Combining Crypto with Biometrics: A New Human-Security Interface

(Transcript of Discussion)

Feng Hao

Computer Laboratory, University of Cambridge

I present my research on combining cryptography and iris biometrics. This is work with Ross Anderson and John Daugman. It is a short talk so I will leave out the technical detail.

The motivation of the research is to incorporate advanced biometric authentication features into cryptography. We find that cryptography lacks the involvement of a human factor. In authentication, you would use a password or a token, but there is no real human factor involved. We studied the iris biometric because it is one of the most reliable biometrics discovered so far. There are however certain issues with the iris biometric. First, it is fuzzy. Second, its storage is quite controversial for privacy reasons. And third, it cannot be kept secret by its very nature. These limitations apply to biometrics in general.

In Unix, you don't store the password in plain text. Instead, you apply a one-way hash function. But you cannot do the same with iris codes because they are fuzzy. If you hash an iris code, it would destroy all the information content. So in our research, we devised a method to map the 2048-bit fuzzy iris code into an exact 140-bit string. This mapping is repeatable with a 99.5% success rate.

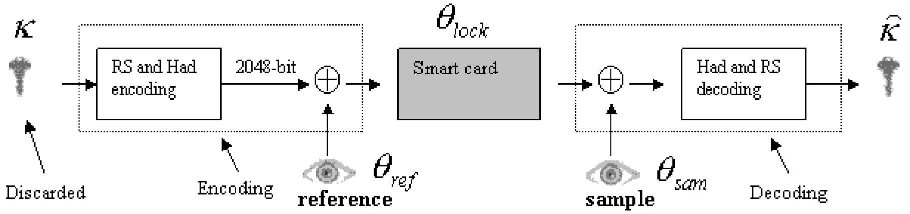
Our technique is based on error correction codes. First I will explain the error characteristics in the iris codes. There are two types of errors. First there are random errors – errors dispersed randomly across an iris code. Second there are burst errors – caused by undetected eyelashes and specular reflections. We devised an error correction scheme to deal with these two types of error. At the top level, we wanted to design an error correction scheme in such a way that it will have the error correction capability at a cutting point to correct errors just enough for authentic users, and not more than that.

We segment the iris code into 32 blocks with 64 bits in each block, and we apply a Hadamard code to correct up to 25% of the bits in each block. This is roughly the cutting point to discriminate between the same eye and a different eye. However, certain errors are clustered in some blocks to give us error blocks. Hence, we have a second layer of error correction using a Reed-Solomon code which corrects these burst errors.

Here is a basic scheme. It is a two-factor scheme. Key reproduction is based on two factors: iris and token. The token is something that we can keep secret, but the iris is not. On the left-hand side of this diagram is the registration part which is also the encoding part. We generate a 140-bit random string, and encode it to 2048 bits. Then we XOR this with an iris code, which is also 2048 bits. The

B. Christianson et al. (Eds.): Security Protocols 2005, LNCS 4631, pp. 133–138, 2007.

result is called the locked code. We store it on a token together with a hash of the key so that we can verify later whether it was generated correctly.

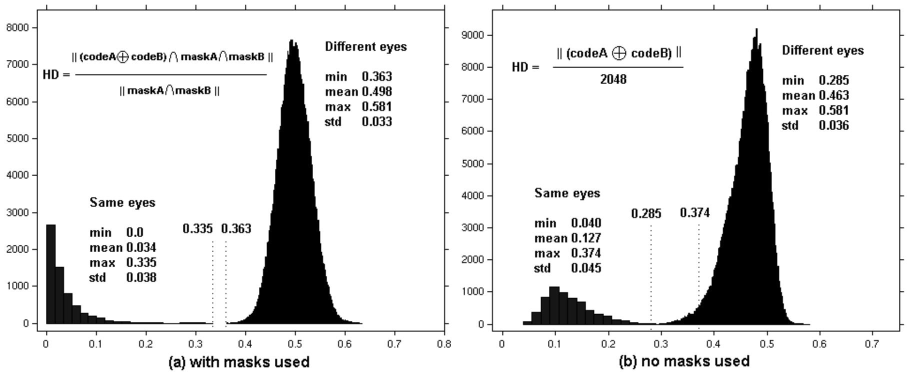


On the right hand side of the diagram is the key generation part, that is the decoding part. We have an authentic iris code which is not identical but close to the reference one. By going through the decoding procedure, we are able to recover this 140-bit key exactly. Here we let e denote the Hamming difference between the two iris codes. If the iris code is authentic, then e has a relatively small Hamming weight. On the other hand, if the sample is not authentic, the Hamming weight will be relatively big. We designed a coding scheme to handle the errors with e up to 27%. Below this threshold we are able to correct the effect of the noise, and recover the exact string. We can check against the hash value which is stored on the token.

What is the performance of this scheme? Here, we must decide the number of error blocks we want to correct using the Reed-Solomon code. We choose a Hadamard matrix size of 64-bit, so we have 64 bits in one block and 32 blocks in total. RS is the number of error blocks we can correct. We find a suitable choice is RS = 6. At this point we can get a key length of 140-bit, and the false rejection rate is pretty low, only 0.5%. This performance is much better than other key generation implementations based on biometrics, like, voice, handwritten signature, fingerprint, and face. The common false rejection rate is 20%. Our 0.5% is much smaller.

Here is the histogram of the database we use. At this stage we haven't included the mask information into our coding scheme because of some technical difficulties. We will do that in our future research. The diagram on the right-hand side is a histogram without using the mask information. As you can see, there is some overlap here and that's why we cannot get zero error rates. Ideally if we can include the mask information, we can surely reduce the error rates even further.

What is the security of this scheme? The initial motivation of our work is that if you have, for example, a biometric ID system, then you will have to store the iris code for each person in a central database. That will cause a lot of privacy concerns. Our thinking is to just store a random sting, and the person is identified by regenerating that random string. The privacy concerns for this approach are much less.



(a) with masks used

(b) no masks used

There are some other applications of our technique. For example, you may want to use the generated key as a cryptographic key. In that case it'd be better to add another factor, a password. We can use a password to permute the Hadamard matrices, so that if an attacker wants to derive this key then he has to compromise all three factors.

Each of the three factors has intrinsic limitations. For example, a password can be stolen, or guessed. As for the token, it can be stolen or reverse-engineered. But it is possible that a password and a token can be kept secret by a careful person. For the biometric, a biometric copy can be stolen. An attacker may be able to obtain your fingerprint. In the case where the attacker has the token only, he may present his own iris and try his luck to see whether he can get through the system. But the probability of that is very small. But the problem is that the iris biometric cannot be kept secret by its nature.

In our 3-factor scheme, if the attacker wants to compromise the key, he has to compromise all three factors.

Mark Lomas: Can I suggest that the problem is not that the biometric can't be kept a secret. The problem is the need for a secure channel from whatever makes the measurements to whatever does the composition which you have just explained. So I think your approach is a sensible one if I can incorporate the crypto part into the camera that does the scan, but not if I'm presented with, say, a smartcard that does the processing and an untrusted camera. Secrecy is not actually the issue.

Frank Stajano: Well it is in a sense. John Daugman has this famous story of the Afghan girl on the cover of National Geographic[1]: you can recover her iris data from just a photograph of her, so in some sense even if the camera does the crypto you can still fool it by presenting a picture of an iris.

Mark Lomas: But if you recall some of John's earlier papers he explains how you work out whether this is a live iris as opposed to just a photograph.

Bruce Christianson: I think Mark's original point is right, the trick is to ensure that the channel between the person and the crypto really does have all

[1] See http://www.cl.cam.ac.uk/~jgd1000/afghan.html

the properties that we want it to have, and this includes checking that you've got a live person, not a head attached to a pump in a hatbox.

Frank Stajano: That's a longer channel, to the iris not just to the camera.

Reply: It's actually not that easy to steal an iris image. You need a special camera. A digital camera doesn't work. It's a near infra-red camera and the image is captured within one meter.

Matt Blaze: So one way to do that is by setting up a biometric scanner that the person is required to use?

Reply: Yes, that is correct.

Mike Roe: In previous talks we've discussed the proliferation of CCTV cameras in the UK, there's even some have been developed which use infrared to see in the dark. As you walk down the street there's already all the technology needed to get you iris image looking at you.

Ross Anderson: But I think the point of this is different. In the real world everything can be faked, forged, guessed, etc. Sure you can end up with a database of the iris of most people in the UK by setting up some strategic cameras, but you won't know their names unless we've also got scanners which read the ID cards in their wallets. You know most of their passwords, because an on-line dictionary is just one mouse-click away, and you can steal large quantities of tokens by subverting some person at the post office. Now all of these have a probability attached. If you take the NSA mathematicians view that if it's not perfect it's no good, and you consider that you will only ever thing about security mechanisms if you're absolutely guaranteed by a complexity theoretic proof that you've 128-bit minimum security, then I suppose this work has no place whatsoever in that lecture course. But that doesn't imply that it has no place here.

Matt Blaze: Ross, I'm sorry, I think you've been replaced with somebody who looks like Ross Anderson. [Laughter] I'll used my biometric scanner just to be sure, but the old Ross never would have agreed to use it, and I know that if you agree to use it, then you are in fact an impostor.

I think the underlying problem is not that it's cryptographically imperfect, or that there's some attack in which we convert every molecule on earth into a super computer, and run it for ten thousand years and break it. The fundamental problem is that it doesn't solve is, I subvert one person using this biometric authentication system, either by collecting their database, or replacing the software that does the collection with something that sends the data to me as well, and then I can use that information to subvert your biometric authentication of these users everywhere else they go.

Ross Anderson: If all I were doing is deriving a single key from somebody's biometric, which is what most schemes do, then that's a fair statement, but in this scheme here we have a different key for every application.

Matt Blaze: But if we can get your biometrics by stealth, if I'm not misunderstanding, this scheme only works if everyone is using it, but if somebody, somewhere, is using a dummy scheme then I can recreate the biometrics for the purpose of fooling you.

Ross Anderson: True. In that case you would fall back on the password. I suppose a security evaluation of a scheme like this is to ask, how good is the best that we can do given the apparent limitations of the underlying technology.

Matt Blaze: The technology is very limited.

Ross Anderson: And if it gets the most bangs that it is possible to get out of that technology then it's a good scheme. But how do we go about systematizing that kind of evaluation?

Alf Zugenmaier: How stable is the biometric in time? Are there problems like, a woman becomes pregnant, the iris totally changes, and will not be recognised again, so the key is wrong for good.

Reply: No, the iris is fixed for a lifetime.

Hoon Lim: Even after a night at the Eagle?

Reply: Yes. Although I also saw reports that some patients had a cataract operation and that changed the iris structure. But I think the main point in our scheme is that the key is random, and completely independent from biometrics. That is quite important because biometrics, by nature, cannot be kept secret. What happens when the biometric copy is stolen? It has no impact on the key because the key is completely independent from biometrics.

Yvo Desmedt: There are people who make artificial irises. All you need is a personal computer and a printer, and an infrared camera, and access to some published papers.

Mike Bond: One of the interesting challenges here is normally in the active counterfeiting scenario: you have anti-counterfeiting measures which mean that you need a really expensive press to print bank notes. Whereas we've got this one unique source of decent irises, ourselves. Now it seems a lot of the anti-counterfeiting protection in irises is currently in the measuring of the irises, whereas once you've got the data from it with your infra-red camera, which costs £1000 today but tomorrow costs £10, that's it. So how do we exploit things that aren't in the measuring phase in verification?

Ross Anderson: There's now a proposal that US Immigration and Naturalisation Service think about egress controls, so that when you leave America on a plane, some pleasant young man or young lady will come up to you, look at your boarding card, scan one of your fingerprints, and if it matches against the fingerprints that you gave when you entered the country then you'll be allowed to leave. So if the future contains a world in which this kind of biometric scanning is done, (which may be evil, but if it's going to happen anyway) then many of the liveness problems can be overcome at a critical level. For example, you can design next generation mobile phone cameras to scan somebody's iris.

Virgil Gligor: I think that all that happens is the border with Mexico will become very popular.

Frank Stajano: Another problem is to keep people out, not not allow them to leave.

Mike Roe: There's scope for some protocol attacks here. You've only got one pair of irises, you don't get a separate iris for each security application. So firstly we've got to secure against the possibility that there is some other protocol using this mechanism that will cause the iris code of everybody in the world to become public. Secondly, you have to design the protocol so you don't break any other protocol that relies upon the mechanism. I think you've got the pieces here to do it, but it's not straightforward.

Ross Anderson: And that's why you have to have statistically independent keys, because you just have to assume that you're using your iris to do many different things.

Bruce Christianson: This is a nice mechanism, particularly the mangling of the iris data with the key and the potential integration of that with access class. But it's going to take a long time to work out some of the other implications, and it certainly brings a whole new meaning to the phrase "cardholder not present" [laughter].