Using Human Interactive Proofs to Secure Human-Machine Interactions via Untrusted Intermediaries

(Transcript of Discussion)

Chris J. Mitchell

Royal Holloway, University of London

It's ironic that these hard problems, such as character recognition, have been known to be hard for a long, long time, and yet almost as soon as people make crypto things out of them, they get solved. Actually it's not quite the way you think because what's happened is that for the examples that get automatically generated, there are special techniques which work just because they've been created deliberately. It's not trivial to produce things that are really hard to solve, and some of the ideas for distorting characters have been quickly broken, but I believe there are some around which are quite robust.

Frank Stajano: Are they written by undergraduate students?

Reply: Yes, that's right, I guess you want things that are automatically generateable, because that's really nice. And there are other techniques apart from distorted characters, I've seen examples where you get six pictures and you're asked to choose which ones of these six show the same person, and they might show several people, but you have to spot that Tony Blair is in three out of those six, and these are the kinds of things that are quite hard for computers to do, or to count objects, you know, count cars in this picture.

Marios Andreou: What about speech, is that feasible, like audio to discern a voice pattern against noise, or something?

Reply: Maybe. You want things that are quite easy for computers to set, these might be quite hard for computers to set, although you could give a computer a large library of photographs with some rules about how to use them.

Jeff Yan: Actually, some CAPTCHAs implement this.

Reply: This talk really stems from the observation that we can use a HIP as a kind of communications channel between the computer and a human, which no other computer can intercept. So if we're using these characters the computer can display a message to the human that the human can read but no other computer can read. So we've got a confidentiality protected channel, and of course it's not very confidentiality protected because human beings who are monitoring the channel can also read it, but not machines, which is kind of nice because, to take

B. Christianson et al. (Eds.): Security Protocols 2006, LNCS 5087, pp. 171–176, 2009.

bank card examples, these malicious attorneys, or whatever one might call them, that might get left in card readers, are typically machines, they're there to make a record of what happened. They can make a record perhaps of the images, but they can't actually read the data in real time, they need a human help to do that. And we can generate also a reverse channel, which is also confidentiality protected, but not authenticated.

So, in principle at least, with HIPs we can get a two-way confidential channel between the user and the computer, even when the communications pass via an untrusted intermediary but the channel is unauthenticated, it's also only confidential against machine eavesdroppers. But I claim that's a potentially useful property.

I've already hinted at the first scenario, where the PIN is sent to a smartcard via a confidential channel, by displaying concealed digits next to the buttons. The problem is, a bad man-in-the-middle could put up its own images, the bad terminal in the middle generates its own distorted images and learns the PIN, because it knows what the images mean that the user is now selecting. But it can't tell the card the PIN because it can't talk to the card, because the card is displaying its own distorted images, which the terminal can't read. But this doesn't offer much security because the first time the card asks for the PIN, the terminal displays its own images, learns the PIN, and then tells the user, oh that PIN failed, and then lets the card prompt for the PIN, so the terminal can learn the PIN, and allow the card to get the PIN, but at the cost of forcing the user to enter the PIN twice.

Ross Anderson: A sensible implementation would see to it that the images that the smartcard puts out would be customised to the user. It's also worth remarking that, again, one of Ueli Maurer's theoretical contributions[1] in 1996, was that you can get an authenticated channel from a confidential channel in general, but it also goes the other way, because if the smartcard sends me some random nonce on this machine confidential channel, and I type that random nonce, I have authenticated by existence.

Bruce Christianson: Or you can go a little bit better than that, the nonce could be a weak key that you then use to do an encrypted key exchange[2]. Now you have a strong key, and the human learning the secret after that point, is no threat.

Ross Anderson: Well my mental arithmetic isn't quite enough to do EKE.

Bruce Christianson: Oh well presumably you've got some PDA, or something like that, that can do it for you?

Ross Anderson: In that case I'll just use it as a card number.

[1] Journal of Computer Security **4**, 1: A Calculus for Security Bootstrapping in Distributed Systems.

[2] "Secure Sessions from Weak Secrets" LNCS 3364, page 190.

Mike Bond: I think there is limited crypto that can still be done in your head; some of the stuff that George Denezis and I talked about last year for secret society protocols would like you to be able to XOR very short numbers[3]. Things like that can be done, though maybe not by everyone.

Reply: Of course you also have the problem that if the machine records the transaction, it can play it back to a human later, and the human will assist the machine to learn the PIN.

Ross Anderson: We did have a system in COPAC which pre-dated EMV by several years, whereby your receipt also contained a MAC of the transaction, which was computed by your card, using a different key from the MAC that was sent off to the bank, but which key was also known to the card issuing bank, and could be produced for third party verification in the event of a dispute. It was interesting to see that this function, as it benefited the customer solely, and the bank not at all, was dropped from the EMV specification.

Mike Bond: I've got a suggestion for usage.

Reply: Oh, good.

Mike Bond: Well not really, it's actually an evil usage scenario. HIPs, and that sort of technology, is I think already deployed wide-scale by spammers for sending things that can't be read by machine, an image of, buy Viagra, and a machine doesn't know what it is, so it has to send it to the human to be read to see if it's spam or not.

Reply: So that's, in a sense, an application of what I'm doing here, but an evil application.

James Heather: It seems to me that there is a danger here because usually when you do something like RSA crypto, you're banking on a fairly well attested Moore's law, to protect you against advances in the technology, and algorithm design, unless somebody is going to break RSA, which looks fairly unlikely. But here you're relying on nobody coming up with something that manages to handle the sorted images, which, as far as I'm aware, could happen any moment, it's just that no-one's really worked out how to do it yet.

Reply: That's true. Of course you could say that about AES.

James Heather: You could, yes, but it seems intuitively less likely.

Reply: Yes, I agree. The first few years of the history of CAPTCHAs has been a sorry tale of broken schemes, but there are some which are still believed to be robust.

[3] "The Dining Freemasons" LNCS 4631, page 258.

Bruce Christianson: I do like Virgil Gligor's idea that you should make the CAPTCHA related to some hard problem in AI, because that way you get a good publication out of it either way.

Reply: I think the problem is, where they are known hard problems, we don't construct general instances of the hard problem, we construct special cases, a bit like knapsack crypto. Knapsacks are very hard, but unfortunately the instances we've created were easy ones.

Ross Anderson: I took your hint and looked up CAPTCHAs in Wikipedia, and about two thirds of the Wikipedia entry consisted of criticisms of CAPTCHAs and accessibility. For example visually impaired people are of course affected, and there's still thirty five thousand deaf/blind adults in the USA, so it looks like there are some social policy issues.

Bruce Christianson: On the other hand if you're colour blind, it's relatively easy to invent a CAPTCHA that only you will be able to do, or at least only people who suffer from your kind of colour blindness.

James Heather: Or use shade, which can simulate colour blindness very easily.

Bruce Christianson: Only if it can solve the underlying problem.

James Heather: Yes, OK, but your underlying problem is then, give me your security.

Ross Anderson: Well that depends on the threat model.

Bruce Christianson: Yes, it depends on the level of authentication that you require.

Ross Anderson: Presumably a mature technology admits customisation so that if you're red/green blind then the CAPTCHA you get at the cash machine will not be solvable except by someone with the same condition.

James Heather: You can put your disability here.

Ross Anderson: And related language for the audio components.

Frank Stajano: Well saying the CAPTCHA in Sanskrit is probably almost sufficient obfuscation.

Matt Blaze: So that suggests the biometric CAPTCHA, so for example, give you an eye chart and ask you to enter the first line you can read.

Ross Anderson: And you get somebody's age by going up the audio scale until he can't hear it anymore.

Reply: Yes.

Ross Anderson: Well hang on, you could use audio masking effects. I wonder if you could do this, I wonder if you could produce an audio signal that would be interpreted differently by a 21 year old and a 41 year old.

Srijith Nair: Just play contemporary music.

Ross Anderson: And you could play Mozart and make the 21 year old go away. It would be possible to use some signal processing trick so that you would actually hear different words spoken depending on your age.

Srijith Nair: Maybe if you used high frequencies.

Ross Anderson: Exactly, with a high frequency component which is tailored to change the performance. Well, a good research project for someone.

James Heather: You can certainly do a similar thing with images.

Srijith Nair: I thought something like that was done in Cambridge a couple of months ago. The intention was you could drive away kids, because kids can hear a particular frequency, but the grown-ups can't hear a particular frequency.

Ross Anderson: There's a guy in Britain who developed a teenager repellent[4] put near his shop, and then they announced it's unethical and he's not supposed to use it[5].

James Heather: You've got no protection with something like a human and a machine operating together, because they could strip out the high frequencies.

Ross Anderson: In that case you put something that's going to intercept the challenges.

Bruce Christianson: It depends a great deal on your threat model. Often what you want is a secure channel between the endpoints, but you don't know in advance who the individual at the endpoint is going to be. Often indeed that subsequent communications is going to happen over some other medium, and we just want the secure channel to establish the key for that. If you could have agreement about where the endpoint was, it's quite good because it means that for the attacker capturing the transaction, and getting a human involved later, it's too late. They can't go back into the agreement. But the hard problem is

[4] `http://en.wikipedia.org/wiki/The_Mosquito`

[5] Although just playing Mozart seems almost equally effective.

ensuring there isn't a false front-end, putting a camera on the screen and relaying it to someone else, and the poor user thinks that's the endpoint.

Ross Anderson: There's almost an isomorphism with these scenarios that David Wheeler was playing around with ten, fifteen years ago, where you used some kind of aid to enter passwords. Various people played with schemes like this, where you have a demonic aid that enables you to answer the challenge from the machine through an untrusted interface, and the bank-end stuff, it strikes me, is much the same in both cases. Such things are relatively trivial to design if they're going to be used one-off; they're much more problematic if the TSB gives the same card to each of its four million customers and they need to customise them all in manufacturing the ten by ten grid with different funny colours.

Bruce Christianson: If you really need to know which customer it is at that stage. . . sometimes you just want to know that you are speaking to a NatWest customer, and the customer wants to know they've got a secure channel before they enter their details.

Ross Anderson: Well I presume the obvious application of this is you can see that you're talking to a real access point, and nobody really knows how to solve that.

Mike Bond: In certain scenarios that you mentioned, say maybe a smartcard where you've got high bandwidth, you're not going over the Internet, could anything be done? Given it's a hassle for them to record an image, why can't it go up from images to full motion video? A smartcard may have a very limited power processor, but can it produce an output we could render at extremely low resolution on some kind of screen that renders something that can just be read by a human. It is a vast amount of data to store under the process, but a little scrolling message, or something, and that could be a race for tagging defence there.

Steven Murdoch: I recall something, probably a year ago, where they proposed that smartcards should render 3-D graphics, which would be even harder than sending CAPTCHAs.

Mike Bond: But I think when removing stuff it would be good. All a potentially evil device has to do at the moment is forward this display, you know, give me direct access to the inputs of anybody else in the display, and I'll make a colour and light show. So that should be what the terminal does now as a feature, it says, don't give me the amount and I'll display it, it's, right, I'm now talking serial to your LCD control.