Selective Disclosure (v0.2)

Ben Laurie (benl@google.com)

May 11, 2007

Abstract

Digital signatures are widely used on the Internet. One application is in identity management, where they may be used to authenticate (that is, prove identity or entitlement) or to make verifiable assertions (e.g. "this person is over 21" or "this person is a UK citizen"). However, traditional digital signatures have implications for privacy – these can be addressed by zero-knowledge and selective disclosure proofs. This paper explores both the need for and the properties of selective disclosure proofs.

1 Summary

There is a growing desire to allow users to be more in control of their online identity, as seen in Microsoft's CardSpace[Cam, Cha06], OpenID[RR06], Project Liberty[Lib], Shibboleth[CCH+04, Shi] and so on.

There is also a desire for users to be able to state things about themselves using identity management systems, for example, their age, their nationality, their credit card numbers, their airline seating preferences and so forth.

And finally, there are reasonable expectations of privacy[1].

Unfortunately, the technologies in wide use to satisfy these requirements are inherently incapable of achieving all of them. This paper explains why, and also describes a family of cryptographic signatures that, although not new, are not widely known.

First, I will explain the need for digital signatures and how they are currently usually provided. Then I will lay out some rigorous requirements for privacy and show how standard digital signatures cannot achieve them. Finally, I will explain selective disclosure proofs and how they achieve the privacy requirements.

It is assumed that the reader has some familiarity with identity management and digital signatures.

2 Assertions and Digital Signatures

An assertion[2] is a statement made about someone or something. It can be thought of as having:

- A subject: the person or thing about which the statement is made.
- A value: the value of the assertion.
- A claimant: the identity of the person or thing making the assertion. This may not always be present.

The details of exactly how these things are represented are not important right now - each system has its own format (for example, X.509[HFPS99] uses ASN.1[asn88], Liberty (and others) use SAML[CKPM05] and so on).

However, no matter how represented, an assertion is of limited value, in many circumstances, unless it can be verified. For example, there's little point in me stating that I am over 21 or a UK citizen unless the person relying on that statement can check that it is actually true.

The usual way to achieve this is through digital signatures: the claimant takes the assertion (expressed digitally, of course) and signs it. He then gives the signed assertion to the subject, who can present it to the relying party as needed. The relying party (RP) can then verify that the signature was made by the claimant - and, if the claimant is one that the RP trusts[3] can then proceed

[1] Although it has been repeatedly shown that the average end user is easily persuaded to give up their privacy, despite professing to care about it[AG05].

[2] Assertions are also often known as attributes or certificates in this context.

[3] By "trust", I mean "is prepared to believe" or, equivalently, "is prepared to rely upon".

safe in the knowledge that the assertion is, in fact, true.

3 Privacy Requirements

Kim Cameron's famous "Laws of Identity"[Cam06] include

> 4. Directed identity
>
> A universal identity system must support both omni-directional identifiers for use by public entities and unidirectional identifiers[4] for use by private entities, thus facilitating discovery while preventing unnecessary release of correlation handles.

This rather opaque language can perhaps be better understood through this excerpt from the explanatory text

> ... a consumer visiting a corporate Web site is able to use the identity beacon of that site to decide whether she wants to establish a relationship with it. Her system can then set up an unidirectional identity relation with the site by selecting an identifier for use with that site and no other. A unidirectional identity relation with a different site would involve fabricating a completely unrelated identifier. Because of this, there is no correlation handle emitted that can be shared between sites to assemble profile activities and preferences into super-dossiers.

The intent of this is quite clear – if I share data with one website, then that data should not be linkable to data I share with any other site I interact with.

What Kim has not captured entirely clearly in his laws are the consequences of this requirement. In particular, the effects the requirements have on assertions: they should not themselves provide an avenue for correlation.

I summarise this all rather more succinctly in my own laws of identity, which I reproduce in full here

> I claim that for an identity management system to be both useful and privacy preserving, there are three properties assertions must be able to have. They must be:
>
> - Verifiable
>
> There's often no point in making a statement unless the relying party has some way of checking it is true. Note that this isn't always a requirement - I don't have to prove my address is mine to Amazon, because its up to me where my goods get delivered. But I may have to prove I'm over 18 to get alcohol delivered.

[4] I would prefer the terms "universal" and "independent" rather than "omnidirectional" and "unidirectional"

- Minimal

  This is the privacy preserving bit - I want to tell the relying party the very least he needs to know. I shouldn't have to reveal my date of birth, just prove I'm over 18 somehow.

- Unlinkable

  If the relying party or parties, or other actors in the system, can, either on their own or in collusion, link together my various assertions, then I've blown the minimality requirement out of the water.

Note a subtle but important difference between Kim's laws and mine – he talks about **identifiers** whereas I talk about **assertions**. In an ideal world, assertions would not be identifiers; but it turns out that in practice they often are.

4 Assertions as Identifiers

So why do assertions turn out to be identifiers? Consider once more what is in an assertion: a subject, a value, a claimant and a signature (of which the last two are optional). If the identity system is respecting privacy, then the subject will be different for each relying party (because the subject will be identified by the unidirectional identifier established with that particular relying party). A naive analysis would lead you to believe that this is good enough - no two relying parties would see the same subject, and therefore no linkage could be established.

But this is not so. Firstly, the value of the assertion will be the same at each relying party. This is bound to be at least partially identifying, or there would be no point in having it (that is, if everyone would have the same value, then you might as well not bother with the assertion at all). For example, if it is my address, then (in my case) that narrows me down to one of four people. If it is my date of birth, then that narrows me down to (approximately) one in 20,000[5] of the world's population. Each assertion I show further reduces the set of possible people that could have shown that assertion until it becomes possible with high probability for two relying parties to work out what their respective "unidirectional" identifiers for me are.

But it is worse than that. In the case where an assertion has a claimant and a signature[6] then the claimant must have generated both versions of the assertion (that is, one for each unidirectional identifier). Because of the nature of the signatures widely used for assertions (RSA[RSA78] and DSA[FIP94]) the signed assertion shown to the relying party is exactly the same as the one the claimant created – that is, bit for bit identical. Therefore it is possible

[5] Assume an average life expectancy of 60 years, then 60 x 365 = 21,900.

[6] Note that an assertion with a claimant but no signature is not worth the paper it is written on.

for the relying party and the claimant to collude in order to link any other "unidirectional" identifiers the user may have.

The situation is even worse if assertions are used as they usually are – that is, bound to my "real name" or some other omnidirectional identifier, like my National Insurance number, for example. In that case, the actual assertion shown is always the same, and so the collusion of the claimant is not even required. Most identity management systems with any pretension at all to privacy fix this problem by having the user present their "universal" assertion with their omnidirectional identifier on it and in exchange give them a temporary assertion with a unidirectional identifier – this can either be done with the original claimant or with some mutually[7] trusted third party. But, of course, whoever issues this temporary assertion can trivially link it to the original assertion, and so we are back to the scenario described above, where relying parties and assertion issuers can collude to link assertions and therefore identifiers.

5 Zero Knowledge and Selective Disclosure Proofs

Now that we have identified the problem, is there a solution? Happily, the answer is "yes". In fact, there are several, but I will describe in detail the one I consider most useful.

Have you ever seen the "Where's Wally?" ("Where's Waldo?" in the US) series of puzzle books? These have a character, Wally, in a stripy red jersey, hidden in a hugely crowded scene. Your task is to find Wally. Suppose you have done so, and you want to prove to me how clever you are without giving away the location of Wally – how can you do this?

First you find a sheet of card that is twice the size of the crowd scene (that is, each side is twice the corresponding side of the picture) and in that card you cut out a Wally-shaped hole. You place the card over the picture such that Wally is showing in the hole, thus demonstrating to me that you know where Wally is without revealing to me where he is[NNR99].

This is a real-world example of a zero-knowledge proof. You have proved your knowledge of something (the position of Wally) without revealing the actual knowledge to me. Zero-knowledge proofs (also known as ZKPs) are also possible with cryptography. For example, I can prove that I have a signature from someone without revealing the actual value of that signature.

Selective disclosure proofs are related to zero-knowledge proofs, in that they hide some of the knowledge they are making proofs from, but not all of it. A classic example of a selective disclosure proof is this: I have a signed assertion stating my date of birth, but all I want to do is to prove I am over 21 - using a selective disclosure proof I can show that I have a signed assertion where the date of birth is before 21 years from today without actually revealing the date of birth.

In the proof I actually prove two things:

[7] That is, by both the user and the relying party.

1. That the date of birth (represented as a number, of course) is less than some particular date.

2. That this fact was signed by some particular claimant.

Of course, this isn't much use unless I can link this proof to my identity, somehow. Fortunately, selective disclosure proofs can also manage that trick, and even without revealing my identity. What happens, in practice, is that I have two groups of signed assertions (at this point it might help to think of them as certificates).

id=1234abcd key=5678efgh

where "key" identifies a public/private keypair for which I have the private key.

id=1234abcd birthdate=25th March 1960

These may have been issued (and therefore signed) by two different claimants. Using selective disclosure I would then prove that

1. I have the private key corresponding to the public key in the first statement.

2. The "id" fields in the two certificates are the same.

3. The date of birth is prior to 21 years before now.

4. Both certificates are signed by their claimants.

An important point to note is that, unlike more traditional certificates (for example, X.509 certificates, or SAML assertions) I do not ever actually *show* the relying party these certificates – what I do is prove that I have them and prove things about them. And, what's more, each time I prove it, the proof is different (and not linkable to the previous proof, even by the issuer of the certificate). This means that the relying party (and everyone else) is denied access to any material that might allow them to link any part of the proof to any other, or to any proof seen at a different time, or to the use of the certificate at any other (or the same) relying party.

If the proofs cannot be linked, then at each interaction instead of gaining an extra piece of information about you all that is gained is an isolated piece of information about someone who cannot be linked to any other isolated piece of information.

Of course, it is important to understand that selective disclosure can do nothing about inherently identifying information: if I want a physical delivery, for example, then I must give an address. That address is likely to limit my identity to one of a small number of people. Similarly information like telephone numbers, email addresses, tax IDs and IP addresses tend to be highly linkable.

Clearly selective disclosure will not obviate the need for users to be well informed about what data is being revealed, and to make choices that help to preserve their privacy - but it does, at least, prevent users from being exposed to less obvious correlation of their personal information.

6 Random Extras

- It is also worth mentioning that using selective disclosure effectively tends to mean rethinking the way things are done. All too often decisions about what users can and cannot do are expressed in terms of their identity: "Ben Laurie is allowed to edit this page". In order to use selective disclosure well it is better to phrase this in terms of entitlement instead: "The owner of this certificate is entitled to edit this page". This allows selective disclosure to minimise (or eliminate, in this case) identifying information.

- I am aware of two selective disclosure schemes that are practical[8]. The first is due to Brands[Bra00] and the second due to Bangerter, Camenisch and Lysyanskya[BCL04]. Both of these have implementations available in the form of PRIME's[pri] Idemix[CH02] and Credentica[cre].

- I said that selective disclosure is not the only way of solving these+ problems. Other mechanisms that may help include zero-knowledge proofs[FFS88, GO94] and blind signatures[Cha82] but none of them are as flexible as selective disclosure proofs. Note that the selective disclosure proofs mentioned above rely on zero-knowledge proofs and blind signatures for their operation.

7 Conclusion

Traditional signatures schemes make it impossible to construct identity management systems that preserve privacy, but the little-known selective disclosure technology rescues us from this dilemma.

All we have to do is start using it!

8 Acknowledgements

Thanks to Adriana Lukas, Cat Okita, Wendy Seltzer and Kymberlee Price for reviewing early versions of this paper. Thanks to James Muir and Dave Walker for comments on earlier (published) revisions.

[8]That is, can be run in a reasonable time on reasonable hardware

References

[AG05] A. Acquisti and J. Grossklags. Privacy and rationality in individual decision making. *Security & Privacy Magazine, IEEE*, 3(1):26–33, 2005.

[asn88] CCITT Recommendation X.208: Specification of Abstract Syntax Notation One (ASN.1), 1988.

[BCL04] E. Bangerter, J. Camenisch, and A. Lysyanskaya. A cryptographic framework for the controlled release of certified data. *Twelfth International Workshop on Security Protocols*, 2004.

[Bra00] S.A. Brands. *Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy.* MIT Press, 2000.

[Cam] Kim Cameron, `http://www.identityblog.com/`.

[Cam06] K. Cameron. The Laws of Identity. *Online document], 2005 May 12*, 2006.

[CCH+04] S. Cantor, S. Carmody, K. Hazelton, W. Hoehn, T. Scavo, and I.D. Wasley. Shibboleth Architecture. *Protocols and Profiles, Working Draft*, 2:22, 2004.

[CH02] J. Camenisch and E. Van Herreweghen. Design and implementation of the idemix anonymous credential system, `citeseer.ist.psu.edu/camenisch02design.html`, 2002.

[Cha82] D. Chaum. Blind signatures for untraceable payments. *Advances in Cryptology: Proceedings of Crypto*, 82:23–25, 1982.

[Cha06] D. Chappel. Introducing Windows CardSpace. *Microsoft Corporation, Redmond, WA*, 2006.

[CKPM05] S. Cantor, I.J. Kemp, N.R. Philpott, and E. Maler. Assertions and Protocols for the OASIS Security Assertion Markup Language (SAML) V2. 0. *Committee Draft*, 4:14, 2005.

[cre] `http://www.credentica.com/`.

[FFS88] U. Feige, A. Fiat, and A. Shamir. Zero-knowledge proofs of identity. *Journal of Cryptology*, 1(2):77–94, 1988.

[FIP94] PUB FIPS. 186. *Digital Signature Standard*, 1994.

[GO94] O. Goldreich and Y. Oren. Definitions and properties of zero-knowledge proof systems. *Journal of Cryptology*, 7(1):1–32, 1994.

[HFPS99] R. Housley, W. Ford, W. Polk, and D. Solo. RFC2459: Internet X. 509 Public Key Infrastructure Certificate and CRL Profile. *Internet RFCs*, 1999.

[Lib] http://www.projectliberty.org/.

[NNR99] Moni Naor, Yael Naor, and Omer Reingold. Applied kid cryptography or how to convince your children you are not cheating. *Journal of Craptology*, 0(1), 1999.

[pri] PRIME - Privacy and Identity Management for Europe, https://www.prime-project.eu/.

[RR06] D. Recordon and D. Reed. OpenID 2.0: a platform for user-centric identity management. *Proceedings of the second ACM workshop on Digital identity management*, pages 11–16, 2006.

[RSA78] RL Rivest, A. Shamir, and L. Adleman. A Method for Obtaining Digital Signatures and Public-Key Cryptosystems. *Communications*, 1978.

[Shi] http://shibboleth.internet2.edu/.