off as Peter.

The lesson here is: who should be verifying the genuineness of the banknote? The prover or the verifier? And the correct answer is: both! If either of them doesn't check with sufficient care, an attack is possible. (NB: if Victoria does not check that she is receiving a genuine banknote, the dual of the above scam, where Moripeter gives Victoria the forged banknote, works equally well.)

This attack scenario also highlights another systems issue we mentioned before: to what extent are we relying on humans to perform additional "implicit" sanity checks? Is it possible for the protocol to run with one machine talking to another machine, in unattended fashion[20]? Assume the crooked machine might exhibit a relaying artifact, e.g. a hi-res screen displaying the banknote, rather than the genuine article. In this case we see that we could in theory run this variant of the protocol in a machine-to-machine setting, provided that both the prover and the verifier contained the approved vending-machine-style technologies for checking that a banknote is not a forgery. Conversely, if we ran this

[19] And if Peter vaguely suspected it was a forgery, he was probably happy to get rid of it by using it in a protocol where it will be destroyed and none will be the wiser—that's an interesting observation about the role of dishonesty in the psychology of scam victims [19] but let's not get sidetracked for the moment.

[20] Imagine for example a car interacting with a barrier, to enter a restricted zone or to pay a road toll or parking charge.

protocol as person-to-machine (a human entering a high-security facility, or a human using an ATM), then it would fall upon the human to perform as careful a check of the authenticity of the banknote as the machine will do. In other words: we do indeed also need unsimulability and untransportability, as well as the strong and weak unclonability that we got from using a banknote!

**Attack: not burning the banknote** Here is another possible attack. Morivictoria asks Peter for a banknote, which he gives her. She pretends to burn it but instead she secretly passes it on to Moripeter. She also asks Peter to lock the serial number with $K_{PV}$, and she gives that answer to Moripeter as well. Now Moripeter can fool Victoria, using the genuine banknote and the $L_{K_{PV}}(S)$ kindly supplied by Peter! To prevent this, we must prevent Morivictoria from being able to reuse the banknote in other runs of the protocol. For example we could say she must cut it in half *and return it to Peter*[21], all strictly under Peter's nose[22]. The interesting problem, here, again, is that the strength of this countermeasure depends on the care with which Peter checks that he received the two halves of the same genuine banknote that he originally supplied—and not, for example, the two halves of a forgery, or of another banknote. But what's Peter's incentive for performing this check? If he is careless and the Moriarty associates succeed in their scam, they are fooling Victoria into opening her door (or giving away her diamond, or whatever) to Moripeter; does Peter lose anything? Not straight away, unless there are external liability issues that penalize Peter for fraudulent use of his authentication credentials. At the baseline level, though, it is Victoria's security (not Peter's) that depends on the care exercised by Peter, and this should be considered a vulnerability. Even though Peter is not actively dishonest, he may not go out of his way in order to protect Victoria, so long as he doesn't lose anything himself by being slightly careless.

This attack scenario explains why we might want to develop a mutual authentication protocol in which the fate of the two parties is more closely entangled than it would be by simply running two instances of the unidirectional authentication protocol one after the other. The reason for wanting a mutual protocol is not to optimize and save on number of messages but rather to bind the incentives of the participants, so that if one of them is sloppy and the other careful then neither gets any benefit from the protocol run (as opposed to the unfair situation in which the sloppy principal is rewarded/protected because the other was careful, and the careful principal suffers because the other was sloppy).

[21] Returning the ashes isn't as good, because Morivictoria might supply the ashes of another banknote and Peter would not be able to notice.

[22] Otherwise another attack would be for Morivictoria to receive the banknote, go to the kitchen to fetch some scissors, pass Peter's note to Moripeter who would then run the protocol by having it cut by the real Victoria; the two halves would be returned by Victoria to Moripeter, then to Morivictoria pretending to have just returned from the kitchen, then to Peter and neither Peter nor Victoria would be the wiser.

### 3.2 Example: accelerometers

In this rather different example, Peter and Victoria have 3D accelerometers that can record, at suitable resolution, a log of the accelerations to which they are subjected. The accelerometers are stuck together and shaken randomly[23] and Victoria checks that the prover could observe the shake. The idea behind this is that "a random shake is unclonable". The protocol runs as follows.

Victoria says: "Give me your accelerometer. Here is mine, too. I stick them together and shake them randomly for $x$ seconds. Now have your accelerometer back. Please lock its log with our common secret and send it back to me."

**Attack: robotic arm** To comply with this request, Moripeter could observe Victoria's shake (the challenge) with his accelerometer, give the precise details to Morivictoria from the accelerometer's log and have Morivictoria reproduce that shake precisely in front of Peter. This last part is practically impossible for a person to do, hence our claim above that "a random shake is unclonable". But what if Morivictoria has a high precision robotic arm that can reproduce the shake to within the required tolerances? Then Peter's accelerometer would record a shake equivalent to that originally performed by Victoria, and Peter would lock it with the secret, and the Moriarty accomplices would win. So this highlights an implicit dependency on Peter being an "alert human" who would spot something amiss if Morivictoria's arm were not of flesh and bones. (But would he actually pay attention to that detail? What if the arm were covered in clothes and appeared to come out of Morivictoria's shoulder?) Thus a machine-to-machine version would not prevent relay attacks.

**Attack: substituting, or tampering with, the accelerometer** Morivictoria could, by sleight of hand, substitute Peter's accelerometer with one into which she downloaded the log communicated to her by Moripeter. No need for robotic arm, but the effect is again that of giving Peter a relayed log instead of the one of the real performance. To guard against this, Peter must ensure that the accelerometer he gets back is really his, and also that it hasn't been tampered with (otherwise Morivictoria could upload the relayed log into Peter's own accelerometer). Once again we raise the warning that we may be relying implicitly on the vigilance of a human Peter and that substitution or tampering might be possible in a machine-to-machine transaction.

**Cameras instead of accelerometers** An alternative might be to monitor the shake with cameras, rather than accelerometers, the intention being that Peter's cameras will never leave Peter's trusted computing base and Morivictoria

[23] The technique of shaking together two objects instrumented with accelerometers was first proposed by Holmquist et al. [13] in the context of device pairing for ubiquitous computing. Later papers [14] perfected the necessary authentication protocols, taking into account error correction and so on.

won't be able to tamper with them. Victoria would then say, without reference to accelerometers: "I'll shake the tip of my finger randomly for $x$ seconds. Please lock the log of the 3D position of my finger with our common secret and send it back to me." Setting aside the interesting but not security-critical computational geometry problem of comparing shake traces taken from different viewpoints, this solution would guard against the last two attacks ("substitute Peter's accelerometer with one containing relayed log" and "upload relayed log into Peter's accelerometer") but would still be subject to the "Morivictoria uses robotic arm" attack.

### 3.3 Example: physical one-way functions

For this third example we use an instance of Pappu's "physical one-way function": a physical object with submicron features that are difficult to replicate exactly and that gives unpredictable but consistent "responses" when "challenged" (illuminated) with a laser. Peter holds the object (or *is* the object—think iris recognition) and Victoria challenges it. The protocol runs as follows.

Victoria shines her laser (in a random way $R$ chosen by her, dictating parameters such as laser frequency, angle, scanning pattern etc) at Peter's POWF object and she records the outcome $O_{Peter}(R)$. Then she tells Peter: "What is the response of your object when illuminated with $R$? Lock the response under our secret and send it to me.".

How can Moripeter answer that question? He will also have a POWF object, but by hypothesis of unclonability it must be different from Peter's. Victoria records $O_{Moripeter}(R)$ and expects $L_K(O_{Moripeter}(R))$ but the Moriarty associates can only produce either $L_K(O_{Peter}(R))$, which has the wrong plaintext inside the outer brackets, or $L_{???}(O_{Moripeter}(R))$, where the correct plaintext is known but the correct key to lock it is unknown to Moriarty.

**Attack: smoke and mirrors** In practice, the Moriarty associates could try to fool Victoria by having Moripeter use a more complex smoke-and-mirrors piece of machinery with its own lasers instead of a regular POWF object. Victoria chooses the laser parameters $R$ and the Moriarty associates, through relay, use these same parameters to interrogate Peter's genuine POWF. They record the response $O_{Peter}(R)$ and then make Moripeter's smoke-and-mirrors machine respond with $O_{Peter}(R)$, rather than with anything physically generated, to Victoria's laser challenge. Then Morivictoria asks Peter to lock the response with $K$, and she relays that to Moripeter, who convinces Victoria with an $L_K(O_{Peter}(R))$ that matches both the shared secret $K$ and the response observed by Victoria.

The two assumptions upon which this attack is predicated are: first, that the Moriarty associates can build a smoke-and-mirrors machine capable of returning arbitrarily chosen laser responses regardless of the laser challenge with which it is illuminated; and second, that Victoria will just shine her laser in the prescribed way without noticing that she is interacting with a smoke-and-mirrors machine rather than with a POWF object. The first of these assumptions is fairly technology-dependent: it concerns the possibility of mounting a specific technical attack against a specific implementation. The second, instead, is once again related to the issue of whether a careful human supervisor will be overseeing the protocol or not[24].

It should also be noted that in practice the attack is much harder than we casually described because Moripeter won't know Victoria's laser parameters $R$ until Victoria actually shines the laser. There is no reason for Victoria to disclose $R$ to the prover before shining the laser. If Victoria only discloses $R$ after having received a laser response from the prover, then Moripeter must perform all of the following difficult tasks:

- figuring out $R$ from the way Victoria shines the laser (instead of being told)
- reproducing those parameters at Morivictoria's end to challenge Peter
- obtaining Peter's POWF response
- relaying that back to Moripeter's smoke-and-mirrors machine

all in real time while Victoria is still operating. If the delay in Moripeter's answer makes Victoria suspicious then this is reminiscent of distance-bounding techniques (all essentially based on measuring whether the response takes longer than would be reasonable), even though conceptually we are still in a different territory. Note that it is very technology-dependent whether it is possible to (a) extract $R$ while Victoria operates her laser and (b) relay the response piecemeal as it unwinds, rather than atomically at the end.

Note that we are now not really discussing the protocol: we are discussing whether or not the proposed special channel has the required unsimulability property.

### 3.4 Example: quantum channel (polarized photons)

This fourth example is even less practical than the previous ones but it is conceptually interesting, since it is based on the inherent unclonability of quantum mechanical states. We leave quantum mechanics to theoretical physicists and we just accept as a black box the assumptions (summarized in the next paragraph) of the BB84 Quantum Key Exchange protocol [1].

Under the assumptions of BB84, Alice the sender can emit photons at various polarization angles that are pairwise orthogonal (say 0, 45, 90, 135 degrees). Her encoding of 0s and 1s into these polarizations is important for BB84 but irrelevant for us. Bob the receiver cannot detect all the potential angles of the incoming photon: he must first choose one of two bases—either the rectilinear one that can distinguish between 0 and 90, or the diagonal one that can distinguish between 45 and 135. If he measures an incoming photon using a base that does not match the photon's polarization (for example measuring a 90-degree

[24] Note that "will be overseeing"—or, better, "is responsible for overseeing"—is quite different from simply "will be present"; in most cases a human will indeed be present, if nothing else to insert the card in the slot, but what matters here is whether the strength of the protocol depends on the degree of care that the human will exercise.

photon using the diagonal base), he will get an incorrect result (either 45 or 135, randomly). The photon is modified by the measurement; so, if eavesdropper Eve listens in on a photon with the wrong base, she "spoils" it for Bob.

We emphasize that we are *not* using (or describing) the BB84 protocol at all—only its underlying physical transmission medium. The BB84 protocol is for building a shared key between Alice and Bob, whereas in our scenario Victoria and Peter already share a key before we even start.

Our protocol runs as follows. Victoria produces a suitably long random string of the symbols {0, 45, 90, 135} and a matching string of the corresponding polarization bases. She sends the second string (of bases) to Peter, locking it with the shared secret[25], and then she sends Peter the actual polarized photons as described in the first string, which Peter can decode correctly by using the bases in the sequence he just received. Then it's Peter's job to send Victoria the string of values he read out, again locked with the shared secret. If Moripeter and Morivictoria splice themselves in, then when Moripeter listens to Victoria's photons he must choose a polarization base to receive each photon, but he won't know the right one because he could not unlock the first message, so he'll get it wrong about half the time and won't be able to tell Morivictoria the correct sequence of photons to retransmit to Peter. Therefore Peter will lock a different sequence of values and, even if they relay that, Victoria will be able to distinguish Peter from Moripeter.

**Attack: relay the photons** An attack here would be for the Moriarty associates to run an optical fibre that shipped Victoria's photons to Peter, without being detected by either. If this were technically feasible, then the channel would lack the required property of untransportability and would not be suitable. However we are as usual assuming that Victoria is sufficiently alert that this attack cannot be mounted without attracting her attention: she would hopefully notice that (Mori)Peter has an extra optical fibre sticking out of the back of his coat.

**Attack: extract the challenge** An over-elaborate and improbable attack sees Morivictoria use Peter as an oracle to check Morivictoria's guess of Victoria's locked sequence of bases. Victoria sends the locked sequence of bases to Moripeter. Morivictoria brute-forces it by trying each possible guess on Peter in turn, as described later. Once she has the correct guess about the bases she gives it to Moripeter, who uses to decode the real photons from Victoria. Morivictoria then sends the same sequence of photons to Peter, who provides the correct locked answer that they can relay to Victoria. (To check each guess, Morivictoria repeatedly sends Peter the same sequence of photons, polarized along the guessed base sequence; if the responses differ, then the guess was wrong, else the guess is shortlisted. She proceeds until only one guess is left.)

This attack relies on (a) the sequence being short enough that brute-forcing won't require years or millennia, (b) Victoria being patient enough to wait for

[25] Note that here we *are* using confidentiality, not just integrity.

the brute-force to take place between her first and second message, and (c) Peter being gullible enough to run the protocol as many times as requested without suspecting anything. It can be thwarted by having Peter include a nonce inside his locked answer so that it is different every time even if the sequence of values is the same[26].

### 3.5 Example: quantum channel (entangled photons)

The other seminal quantum cryptography protocol, E91 [10], uses a different underlying mechanism for quantum key establishment: an entangled pair of photons. This mechanism, too, can be used to build another protocol in our family.

Under the assumptions of E91[27], some external source can prepare pairs of entangled photons and send one photon of the pair to Alice and one to Bob. Each photon can be measured using either a "blue" or a "red" machine and the outcome will be either 0 or 1. If Alice and Bob measure the two photons of an entangled pair using same-coloured machines, the outcomes will be the same; if they measure them with differently-coloured machines, they will be unrelated. Once again, we are not describing or using the E91 protocol—just its physical assumptions.

Our protocol runs as follows. Victoria generates $n$ pairs of entangled photons and sends one photon from each pair to her correspondent (either Peter or Moripeter—she doesn't know yet, but with our protocol she will be able to tell). Then Victoria sends Peter, over the standard channel, a random string of $\{$red, blue$\}$ symbols—one for each of the entangled photons. Peter must then measure each photon with the machine of the specified colour and communicate the result to Victoria over the standard channel. Victoria performs the same measurements on her own photons and checks whether they match, which they should if there is no man in the middle.

In case of relay attack, Moripeter won't be able to obtain the "challenge" string of reds and blues and therefore won't be able to perform the correct measurements even though he has the genuine photons that are entangled with Victoria's. Meanwhile Peter, who can perform the prescribed measurements, will be doing so on photons that are entangled with those of Morivictoria, not of Victoria, and therefore his answers won't match those of Victoria, who will detect the difference.

Note how easy it is to specify and describe a protocol that won't work, even if we can rely on seemingly all-powerful unclonable features such as entangled photons. Victoria generates $n$ pairs of entangled photons and sends one from each pair to Peter. Then she also sends Peter, over the locked channel, a challenge consisting of a string of randomly chosen red and blue symbols. Peter must measure the entangled bits using machines of the prescribed colours and then report

[26] Of course this relies on the cryptographic implementation of "locking" not leaking information about the fact that two ciphertexts might correspond to plaintexts that share a long common portion.

[27] Or rather its simplified description, by Ekert himself, at http://pass.maths.org.uk/issue35/features/ekert/2pdf/index.html/op.pdf.

the answers to Victoria over the locked channel. But here the Moriarty associates relay the challenge from Victoria to Peter, let Peter do the measurements, relay the measurements from Peter to Victoria and appear indistinguishable from the case in which Peter answered directly.

Could you spot the subtle difference between this (broken) protocol and the almost identical one that instead works? Stop reading if you haven't... In the working protocol, Victoria sends the photons to the guy in front of her; in the broken one, she sends them to "Peter"[28].

Note that the "relay the photons" attack (cfr 3.4) applies to this setting as well, with the same caveats.

### 3.6 Why our multichannel approach works

The key insight of our approach is that the standard channel (think radio) connects Victoria to Peter (even if she doesn't know where he really is) and that the special unrelayable channel connects Victoria to the principal in front of her. Victoria challenges Peter over the standard channel and Peter issues conceptually the same response over both channels. The Moriarty associates can only get it right on one channel at a time (they can relay the standard channel or they can "prove presence" over the unrelayable channel[29]) but they can't issue a consistent response over both. All the protocols shown so far are variations of this principle.

## 4 Conclusions and further work

We presented a novel paradigm: a family of multichannel protocols featuring a special channel that is unrelayable. We discussed the properties of unrelayable channels and illustrated possible channels and protocols with imaginative (if not always realistic) examples, chosen to explore the subtleties of the possible attacks, including the crucial role of the human principal in checking for unexpected hardware. We trust readers will recognize this framework as a conceptually new approach to developing protocols that prevent relay attacks.

What we need next is one or more robust and practical implementations of the unrelayable channel, using appropriate physical phenomena and transducers, and suitable protocols from this family to accompany them. Another useful contribution would be a formal analysis of the properties of these protocols.

We see great potential in this new line of authentication protocol research and hope that others will join us in bringing it to fruition for real-world applications.

[28] The broken protocol is thus also impossible to implement: Victoria doesn't know which principal is Peter (whole purpose of protocol); so how could she send *him* the photons?

[29] ...which, for all its wonderful properties, does not need to be particularly versatile: for example, you may not even be able to choose what bits the source will transmit!

# References

1. C. Bennett and G. Brassard. “Quantum cryptography: Public-key distribution and coin tossing”. *Proc IEEE ICCSSP*, 1984.
2. T. Beth and Y. Desmedt. “Identification Tokens — or: Solving the Chess Grandmaster Problem”. *Proc CRYPTO 90*, LNCS 537.
3. S. Brands and D. Chaum. “Distance-Bounding Protocols”. *Proc EUROCRYPT 93*, LNCS 765.
4. B. Christianson and J. Li. “Multi-channel Key Agreement using Encrypted Public Key Exchange”. *Proc Security Protocols Workshop 2007*, LNCS 5964.
5. J. Clulow, G. Hancke, M. Kuhn and T. Moore. “So Near and Yet So Far: Distance-Bounding Attacks in Wireless Networks”. *Proc ESAS 2006*, LNCS 4357.
6. J. Conway. *On numbers and games*. Academic Press, 1976.
7. Ivan Damgård, Jesper Buus Nielsen and Daniel Wichs. “Isolated Proofs of Knowledge and Isolated Zero Knowledge”. “Proc. EUROCRYPT 2008”, LNCS 4965.
8. Y. Desmedt, C. Goutier and S. Bengio. “Special Uses and Abuses of the Fiat-Shamir Passport Protocol”. *Proc CRYPTO 87*, LNCS 293.
9. S. Drimer and S. Murdoch. “Keep your enemies close: distance bounding against smartcard relay attacks”. *Proc USENIX Security 2007.*
10. A. Ekert. “Quantum cryptography based on Bell’s theorem”. *Physical Review Letters*, **67**(6):661+, 1991.
11. G. Hancke. “Security of proximity identification systems”. Tech. Rep. 752, University of Cambridge, 2009.
12. G. Hancke and M. Kuhn. “An RFID Distance Bounding Protocol”. *Proc IEEE Securecomm ’05.*
13. L. Holmquist, F. Mattern, B. Schiele, P. Alahuhta, M. Beigl and H. Gellersen. “Smart-Its Friends: A Technique for Users to Easily Establish Connections between Smart Artefacts”. *Proc UbiComp ’01*, LNCS.
14. R. Mayrhofer and H. Gellersen. “Shake well before use: Intuitive and Secure Pairing of Mobile Devices”. *IEEE Trans Mobile Computing*, **8**(6):792–806, 2009.
15. J. McCune, A. Perrig and M. Reiter. “Seeing-Is-Believing: Using Camera Phones for Human-Verifiable Authentication”. *Proc IEEE Security and Privacy 2005.*
16. L. Nguyen and A. Roscoe. “Authentication protocols based on low-bandwidth unspoofable channels: a comparative survey”, 2009 (manuscript).
17. R. Pappu, B. Recht, J. Taylor and N. Gershenfeld. “Physical One-Way Functions”. *Science*, **297**(5589):2026–2030, 2002.
18. D. Pavlovic and C. Meadows. “Deriving Authentication for Pervasive Security”. *Proc ACM ISTPS 2008.*
19. F. Stajano and P. Wilson. “Understanding scam victims: seven principles for systems security”. Tech. rep. 754, University of Cambridge, 2009.
20. F. Wong and F. Stajano. “Multi-channel Protocols”. *Proc Security Protocols Workshop 2005*, LNCS 4631. See also the extended and revised version in *IEEE Pervasive Computing* **6**(4):31–39, 2007.