Identities Evolve

Why federated identity is easier said than done

Stephen Wilson
Lockstep Consulting Pty Ltd, Australia
swilson@lockstep.com.au

AusCERT 2011 Conference: "Overexposed" May 2011, Gold Coast, Australia

Abstract

Why does digital identity turn out to be such a hard problem? People are social animals with deep seated intuitions and conventions around identity, but exercising our identities online has been hugely problematic. In response to cyber fraud and the password plague, there has been a near universal acceptance of the idea of Federated Identity. All federated identity models start with the intuitively appealing premise that if an individual has already been identified by one service provider, then that identification should be made available to other services, to save time, streamline registration, reduce costs, and open up new business channels. It's a potent mix of supposed benefits, and yet strangely unachievable. True, we can now enjoy the convenience of logging onto multiple blogs and social networks with an unverified Twitter account, but higher risk services like banking, e-health and e-government have steadfastly resisted federation, maintaining their own identifiers and sovereign registration processes.

This paper shows that federated identity is really a radical and deeply problematic departure from the way we do routine business. Federation undoes and complicates long standing business arrangements, exposing customers and service providers alike to new risks that existing contracts are unable to deal with. Identity federations tend to overlook that identities are proxies for relationships we have in different contexts. Business relationships don't easily "interoperate". They can't be arbitrarily tweaked to suit different contexts, because each relationship has evolved to fit a particular niche. While the term identity "ecosystem" is fashionable, genuine ecological thinking has been lacking in contemporary identity theory. The alternative presented here is to conserve business contexts and replicate existing trusted identities when we go from real world to digital, without massively re-engineering traditional business practices.

The password plague and 'token necklace' have elicited a sort of broad moral panic, yet they are essentially just human factors engineering problems. Traditional access control was devised for and by technicians; consumer authentication demands better user interfaces. The real problem lies not in identity issuance processes but rather in the way perfectly good identities once issued are taken 'naked' online where they're vulnerable to takeover and counterfeiting. If we focussed on conserving context and replicating existing real world identities in non-replayable forms, most routine transactions could take place safely online, without the incalculable cost of re-engineering proven business arrangements.

Federated identity: An unproven orthodoxy

The past decade is littered with earnest identity initiatives that failed to get off the ground and security industry consortia that over-promised and under-delivered. We've endured endless deconstructions of "trust" and theoretical dissertations on "identity" but none of this work has led to the sort of breakthrough that's desperately needed. Online identity fraud continues to grow. The direct cost is hundreds of billions of dollars globally; the indirect cost includes a malaise inhibiting such truly transformative initiatives as e-health.

In spite of its conspicuous failures and the revolving door of technical working groups, Federated Identity has become an orthodoxy. The US federal government's proposed National Strategy for Trust Identities in Cyberspace (NSTIC) takes federation as a given. Its central tenets, such as the pigeonholing of identification risk into four generic "levels of assurance", have been expeditiously standardised by numerous governments, and expressed in technical standards like SAML and OAuth, but not yet widely realised at more than the trivial Level 1. So we can now enjoy the convenience of logging onto multiple blogs and social sites with a Twitter, Facebook or LinkedIn account, but higher risk services like banking, government and healthcare stand apart, steadfastly maintaining their own identifiers and sovereign registration processes.

From time to time in this paper I'll look at social logons but the main concern is online business needing the higher levels of assurance. I contend that the great majority of economically important transactions elude federated authentication, and yet this is where the high priority problems lie, and where progress on digital identity has been wanting.

The federated identity movement is propelled by four forces:

1. In general terms, losses are mounting from cybercrime, a good deal of which is attributable to mis-authentication, including especially phishing, pharming, identity theft, ID takeover, and Card Not Present fraud.
2. We are suffering an ever worsening password plague and unwieldy 'token necklaces'.
3. We are burdened by the need to re-register at a great many sites to get the most basic online services.
4. Some players are energised by a vision of monetising their customer relationships, generating various forms of new business from allowing identities to be re-used by their customers in new settings.

Rolling too many objectives up into the one IT solution is so often a folly. In general, most federated identity models involve re-using one organisation's *authentication device* and *enrolment protocol* to support transactions with another organisation. Federation thus strives to kill two birds with one stone: improving security by giving users easier access to two factor authentication—which is a *technology problem*—and streamlining enrolment—which is a *business process problem*. One plank in my argument that federated identity is far harder than it first appears, is that its business process implications swamp the technology aspect, which is actually quite straightforward.

The roots of modern identity thinking

The conventional wisdom of identity management has been handed down by computer scientists, forged by the experience of 1970s era authentication. As modern cyberspace exploded from the late 1990s—and before the IT industry got around to improving much on password technology—we were besieged with earnest deconstructions of "trust" and theoretical dissertations on "identity", almost all written by engineers. The Comp Sci perspective has led to a sterile taxonomy and arbitrary formulations around identity that are reasonable to technicians but don't actually jive with business customs.

Let's look a little more deeply at one instance: the orthodox separation of "authentication" and "authorisation". IT security practitioners insist that establishing the identity of a Subject is strictly separate from defining what they can do. While these certainly can be seen as conceptually different questions, most infosec practitioners go further and insist that you cannot authorise anyone unless and until you have authenticated them. The primacy of authentication over authorisation has distorted the study of identity. The paradigm where authentication always comes first insinuates that we each have just one main identity. It's not a good starting point if we seek to maintain privacy as well as security online.

This perspective seems to have come from the world of network login and access management, where computer users typically have one static account name and password per system, and where a changeable Access Control List (ACL) governs what resources they can reach.

When authorisation is more important than authentication

Early in the development of PKI, we saw the authentication-authorisation divide enshrined in the idea that X.509 certificates asserted identity, and other mechanisms (such as ACLs, or secondary Attribute Certificates) would convey rights and privileges. For many years, commercial CAs concentrated on issuing general purpose identity certificates, and left it to Relying Parties to manage what any given Subject, once authenticated, was allowed to do. By presuming that the one identity certificate would suffice across many settings, this model put upward pressure on the quality of the identification process, resulting in unusual burdens on users. For instance, in Australia the early government operated healthcare CA required doctors to apply for certificates, presenting passport-equivalent photo IDs at a post office, a burden of proof they were unaccustomed to. Doctors rebelled, on the basis that if they were already bona fide medical practitioners known to the department of health, why should they jump through hoops in order to act as doctors on line? Healthcare PKI would have succeeded 10 years sooner if only they had issued dedicated certificates congruent with existing medical registration rules, instead of general purpose certificates with arbitrary new identification rules. One of the deep problems in early PKI was described thus by the IETF PKIX committee chair Stephen Kent :

"*For many big CAs, there is an assumption that a single certificate is all a user should need. This assumes that one identity is sufficient for all applications, which contradicts experience. For personal privacy and security, multiple independent certificates per user are preferable*" [1].

Instead of thinking of authentication and authorisation as orthogonal, an alternative view is that professionals like doctors are not really authenticated at all by their Relying Parties in the infosec sense of the word, but instead are accepted in business on the basis of their credentials alone. In most routine real world transactions, parties rely on authorisation and not identity. Consider that pharmacists dispensing prescriptions don't "know" (let alone "trust") the doctors. Likewise, investors don't "know" their financial advisers, nor a company's auditors; airline passengers don't "know" the airframe safety inspectors; bank customers don't "know" the tellers; employees don't "know" who signs their pay cheques. Yet the Subjects in these transactions are not total strangers to the Relying Parties; they each have a clearly defined identity in a certain context. So we can conclude that *identity-in-context is precisely the same thing as authorisation*. Throughout this paper, I shall seek to move away from the absolutist approach to authentication.

The foreign language of identity management

What does federation mean?

For something that has become a *de facto* standard and marketed so widely, it is curious that federated identity remains poorly defined. The NSTIC draft discussion paper states that "*an identity federation allows an organisation to accept and trust external users authenticated by a third party*" [2]. Similarly, a widely quoted description is that identity federation enables "*portability of identity information across otherwise autonomous security domains*".[1] Australia's National E-Authentication Framework (NEAF) uses the same wording [3]. The theme of portability is also central to the OAuth protocol, as we shall see below.

A less consistent theme in identity management is decentralisation. It is held by many that "*federated identity allows users to link elements of their identity between accounts without centrally storing all of their personal information*".[2] At one time, the early Liberty Alliance of identity vendors used that phrase on its homepage, yet at some point between 2007 and 2008, the wording vanished, replaced by a vision that was less specific and more oriented towards convenience: "*enable a networked world based on open standards where consumers, citizens, businesses and governments can more easily conduct online transactions while protecting the privacy and security of identity information*".[3]

Furthermore, the OECD is quite at odds with federated identity being decentralised. The OECD's Identity Management Primer for Policymakers states that "*with the 'federated' model, service providers do not aggregate their account information, but rather establish a central 'identity provider' that keeps track of which user identifiers correspond to the same user. In other words, federation links up previously unlinked identifiers*." [4]

---

[1] See e.g. http://www.infosectoday.com/Articles/Secure_Service-Oriented_Computing/Secure_Service-Oriented_Computing.htm (accessed 6 April 2011).

[2] See e.g. http://www.sun.com/software/products/identity/standards/liberty.xml (accessed 6 April 2011).

[3] http://www.projectliberty.org/liberty/about (accessed 4 April 2011).

Identity: both fuzzy and familiar

We shouldn't be surprised that federation is a slippery concept when even the word "identity" means different things to different people.[4] I believe it's futile quoting dictionary definitions (in fact, when a perfectly ordinary word attracts technical definition, it's a sure sign that misunderstanding is around the corner).

Instead of forcing precision on the term, we should actually respect its ambiguity! Consider that in life we are completely at ease with the complexity and nuance of identity. We understand the different flavours of personal identity, national identity and corporate identity. We talk intuitively about *identifying with* friends, family, communities, companies, sports teams, suburbs, cities, countries, flags, causes, fashions and styles. In multiculturalism we know about co-existing cultural identities; the idea of "multiple personality syndrome" makes perfect sense to lay people. Identity is not absolute, but instead dilates in time and space. Most of us know how it feels at a school re-union to no longer identify with the young person we once were. And it seems clear that we switch identities unconsciously, when for example we change from work garb to casual clothes, or wear our team's colours to a football match.

Yet when it comes to digital identity—that is, knowing and showing who we are online—we have made an embarrassing mess of it. Information technologists have taken it upon themselves to redefine the meaning of the word, while philosophically they don't even agree if we *should* possess one identity or more.[5]

While words are critically important, I do not propose any new definitions in this paper. I would simply ask readers to go with plain everyday language for describing digital identity. Let's think of identity as *how someone is known*. In life, people move in different *circles* and they often adopt different guises or identities in each of them. We have circles of colleagues, customers, fellow users, members, professionals, friends and so on—and we often have distinct identities in each of them. The old saw "don't mix business and pleasure" plainly shows we instinctively keep some of our circles apart. The more formal circles—which happen to be the ones of greatest interest in e-business—have procedures that govern how people join them. To be known in a circle of a bank's customers or a company's employees or a profession means that you've met some prescribed criteria, thus establishing a *relationship* with the circle.

---

[4] "Identity 2.0" doyen Dick Hardt for one acknowledged that he and others including Phil Windley and Bob Blakley are "disagreeing on what is identity". Hardt contends that identity is 'what I say about me, and what others say about me', and therefore equates—literally—identity with reputation (see his famous, beguiling and ultimately utopian presentation at the 2005 Open Source Convention at http://www.youtube.com/watch?v=RrpajcAgR1E). He believes that *any* personally identifiable information counts as identity data http://identity20.com/?p=47, presumably because he feels his identity to be a sort of singular synthesis of all those stories. Windley and Blakley object, insisting that transaction data and reputation be separated from identity. See the blogs at http://identity20.com/?p=47, http://www.windley.com/archives/2006/01/owning_identity.shtml, and http://notabob.blogspot.com/2006/01/on-absurdity-of-owning-ones-identity.html. Tellingly, this level of fundamental discord has not stopped various players from proceeding to productise their philosophies and to assume labels such as "Identity 2.0".

[5] http://michaelzimmer.org/2010/05/14/facebooks-zuckerberg-having-two-identities-for-yourself-is-an-example-of-a-lack-of-integrity (accessed 8 April 2011).

Kim Cameron's seminal *Laws of Identity* define a Digital Identity as "a set of claims made by one digital subject about itself or another digital subject" [5]. This is a *relativistic* definition; it stresses that context helps to grant meaning to any given identity. Cameron also recognised that this angle "does not jive with some widely held beliefs", especially the common presumption that all identities must be unique in any one setting. He stressed instead that uniqueness in a context might have featured in many early systems but it was not necessarily so in all contexts.

So a digital identity (or equivalently a numerical identifier) is essentially a *proxy* for one's identity in a given circle; it *represents* someone in that circle. Digital identity is a powerful abstraction that hides a host of formal complexities, like the identification protocol, and the terms & conditions for operating in a particular circle, fine tuned to the business environment. All modern identity thinking stresses that identity is *context dependent*; what this means in practical terms is that an identifier is usually meaningless outside its circle. For example, if we know that someone's "account number" is 56236741, it's probably meaningless without giving the bank/branch number as well (and that's assuming the number is a *bank* account).

I contend that plain everyday language illuminates some of the problems that have hampered progress. One of these is "interoperability", a term that has self-evidently good connotations but which passes without a lot of examination. What can it mean for identities to "interoperate" across contexts? People obviously belong to many circles at once, but the simple fact of membership of any one circle (say the set of chartered accountants in Australia) doesn't necessarily say anything about membership of another. That is to say, relationships don't "interoperate", and neither in general do identities.

The risks in metaphors

Identity, as understood as a "set of claims" in the *Laws of Identity*, is a *metaphor*. Intellectually this is a powerful formulation, and I have no essential objection to it, but unfortunately the word *identity* in day-to-day use is suggestive of a sort of magic property that can be taken out of one context and applied in another. So despite the careful framing of the *Laws*, many people still carry around a utopian idea of a singular digital identity based on a different metaphor: the *passport*. The tacit belief in the possibility of a universal digital passport has been a long standing distraction, and terribly unhelpful, for there is no such thing in the sense the word is used by technologists.

Ever since the early days of Big PKI, there has been the beguiling idea of an all purpose credential that will let its bearer into all manner of online services, and enable total strangers to "trust" one another online. Later Microsoft of course even named an early digital identity service "Passport", and the word is still commonplace in discussing authentication products. The idea is that the passport allows you to go wherever you like, yet the thing that the metaphor alludes to doesn't exist.

A real world passport simply doesn't let the holder into any country. To begin with, a passport is not always sufficient; you often need a visa. Then, you can't stay as long as you like in a foreign place; some countries won't let you in at all if you carry the passport of an unfriendly nation. You also need to complete a landing card and customs

declarations specific to your particular journey. And finally, when you've got to the end of the arrivals queue, you are still at the mercy of an immigration officer who usually has the discretion to turn you away based on any other evidence they may have to hand. As with business transactions, there is much more to border control than identity. So if we could create the universal digital identity, we should call it something other than "passport"!

Metaphors are more than wordplay; they are used to teach, and once learned, simplistic mental models like "electronic passport" can be deeply unhelpful. The dream of general purpose digital certificates is what derailed PKI, for they are unwieldy, involve unprecedented mechanisms for conferring open-ended "trust", and are very rarely useful on their own. Great time and money was squandered on the electronic passport goose-chase when all along PKI technology was better suited to closed communities of interest. What matters in most transactions is not personal identity but rather, credentials specific to the business context.

Yet with "open" federated identity frameworks, we're unwittingly repeating many of the missteps of early PKI, because people have been seduced by a metaphor based on something that doesn't exist. The well-initiated get that the *Laws of Identity* and worthy schemes like NSTIC all involve a plurality of identities tuned to different contexts. Yet NSTIC in particular is easily confused by many with a single new ID,[6] a misunderstanding actually exacerbated by the strategy's own champions when they use terms like "interoperable" without enough care, and casually imagine that a student in future will log in to their bank using their student card (see below).

The tenets of federated identity

Risks and "trust" can be categorised into discrete Levels of Assurance (LOA)

Surprisingly quickly, IDAM practitioners and many government policy makers have settled on a generic quaternary (four bin) categorisation of transaction risk and of quality of enrolment for an authentication device. The idea is to characterise the seriousness of a transaction in terms of LOA 1, 2, 3 or 4 and then match the LOA of the party you're planning to do business with. Quaternary LOAs are codified quite precisely in the US Government's e-authentication guide [6] and described somewhat loosely in the Australian National Electronic Assurance Framework (NEAF) [3] and the Kantara Identity Assurance Working Group [7].

It is likely that the idea of LOAs was inspired by risk management methodologies that matured through the '90s and '00s to result in such standards as ISO 31000. These involve gauging the severity and frequency of anticipated adverse events, and combining them to deduce a rolled-up risk rating for each event on an ordinal scale such as {*Negligible*, *Low*, *Medium*, *High*, *Extreme*}. A powerful feature of this approach is that each enterprise is empowered to create its own internal criteria, and to set its own policy for what level of

[6] http://www.washingtontimes.com/news/2011/jan/13/obamas-internet-passport and http://www.smh.com.au/technology/technology-news/no-country-for-cyber-outlaws-20110121-19zym.html (accessed 8 April 2011).

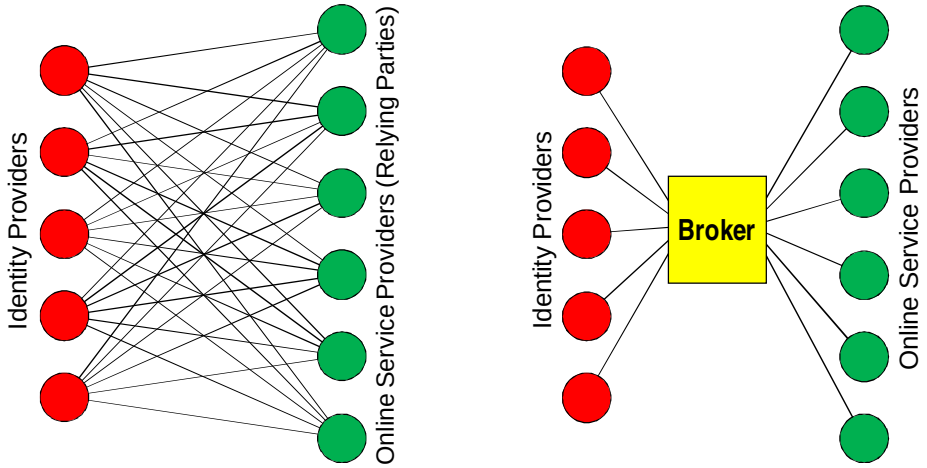
risk is acceptable for each anticipated threat. So some organisations will not tolerate residual risks that are higher than *Low*, while others will live with *Medium* risks on a case-by-case basis with special contingency plans. As a result, risk determinations made against ISO 31000 and the like *are not transferable between organisations*. Simply saying that a certain event—for example a compromised user account—has a risk rating of "Medium" tells someone outside the organisation nothing at all about the details of the threat, how it might be mitigated, its impacts, or even its expected likelihood. And yet the *authentication* Level of Assurance model would have us pick and choose externally issued identities based on a rolled up rating of LOA 1, 2, 3 or 4. There really cannot be any definitive assurance that all LOA X credentials issued by all IdPs are equivalent, nor that they will satisfy the detailed needs of all Relying Parties conducting LOA X transactions.

Identities should "interoperate".

It is expected that an identity issued in one context should be somehow reusable in other contexts. "Interoperability" is a famously slippery concept in information security;[7] you would think that by now we would have learned to use the term with more precision. Instead, interoperability of identity usually reflects an intuition that if I am known in one community then I should be more or less knowable in others. The idea is bolstered by the experience that we all seem to go through much the same registration process with new websites time after time. So we feel that the fact of registration with one service ought to be recognisable across others. But like so many intuitions, this one turns out to be wrong.

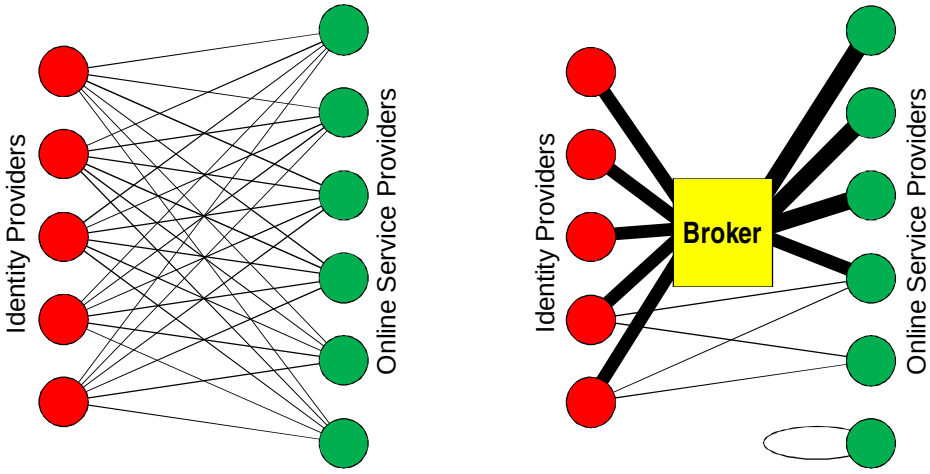
The business case for many *authentication brokers* is in fact just that: if identities one has established with identity providers are rendered reusable with multiple service providers, through the agency of a trusted intermediary, then the total number of enrolments will decrease, and time and effort will be saved. The benefit is typically illustrated by way of 'before' and 'after' schematics like these:

[7] The CEO of the Australian Payments Clearing Association reflected on their experience with PKI through the 1990s thus: "*Interoperability is something of a will-o'-the-wisp. You think you understand what people mean by it, and then quickly realise that you don't. In my experience, it's possible when discussing interoperability to be at cross-purposes for all of the time. Interoperability between members of the same PKI is axiomatic. Certificates issued by one bank should be recognizable by another. Interoperability becomes an issue when it is between different PKIs … But this still leaves the basic question of interoperable in respect of what?".* [8]



Each link in the graph joining service and identity providers represents a bilateral legal arrangement. Introducing the broker appears to reduce the total overhead. Yet the right hand scenario is only less expensive overall if the arrangements before and after are comparable—but they are not. An authentication broker's central promise is that identities will be useable for current and future online services over which the IdPs have no control and with Relying Parties with whom they have no relationship. This undertaking defies conventional risk management practices. The authentication broker has to construct *pro forma* contracts with IdPs and RPs that circumscribe risks of the unknown. When providers that want to join up engage their lawyers to review these unprecedented contracts, what will they make of them? For starters they will want to know how liability is to be managed if an error made by one IP can damage untold SPs.

Identity federation takes existing carefully crafted arrangements, in which businesses know their customers for the purposes of known applications, and breaks them open so that strangers with no prior relationship can also transact with those customers. The cost of having lawyers even come to grips with this situation, let alone negotiate novel contracts, is great and difficult to constrain. Therefore a fairer depiction of life before and after introducing an authentication broker looks more like this:



The multilateral arrangements between the broker and each player are far more complicated than any existing siloed bilateral relationships—as indicated by the heavier links—and the total cost of a small number of complex contracts is likely to exceed that of even a much larger number of traditional simple ones. And no authentication broker can ever be as universal as suggested by the idealised diagram; the one authentication broker can never meet the liability requirements of all players. So inevitably, some Service Providers will continue to deal with Identity Providers on their own terms; others will use a mix of federated identity arrangements and specific ones; and some will continue to issue identities for themselves.

Identity Providers and Relying Parties are separate roles

The Identity Metasystem teaches that identity provisioning and service provisioning (that is, identity processing) can be separated, and strongly implies that they *should* be. Organisations like banks are actively encouraged to re-imagine themselves as IdPs in their own right, and as Relying Parties, so that they can (a) act as IdPs to other institutions and so generate new business, and (b) rely upon externally issued identities so as to streamline customer origination. Yet both of these directions are radical departures from how banks work today. For a bank to have its identities accepted by other organisations necessitates some modification of those identities. Even the most minor change to the way a bank vets its customers, and to the detail of the customers' Ts & Cs, has complex legal ramifications. When it comes down to the detail of issuing identities for external use, banks struggle with managing misidentification related risks outside their own business silos. And the main problem with adopting identities issued by others is that Know Your Customer (KYC) regulations today generally demand that banks do their own identity proofing. And so the legal complexity swamps the benefits that might be gained from extra revenue streams and shrunken identity proofing.

To put that another way, while we all agree that banks are IdPs and governments are *examples* of IdPs, it is not logically the case that they can be the *same* IdPs.

Users need a choice of Identity Provider

The term "user centric" identity means more than designing authentication systems to be user friendly; it is taken by some to expressly mean that users should be able to pick and choose freely from a range of IdPs for the services they wish to access. This has turned out to be natural in the case of logging on to blogs and social media sites, where many different identities are often acceptable. However this is a near trivial use case,[8] all about convenience and attracting as much traffic as possible. The consequences of misidentifying a visitor are negligible. For more serious transactions like banking, government and healthcare, offering users a choice of authentication is a radical step. When so much transaction risk is borne by the Relying Party, it is complicated and indeed unprecedented for the user to have much say in the way they will present their credentials. To illuminate this problem, simply consider that some retail merchants

[8] And it has led to a fresh difficulty dubbed the "NASCAR problem" where site visitors are confronted by a profusion of colourful little logos for the various social identities they can exercise, similar to the tangle of advertising that bedecks American racing cars.

accept American Express cards while others don't, and there is precious little that an Amex customer can do to change that.

The Federated Identity experience

Successes

Most of the successful identity federations to date have been catalogued by the Kantara Initiative.[9] They include the Certipath Bridge CA for the aerospace industry, the SAFE Biopharma scheme for the pharmaceuticals sector, the Canadian federal B2C PKI, the US Federal Identity Credential and Access Management program (FICAM), and numerous tertiary education sector access control schemes (typically based on Shibboleth). Tellingly, these are all PKIs. More importantly from the legal and business perspectives, all of these federations are tightly bounded within a specific sector with common rules and standards set by an accepted central policy authority. Within each federation, all participants start out with credentials that are already highly congruent.

The major cross-sector success stories are found to Scandinavia. A good example, studied closely by the Australian banking sector, is Sweden's *BankID*[10] consortium of banks that have agreed on a particular PKI solution for issuing electronic credentials. Over two million BankID credentials have been issued, and may be used for lodging tax returns as well as many other government transactions. The interoperability of private sector BankIDs with government is underpinned by legislation, which in turn was enabled by Sweden's "long tradition of government relying on identities provided by other organisations such as the post office" [9]. After its own citizen e-ID card failed to gain ground, the Swedish government in effect outsourced administration of electronic credentials to the BankID consortium.

While I am focused here on higher value e-business identities, special mention must be given to the currently successful social IDs, especially Facebook Connect and LinkedIn Identity. These each operate almost seamlessly thanks to the OAuth protocol that permits services to share identity data, with the user's blanket consent. On its face, a Facebook identity is an unverified nickname, useful for logging on to sites that don't really care who you are. Yet the advent of the *social graph* adds another dimension to the synthesised verisimilitude of Facebook identities, which gain in stature as they are exercised over time across the community (even if this process is somewhat chaotic and blends reputation with identity in ways that are not yet clear). The LinkedIn Identity[11] shows promise of a stronger pedigree still because its users are much more likely to be using their real names, the social graph will be hardened by a more demanding professional community, and a good proportion of users are paying for the service which further enhances the identity verification.

---

[9] http://kantarainitiative.org/confluence/display/bctf/Implemented+Trust+Frameworks (1 April 2011).

[10] http://www.bankid.com (accessed 5 April 2011).

[11] http://info.gigya.com/LIIdentity.html (accessed 11 April 2011).

Failures

The Trust Centre

The Trust Centre was an initiative of the Australian banking sector, founded by Westpac in 2006, with the aim of combating phishing and identity theft. The plan was for an independently operated utility to provide identity management solutions to aid enrolment, authentication, authorisation and forensics.[12] In December 2006, Online Banking Review reported that seed funding was raised to establish the entity, and published an interview with the new COO Ben Forrest. Over the long term Forrest envisaged a "federated model of cross reliance that would issue a credential from one bank to another" which would enable a user interface "consistent across numerous platforms and experiences, for example banks, government, e-commerce, eBay etc." [10]. Others promoted the Trust Centre as "a hub for identity verification services in the Asia-Pacific region" [11]. At first each bank would continue to issue its own credentials and authentication devices. The immediate challenge was said to be getting the financial services industry to agree on the way forward in a collaborative environment [10].

After at least a year's hard work, in late 2007 the Trust Centre team broke up unceremoniously. The only public statement was released by Westpac which said "the vision still holds great promise, however the industry is not yet ready to embrace this vision to enable it to become a reality".[13] The reasons for the Trust Centre's failure have never been publicly accounted for satisfactorily. One of the four major banks had from the outset declined to be involved, because they argued that "security can deliver a competitive advantage". Reluctance on the part of one institution was extrapolated by some commentators to indicate discord across the whole group. I cannot contribute to the post mortem, but I observe that the challenge of even agreeing on a pathway towards federation was evidently too great for a small number of very similar organisations working in a highly regulated environment.

Internet Industry Association 2FA Framework

In 2005, the Internet Industry Association (IIA) developed a framework for an industry-based Two Factor Authentication scheme.[14] With the support of the Australian government, the IIA assembled a task force and commissioned detailed study of the legal risks, risk management strategies and IT architecture required to support the use of a range of authentication devices across multiple online service providers. The IIA's vision was to enable customers to re-use any "credible" authenticator at any participating e-merchant; the scheme would be technology agnostic, to increase choice, and to foster competition amongst authentication providers and solutions vendors. The objectives

---

[12] http://www.bankingreview.com.au/2006/11/banks-collaborate-to-protect-trust-in-online-banking.html (accessed 5 April 2011).

[13] See *Westpac exits the Trust Centre* 27 November 2007 http://www.finextra.com/news/fullstory.aspx?newsitemid=17782 (accessed 5 April 2011).

[14] Refer to the report by Patrick Gray of the news site "Risky Business", *Australia's neglected national 2FA scheme*, including an interview with IIA Chief Executive Peter Coroneos http://risky.biz/netcasts/risky-business/risky-business-119-australias-neglected-national-2fa-scheme 14 August 2009 (6 April 2011).

were to save cost, reduce the burden of the 'token necklace' on users, and improve overall authentication quality in the face of mounting cybercrime.

Sadly, having done the analysis and designed agreements and an architecture, the IIA task force was unable to sign up industry participants to trial the 2FA proposal. The organisation observed at the time that banks seemed reluctant to allow re-use of their authentication devices, and online services baulked at adding extra logon interfaces to their websites. In short, the IIA assessed that it was too early to implement shared authentication.

VANguard and AGOSP

"VANguard" is an Australian whole-of-government service built and provided by the Department of Innovation, Industry, Science & Research to support B2G transactions with user authentication, digital signature verification and time-stamping.[15] A Security Token Service (STS) and Single Sign-On service are still on the drawing board. VANguard at one time was more ambitious. Through c. 2006-07 VANguard management made presentations to local government and other organisations describing plans to build an "authentication brokerage service" that would verify transactions with agencies against third party credential issuers including commercial Certification Authorities; a schematic of the broker was presented at a software conference in 2008 [12]. There are no traces of this vision in the current VANguard offering. Even the STS is envisaged to be a closed service, with no mention of external tokens being accepted.

The Australian Government Online Service Point (AGOSP) is a sign-on service at the government's main portal *australia.gov.au* that switches visitors through to other agencies.[16] I believe that AGOSP at one time held similar ambitions as VANguard, namely that citizens who already have authentication credentials such as OTPs from their banks should be able to present them to a hub and after verification, be passed through to the government portal. AGOSP today has no links to private sector identity providers.

OpenID

OpenID became the poster child for federated identity. In launching NSTIC, Whitehouse security adviser Howard Schmidt in January 2011 blogged "*imagine that a student could get a digital credential from her cell phone provider and another one from her university and use either of them to log-in to her bank, her e-mail, her social networking site, and so on, all without having to remember dozens of passwords*".[17] This idea clones OpenID.

Support for OpenID was dropped by influential early adopter 37signals in January 2011. Unease with OpenID had actually been simmering for some time.[18]

---

[15] See http://vanguard.business.gov.au (accessed 6 April 2010).

[16] http://www.finance.gov.au/e-government/service-improvement-and-delivery/agosp.html (accessed 6 April 2011).

[17] http://www.whitehouse.gov/blog/2011/01/07/national-program-office-enhancing-online-trust-and-privacy (accessed 8 April 2011).

[18] http://blog.wekeroad.com/thoughts/open-id-is-a-party-that-happened (accessed April 8 2011).

Cardspace

In February 2011, in what was perhaps the most unsettling identity news in recent years, Microsoft declared that it was not releasing any further updates to Windows Cardspace.[19] It remains to be seen how information cards will be developed in the various open source communities and other identity initiatives. At the time of writing, two months on from the shock announcement, there has been no official response from Kantara, Higgins[20] nor the Information Card Foundation.[21]

A fair amount of analysis has been aired about the significance of the OpenID and Cardspace developments.[22] Reasons range across company politics, user interface design, and a lack of applications (which rather begs a question). Meanwhile "identity 2.0" continues to struggle; in March 2011 Dick Hardt announced that his password management tool *sxipper* was being discontinued.[23] Some commentators have resigned themselves to "the death of user centric identity" at least for the time being, with the rise of Facebook Connect signalling that the "identity wars" have been lost[24] and that some sort of identity monopoly is looming.

An alternative and unifying explanation recognises all these cases as variations on the basic theme of identity sharing, and concludes that something is amiss with the concept.

Analysis

It is often said that identity management is 'not a technology issue'. The statement is both right and wrong. The biggest challenges in federated identity are certainly not technological; rather, they relate to risk allocation in an unprecedented joined-up matrix which changes the legal fundamentals of how we do business. This challenge is of the model's own making. On the other hand, the pressing problems in digital identity management today really are technologically straightforward. We urgently need consistent, easier-to-use, more secure authentication methods.

If we take a closer look, we can see that nothing like federated identity has ever been done before. The proposition that banks, phone companies, universities and governments should act in the open as "Identity Providers" to support transactions they have nothing to do with is not something these institutions have ever seriously contemplated outside their own closed business contexts. Federation implies widespread changes to business rules and risk management arrangements, which lawyers and legislators have yet to come

---

[19] http://blogs.msdn.com/b/card/archive/2011/02/15/beyond-windows-cardspace.aspx (7 April 2011).

[20] http://eclipse.org/higgins (accessed 11 April 2011).

[21] www.informationcard.net (accessed 11 April 2011).

[22] See blog posts by Microsoft's Mike Jones http://self-issued.info/?p=458 and Kim Cameron http://www.identityblog.com/?p=1164, Craig Burton http://www.craigburton.com/?p=3128, Wired's Scott Gilbertson http://www.webmonkey.com/2011/01/openid-the-webs-most-successful-failure and Forrester's Eve Maler http://blogs.forrester.com/eve_maler/11-02-03-openid_successful_failures_and_new_federated_identity_options.

[23] http://dickhardt.org/2011/03/putting-sxipper-down (accessed 1 April 2011).

[24] See http://netmesh.info/jernst/big_picture/the-death-of-user-centric-identity-for-now (8 April 2011).

to grips with. Consider the banks' long established and highly regulated KYC protocols for identifying customers; introducing new third party identity providers and new enrolment pathways is a true paradigm shift, demanding untold revision of conventions, contracts and legislation.

The greatest challenge in federated identity is getting service and identity providers, accustomed to operating in their own silos, to accept risks incurred by their members doing business in foreign silos. This is where the term identity can detract from the reality that the "identities" issued by banks, government agencies, universities, phone companies, merchants, social networks and blog sites are really proxies for the arrangements to which members have signed up.

This is why identity is *so very* context dependent, and so why some identities are so hard to federate. The "identerati" have in the past got very close to the truth. In 2002, Darryl Greenwood observed:

> *All identity is 'local'. That local nature—the set-point in a closed and contextualized community—carries with it a tremendous amount of implicit information about the nature, scope and proper usage of the identity information. The further away from a pre-specified business context an identity credential becomes, the less valuable it is.* [13]

The intellectually compelling *Laws of Identity* speak of deep truths about digital identity and context, and they forcefully make the case for each of us exercising a plurality of identities, never just one. The *Laws* expose the abstract roles of Identity Provider and Relying Party hidden within what organisations like banks and governments do for their customers. Yet few if any of these sorts of institutions have been convinced by the *Laws* to expand these roles, mainly because nobody has yet worked out how to allocate liability in multilateral brokered identity arrangements, without re-writing the contracts that currently govern how we buy, bank and access government and health services.

So the problem with the *Laws* is not that they are wrong; it's that they are overly abstract. And the problem with the Identity Metasystem is that it imposes new intermediaries in business relationships that until now have been closely managed in a familiar bilateral way. A parallel example of over-engineering is seen in a particular new generation identity product *U-Prove*. Developed by Stefan Brands as an offshoot of his PhD dissertation and book *Rethinking PKI*,[25] U-Prove incorporates sophisticated new zero knowledge cryptographic algorithms to enable users to "prove unanticipated properties of protected identity assertions".[26] It's a brilliant idea, soundly implemented, but in reality, there are as yet very few mainstream e-business use cases that involve mutual strangers; as discussed below, most business is done on the basis on *anticipated* assertions.

It is something of a tragedy that the *Laws of Identity* have not yet encouraged many organisations to recast themselves in the identity marketplace. My personal experience is

---

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

[26] Much of the original U-Prove marketing materials have disappeared from the public domain in the past few years following Microsoft's acquisition of the technology in 2008. Some descriptions of the early SDK remain e.g. http://www.timberlinetechnologies.com/products/devkit.html (accessed 7 April 2011).

that when push comes to shove, banks in particular just don't see themselves as "identity providers". They are inherently conservative institutions, understandably rather easily put off by legal complexity that falls outside their main business activity.

Privacy in the federated identity environment

Privacy is held to be a headline benefit of federated identity, yet the sheer novelty of the Identity Metasystem in particular brings new risks to do with excessive aggregation of personal information, and potential instability. The NSTIC discussion paper states:

> *The Identity Ecosystem protects anonymous parties by keeping their identity a secret and sharing only the information necessary to complete the transaction. For example, the Identity Ecosystem allows an individual to provide age without releasing birth date, name, address, or other identifying data.* [2]

It is inherently good for privacy and security that NSTIC and identity federation in general deprecate any one master ID. Yet NSTIC introduces unusual intermediaries and new collections of personal information, the risks of which are not yet accounted for.

Any decentralisation in NSTIC is largely illusory: large volumes of personal information will be disclosed to Identity Providers, many of which will be start-ups. Left to their own devices, IdPs will accumulate extra data about the service providers to whom they have released each user's details, including anonymous attributes like age. These aggregations may make IdPs enormously valuable. As was the case with Big PKI a decade and a half ago, NSTIC's IdPs are likely to be start-up companies, or new business units of banks, existing business information brokers and the like. Even if these organisations are themselves scrupulous with privacy, there's the risk of hostile takeover leading to breaches and secondary exploitation.

With so many intermediaries joined into what have always been bilateral exchanges between customers and services, anonymity becomes a more technical matter in federated identity. So-called "verified anonymity" involves registering with third party providers, handing over personal information to them, only so that it may be *withheld* from Service Providers. It's more than ironic that in minimising disclosure between individual and Service Providers, the Identity Metasystem necessitates new disclosures of PI to IdPs.

Revisiting the identity security problem

By far the most economically important transactions on the Internet occur between parties that already have a local "metasystem" in place. Different sets of transactions—such as retail payments; business-to-business payments; healthcare orders, results, prescriptions and discharge summaries; accountancy; share trading; superannuation and funds management; e-invoicing; and all manner of private corporate intranets—are each undertaken within respective risk management and legal arrangements. In general these arrangements involve registration protocols, formal credentials, terms & conditions, and liability allocation.

The types of identifiers used to authorise all of these transactions are known in advance; they are *anticipated* identity assertions. Parties in each respective context know precisely where they stand, because they're participants in a scheme of some sort setting the context in which they operate. Everyone knows their roles & responsibilities *before* they transact, and indeed even before they've installed whatever application software and authentication devices as mandated by the service providers and concerned.

A great deal of identity fraud and cybercrime result directly from the takeover (or counterfeiting) of identifying information. These vulnerabilities relate to mechanical weaknesses in the way we present our sensitive data. They're technologically straightforward issues; they have nothing at all to do with "trust". We actually identify people well enough in the real world. Sometimes our identification breaks down, but not so often that the entire paradigm needs shifting. Instead, the authorities that oversee each transaction context are continuously monitoring fraud rates and other risks as well as performance, and tweaking all sorts of variables to optimise a mix of objectives peculiar to the business environment. Of special interest to us in this paper are the identification processes and parameters; as time goes by, these are steadily improved by financial regulators, universities, professional associations, employers and so on, in response to changing conditions and emerging threats.

Where did the preoccupation with trust came from?

One of the maxims of Internet security comes from Peter Steiner's famous New Yorker cartoon: *On the Internet, nobody knows you're a dog*.[27] Technologists especially have latched onto the saying and given it deeper meanings, particularly relating to trust. Through the rest of the 1990s and beyond, hoards of people became preoccupied with trust as a precondition for e-business. Untold effort has been spent researching, debating, deconstructing and redefining trust, as if the human race had never really understood the concept until Internet technologists came along with deep new insights. But there never really was a trust problem, and nothing on the Internet so far demands a rethink of identity.

We should read the Steiner cartoon as being about *fidelity* not trust. It goes without saying that nobody would trust a dog. The challenge in transacting with someone online is more prosaic: we must be able to tell faithfully what they represent, or what they are trying to assert.

There is an old Italian proverb that neatly sums up most business:

Fidarsi è bene; non fidarsi è meglio.

Or "To trust is good; not to trust is better". This would make a much better defining slogan of Internet sociology. In this light, the transition from real world to digital identity need not be so daunting, for trust is moot and technologists can stop fretting that the concept of identity needs re-defining. Instead, let us focus on taking the perfectly good IDs we have in the real world and taking them online in a smarter, safer form.

[27] See *The New Yorker*, 5 July 1993. Interestingly, the cartoon predates almost all e-commerce and cybercrime as we know them today.

Simplifying assumptions

We need to avoid complicated abstractions and generalisations about identity, and instead focus on simplifying assumptions.

Assumption: There aren't many strangers in real life business

The idea of "stranger-to-stranger" transactions is implicit in open identity theory, as it was in Big PKI, yet it's something of a false god. A review of the PKI accreditation market in 2002 highlighted an important disconnect: *"[During] the Internet boom there was a belief that e-business was going to release a massive pent-up demand to conduct stranger-to-stranger commerce. But truly un-vetted business introduction is rare*." [14]

Most e-business automates routine transactions between parties that already know—or know of—each other, in that they are subject to an over-arching set of arrangements, like a credit card agreement or supplier contract, or the legislation that governs a regulated sector such as healthcare. The first and foremost aim of most digital identities should be to faithfully represent existing real world credentials, allowing them to be exercised online without changing their meaning or their terms & conditions.

Assumption: There are no surprise credentials

As discussed, the novel identity product U-Prove has the objective of proving "unanticipated properties of protected identity assertions". That is, two strangers can use this solution to work out what they need to know about each other in real time before they transact. That's obviously very powerful but it's not what we really need right now.

Unanticipated identity assertions are quite academic. The vast majority of assertions in mainstream business are instead *anticipated*, and are completely worked out in advance of designing and implementing the transaction system. When you go shopping for instance, the merchant anticipates you will present a credit card number (so much so that they invest thousands of dollars in card processing infrastructure). When you log onto the corporate network, the relevant identity assertion is anticipated to be your employee number. When a doctor signs a prescription, the relevant assertion is their medical provider number, and pharmacists anticipate that number (after all, they can't read the typical doctor's signature!). Just think for a moment of the huge cost and tiny benefit of reengineering doctor-pharmacist arrangements so that some alternative unanticipated assertion could be presented in place of a provider number to authorise a prescription.

In almost all cases, the transaction context pre-defines what identity will be relevant, and we arrange ahead of time for the parties to be equipped with the right one. There may be interesting use cases where strangers can use U-Prove to strike up new relationships in cyberspace, but I simply argue that for most routine e-business today, the practical identity needs are more prosaic and more simply solved.

Assumption: Relying Party and "Identity Provider" are often the same

The central generalisation in the Identity Metasystem, and its progeny like the Open Identity Exchange (OIX) Framework [17], and NSTIC, is that Identity Providers are

separate from Service Providers. This may be perfectly true in the abstract, but it plays into the flawed intuition that the identity you have with one bank for instance should be readily recognisable by another. When you take an identity outside of its original context and try to make sense of it in other settings, you break its original Ts&Cs. Worse, you undercut any risk analysis that was done on the original issuance process. If a bank doesn't know how its customers are going to use their IDs, how can it manage its risks?

In reality, when the Relying Party is the Identity Provider, it retains closed-loop control over identification risk management and transaction risk management. This is the natural state of affairs in business and it does not yield easily to earnest efforts to 'break down the silos'. In many cases it will streamline digital identity (and minimise total cost of ownership) if we simply let certain Relying Parties continue to act as siloed Identity Providers.

The natural history of identities

Things are the way they are because they got that way.

Gerald Weinberg

The pressing problems of ID theft and fraud really are technologically straightforward, but we sorely need a better frame for understanding digital identity. We need to adjust the paradigm to promote better understanding that a plurality of identities is the natural state of being. First let's take a fresh look at the problem frame.

The term "ecosystem" has become fashionable in IT, as a sexy euphemism for "marketplace". With a politically correct ring to it, the word is seemingly used to lift the conversation above the hurly burley of competition and to attract more active government support (that is, stimulus). But those who like the term should heed the fact that the strongest ecosystems evolve naturally, rather than being designed. Truthfully, the ecosystem expressed in NSTIC for instance is an elaborate IT architecture with predefined often novel roles for all players. And at the time of writing, NSTIC was not even complete, with many anticipating that new legislation will be needed to allocate liability. Until the question of liability is resolved, nobody knows if the system is sustainable by the private sector, making the prefix "eco" seem a little premature.[28]

But what if we actually *thought ecologically* about the identity problem? A good starting point is the rich plurality of identities we already have in the real world. Where did they all come from?

The origin of identities[29]

If digital identity is a proxy for a relationship one has with a community of interest, then the natural explanation for the variety of identities is that we each exercise a variety of relationships. We have long lived with multiple connections, but cyberspace has

[28] It was recently announced that the complete NSTIC proposal will be launched on 15 April 2011.

[29] With apologies to Charles Darwin.

presented a dizzying array of new digital services each of which by default represents a new relationship and potentially a fresh identity. The enormous inflation of identities in the past decade was largely artificial. Most media and blog sites are inherently impersonal; the only reason their providers force us to register is so they can strike up a new commercial relationship of some sort. They know most users register under some duress and use false names to protect their privacy, but it's a numbers game. They hope to attract a proportion of bona fide registrations, and improve the figures by enhancing the relationship over time.

Social logon was a godsend. Ostensibly it offered faster, almost seamless re-registration based on one's existing account with Google, Facebook or Twitter.[30] Reducing friction would lift registration rates, and there was a fair chance registration fidelity would be improved if users employed their favourite and much-worked on social identity.

But what of our other more 'serious' identities, like bank accounts, credit cards, employee IDs and health identifiers? Are these too amenable to neat frictionless federation as Howard Schmidt imagines? Experience shows that breaking open the identity silos is harder than it looks, so we should look at the forces that have shaped all these identities before we tried taking them online.

Identities evolve

While the identity industry has appealed for a new ecosystem to be constructed, it seems oblivious to the existing ecology of business which has spawned specific arrangements for managing risk in different sectors and communities-of-interest.

Let us remember *why* there is almost always a formal protocol by which an individual joins a community-of-interest, whether it be a company, a professional association or a credit card scheme: it is to help the community manage its business risks. Some of these registration protocols are set freely by employers, merchants, associations and the like; others have a legislated element, in regulated industries like aviation, healthcare and finance. From case to case, the protocols are fine-tuned over time to cope with changing conditions. That is, they evolve.

The conventions, rules, professional charters, contracts, laws and regulations that govern how people do business in different contexts are examples of what are called *memes*; namely, heritable units of cultural transmission or of imitation [15]. The idea of memes was inspired by modern genetics and neo Darwinism. Over the past thirty years, a new scientific discipline has emerged that uses memes to explain the development of cultural phenomena.

I suggest we look at how identities may be built from memes and how they can have evolved in different contexts to minimise risk. Identity memes will have literally been passed on from one generation to the next. Consider for example the wide use of drivers licences to prove identity in retail transactions: this has the classic hallmarks of a meme.

---

[30] Or Microsoft Live, Yahoo, Amazon, Tumblr, Typepad, Posterous, Wordpress, Blogger, Bit.ly, Slashdot, Del.icio.us, Bebo, LinkedIn or MySpace, amongst others.

Roads & traffic authorities typically wish the situation had not arisen where permits to operate motor vehicles were coopted for general identification purposes. It is not clear that licence issuers ever officially sanctioned this practice, but it has obviously been widely mimicked across many different sectors, and slowly adapted and varied in many *ad hoc* ways. In contrast, other identity memes are more formally constructed and transmitted, often by legislated mandates like the Australian 100 point check for opening a bank account.

As business environments change, risk management rules in response change too. And so identity management processes and technologies are subject to *natural selection*. An ecological treatment of identity recognises that selection pressures act on all those elements. For instance, to deal with increasing money laundering and terrorist financing, many prudential regulators have tightened the requirements for account opening. To deal with ID theft and account takeover, banks have augmented their account numbers with Two Factor Authentication. The US government's PIV-I rules for employees and contractors were a response to Homeland Security Presidential Directive HSPD-12. Cell phone operators and airlines likewise now require extra proof of ID. Medical malpractice in various places has led hospitals to tighten their background checks on new staff.

By the same token, some environmental pressures act to actually weaken identity practices. For example, heightened privacy awareness is leading to some employers collecting less identifying information from new staff when they join up than they might otherwise prefer.

While the context dependency of digital identities is widely known, nevertheless it seems federated identity projects have repeatedly underestimated the strength of that dependence. The federated identity movement is fuelled by an optimism that we can change context for the IDs we have now, and still preserve some recognisable and reusable core identity, or alternatively create a new smaller set of IDs that will be useful for transacting with a superset of services. Such "interoperability" has only been demonstrated to date in near-trivial use cases like logging onto blog sites with unverified OpenIDs, Facebook or Twitter handles. More sophisticated re-use of serious identities across context has foundered; the Australian Trust Centre couldn't bring federation to life even amongst highly regulated banks working to the same identification protocols!

If we think ecologically, we can explain the surprising power of context. A better word for it may be *niche*. This term properly evokes the tight evolved fit between an identity and the setting in which it is meaningful. In most cases, if we want to understand the natural history of identities, we should look to the existing business ecosystem from where they came. As with real life ecology, characteristics that bestow fitness in one niche can work against the organism or digital identity in another. Thus the derided identity "silos" are a natural and inevitable consequence of how all business rules are matched to particular contexts. The environmental conditions that shaped the particular identities issued by banks, credit card companies, employers, governments and professional bodies are not fundamentally changed by the Internet. As such, we should expect that when these identities transition from real world to digital, their properties—especially their "interoperability" and liability arrangements—cannot change a great deal. It is only the pure cyber identities like blogger names, OSN handles and gaming avatars that are highly

malleable, because their environmental niches are not so specific. Noting how quickly social identities like Facebook Connect have spread far and wide, in a very real sense we can liken them to *weeds*.

Taking a digital identity (like a mobile phone account) out of its natural niche and hoping it will interoperate in another niche (like banking) can be compared to taking a salt water fish and dropping it into a fresh water tank. If NSTIC is an ecosystem, it is artificial. As such it may be as fragile as an exotic botanic garden or tropical aquarium. Full blown federated identity systems are going to need constant care and intervention to save them from collapse.

The way forward: Identity conservation

I am large. I contain multitudes.

Walt Whitman

All politics is local

Tip O'Neill

So, the real problem lies not in existing identity issuance processes; rather it is to do with the way perfectly good identities once issued are taken 'naked' online where they're vulnerable to takeover, counterfeiting and other misadventures. For the most part, there is no need to redo identity management. We all know that the hardest part of any digital transformation project is process reengineering, not technology, and so the reason so many digital identity schemes struggle should now be plain to see. If we focussed on conserving context and faithfully replicating existing real world identities in non-replayable forms, most routine transactions could take place safely online, without the incalculable cost of re-jigging mature business practices.

One of the most elegant and robust ways to render a digital identity non-replayable is to bind it by digital signature to relevant transactions. This is how the most secure modes of the Card Authentication Protocol work for paying by Chip-and-PIN card online; the customer inserts their card into a portable standalone reader, enters certain details of the payment together with their PIN, and a private key within the chip transforms the data into a unique and non-reversible code.

Fully fledged digital signatures using X.509 certificates and the like are not only non-replayable; they bind rich context information to the signed transactions too. This is perhaps the greatest untapped power of PKI. Public key certificates always include a "Policy Object Identifier" which points to a detailed specification of what each type of certificate is for, the conditions under which it was issued, the applications it is intended for and so on. In short, the Certificate Policy sets the context for the identity.

Context as such has been terribly bland in Big PKI. Historically, commercial CAs issued a limited range of general purpose certificates; the only really interesting aspects of the orthodox Certificate Policy was the Level of Assurance, the name of the issuing CA, the liability limits they usually imposed, and any warranty they were prepared to offer. Yet the

Policy can convey so much more. When special purpose digital certificates are issued in a closed community of interest, the Policy can map the precise relationship between the Subjects and the CA acting on behalf of the community. So for instance, one type of digital certificate might convey the fact that the Subject is an accredited surveyor in the state of New South Wales with a given licence number; another can represent that the Subject is a Director of a company with a certain securities commission registration number (note that the same individual might carry both of these certificate types, using one them to sign survey reports and the other to sign company returns). By convention, unique X.500 Object Identifiers for these different 'species' of certificate can be globally registered. Relying Party software in different contexts can easily be configured to look for the anticipated Policy that signals a party's authority to transact.

And so digital certificates provide the means for context to be conserved and unambiguously bound to identifiers. Some of my preceding practical investigations of contextualised digital certificates have led to working prototypes of anonymous e-health transactions [18] and anonymous e-voting [19] which further demonstrate the de-coupling of digital identities in different circles.

Further work

The rare successes in federated identity—from the recognition of bank issued IDs for e-government in Scandinavia, through to Facebook Connect—show that some digital identities do work well in multiple niches. On the other hand, Australian bank identities have resisted re-use even by other banks. If it were possible to examine any given digital identity and gauge its ecological agility, we could at the very least avoid yet more costly repeats of futile federation ventures. We could further learn how to optimise the interoperability of new identities, especially social IDs, as well as portable bank account numbers.

If digital identities do evolve as suggested here, then it should be possible to work out their *phylogeny*; that is, the natural history of the important features of an identity as they change over time. In different business ecosystems, different selection pressures will apply depending on the nature of the risk environment. Risk managers in recent years across different sectors have had to respond to developments such as high quality fake passports and similar breeder documents, magnetic stripe card skimming, Card Not Present fraud over the Internet, "identity theft" (according to various meanings of the term), tightening Anti-Money Laundering and Counter Terrorism Financing regulations, privacy laws, consumer power and greater demands on service quality, globalisation of financial services and the advent of pure-play virtual retail banks, mobile devices, virtual worlds and the financial risks that go with "Real Money Trades" [16], and the unpredictable influence of social media.

In turn, the heritable and variable (i.e. *mutable*) features of digital identities include the following:

— the particular facts about a Subject that are collected and checked at registration, such as name, date of birth, residential location, country of birth, education, qualifications, affiliations, test results, personal references, credit history, and so on;

— the way in which those facts are presented and verified at registration, such as in-person checking of photo IDs, or remotely with help of electronic verification services;

— other credential fulfilment security measures, such as in-person collection of new ID cards, password defaults, distribution of initial password by mailer and so on;

— the terms & conditions to which a Subject agrees to be bound in exercising their identity, such as sanctioned uses and proscribed uses;

— the lifetime of an identity, before renewal is required;

— the number of authentication factors used to demonstrate a Subject's connection with a presented identity (something you know, something you have and so on);

— the characteristics of a second factor physical device, such as one time password generation triggered by event or by time, tamper resistance of the device, PIN match-on-card, cryptographic key length, digital signature (to impress identity data upon a transaction to render it unique to the originator), cryptographic algorithm and so on;

— personal security policy, such as password strength and password rollover frequency.

One of the most notable recent changes in identity is the move away from in-person proofing especially in banking. Risk management is increasingly multi-dimensional, and identification at registration time is no longer the main lever by which service providers can combat fraud. The advent of better real time intelligence in transaction monitoring, and multiple authoritative data sources, has helped to enable purely online origination of new bank accounts in some jurisdictions; the traditional emphasis on bank tellers verifying static identification documents presented in person is giving way to dependable ways of establishing bona fides remotely.

All of these mutable features can be regarded as memes. In principle, a phylo*memetic* analysis should be possible to elucidate the stability and specificity of every aspect of a given digital identity. I suggest that an important next step would be to conduct a careful memetic study of representative identities taken from the wild, to uncover their roots in different business ecosystems, and the degree to which they have been adapted to suit particular environmental conditions. If the right set of features and matching memes can be worked out, then a robust phylomemetic family tree of digital identities could be derived. In turn, we could better understand when a given existing identity is amenable to reuse in next contexts, whether it might be "memetically" engineered to optimise its performance, or alternatively if it's best to breed a whole new identity to meet the e-business needs at hand.

Conclusions

I have presented an alternative theory to explain the under-performance of federated identity, as exemplified by such experiences as the Australian Trust Centre, Cardspace and OpenID. A crucial contention of mine is that the Identity Metasystem is over-engineered relative to the pressing problems of identity fraud and cybercrime today. The Identity Metasystem is a grand attempt to solve stranger-to-stranger "trust" and to allow parties to confirm one another's unanticipated identity assertions. Yet these are almost academic problems. The most economically important transactions on the Internet occur between parties that relate to one another according to extant *business* metasystems. All serious business—and therefore most e-business—takes place within overarching risk management and legal arrangements involving specific registration protocols, formal credentials, terms & conditions, and liability allocation. The analysis and design of all business systems anticipates the risks specific to their respective environments, and leads to appropriate identification protocols and participation rules. Parties in these different transaction contexts know precisely where they stand. They know their roles and responsibilities before they transact, even before they've installed whatever extra software and authentication devices that are required.

The price we pay for this crystalline certainty is that each of our different identities is brittle. Serious digital identities are highly context dependent, which is exactly what the *Laws of Identity* teach us. On the other hand, the utopian Identity Metasystem tries to teach us to bend those identities to suit other contexts. In practice, highly specific identities simply break when taken out of context, for their antecedent risk management arrangements cannot be easily modified. Identity federation takes the essentially technological problems of ease-of-use and security of digital identities, and inadvertently turns them into unprecedented legal and business process problems.

The fresh explanation for the resistance of most digital identities to federation is that identities have evolved to fit a special niche. Each one of the business identities we have—with banks, government agencies, employers, professional associations and so on—is really a proxy for the relationship we have in each context. These relationships have rules and conditions that have evolved over long periods of time in response to various risks peculiar to each case. Digital identities are therefore even more context dependent than the *Laws of Identity* imply. Different environments have bred the precise form and memetics of each identity, and it is no simple matter to take a stable form and expect it to work properly in any other niche.

If we appreciate identities as having evolved, then we should have more success with digital identity. It will become clearer which identities can federate easily, and which cannot, thanks to the way they have descended in real world business ecosystems.

About the author

Stephen Wilson is an independent analyst, consultant, commentator and inventor who specialises in digital identity and privacy. In 2004 he established the Lockstep Group to provide research and advice regarding authentication, and to develop innovative new cryptographic solutions to address identity theft and anonymity.

References

[1]. *Global PKI: Status, Trends and the Future*, Stephen Kent, Taipei International PKI Conference, Taipei, 2005 http://www.a-men.com.tw/a-men/dowload/01_Stephen%20Kent.pdf (accessed 8 April 2011).

[2]. *National Strategy for Trusted Identities in Cyberspace*, (DRAFT) June 2010 http://www.dhs.gov/xlibrary/assets/ns_tic.pdf (accessed 4 April 2011).

[3]. *National e-Authentication Framework*, Australian Government Information Management office, January 2009 http://www.finance.gov.au/e-government/security-and-authentication/docs/NeAFFramework.pdf (accessed 5 April 2011).

[4]. *The Role of Digital Identity Management in the Internet Economy: A Primer for Policy Makers* OECD Working Party on Information Security and Privacy, DSTI/ICCP/REG(2008)10/REV1, 2009.

[5]. *The Laws of Identity*, Kim Cameron, Microsoft, November 2005 http://www.identityblog.com/stories/2005/05/13/TheLawsOfIdentity.pdf (accessed 4 April 2011).

[6]. *Electronic Authentication Guideline* SP 800-63 v1.02, Bill Burr, Donna Dodson & Tim Polk, NIST NIST Special Publication 800-63, http://csrc.nist.gov/publications/nistpubs/800-63/SP800-63V1_0_2.pdf (accessed 5 April 2011).

[7]. *Identity Assurance Framework: Assurance Levels*, Britta Glade, Kantara Initiative Identity Assurance Work Group, 2009 http://kantarainitiative.org/confluence/download/attachments/38371432/Kantara+IAF-1200-Levels+of+Assurance.pdf (accessed 8 April 2011).

[8]. *Trust and Digital Certificates*, Peter Smith, Australian Payments Clearing Association, 6th Payment Systems International Conference (Belgium, 2000).

[9]. *Different countries – different paths: Extended comparison of the introduction of eIDs in eight European countries*, Herbert Kubicek & Torsten Noack, in Identity in the Information Society (*Special Issue: The Diversity of National E-IDs in Europe*), Vol 3, No. 1, pp 235-245, 2010.

[10]. *Trust Centre gains momentum* Charis Palmer, Online Banking Review Dec 2006.

[11]. *Westpac ID project fails to win support*, Julian Bajkowski, CIO New Zealand (online) 30 November 2007; http://tinyurl.com/trust-centre-fails (accessed 5 April 2011).

[12]. *Performance modelling for e-government Service Oriented Architectures* Paul Brebner, Liam O'Brien & Jon Gray, NICTA, paper presented to 19th Australian Software Engineering Conference, Perth, 26-28 March 2008 http://www.aswec2008.curtin.edu.au/IndustrySlide/Brebner%20on%20Performance%20(slides).pdf

[13]. *XKMS Study Point: Technical Standards in the Context of Business and Legal Architecture*, Darryl Greenwood, MIT, 1st Annual PKI Research Workshop, Gaithersburg 2002.

[14]. *Trust Services – A Market Appraisal*, Rohan Freeman, Mack Interact 2002 http://www.tscheme.org/library/tSi0156_01%20TSP%20market%20status%20report.pdf (accessed 8 April 2011).

[15]. *The Selfish Gene* Richard Dawkins, Oxford University Press, 1976.

[16]. *Virtual worlds; Real money*, European Network & Information Security Agency, 2009, available at www.enisa.europa.eu.

[17]. *The Open Identity Trust Framework (OITF) Model*, Mary Rundle (ed) March 2010. Open Identity Exchange http://openidentityexchange.org/sites/default/files/the-open-identity-trust-framework-model-2010-03.pdf (accessed 4 April 2011).

[18]. *A novel application of PKI smartcards to anonymise Health Identifiers*, Stephen Wilson, AusCERT2005 Refereed Academic Stream, 2005 http://conf.isi.qut.edu.au/auscert/proceedings/2005/wilson05novel.pdf.

[19]. *An easily validated security model for e-voting based on anonymous public key certificates*, Stephen Wilson, AusCERT2008 Refereed Academic Stream, 2008 http://conference.auscert.org.au/conf2008/Proceedings-SETMAPE.pdf.