Identity and Privacy

Have you chosen an Identity Provider lately?

June 2014

Summary

In this white paper, we look at the changing world of identity provision. Digital identity is evolving from a “retrospective” model to an increasingly predictive one, based on behavioural data as much as traditional credential. There is also a shift from siloed credentials towards more transferable assertions of identity and attributes. Potentially, these changes offer the individual more choice and power – but only potentially. Emerging identity models also contain hidden pitfalls, of which users need to be aware if they are to influence the market through effective exercise of choice.

Context

The world of digital identity continues to evolve at pace. The concept of digital identity itself now encompasses several different forms, principally the following three:

- traditional “retrospective” identities, where you go through a trusted enrollment process to receive, from a third party, a credential that you can present later on to authenticate yourself;
- low trust “self-asserted” identities, where the third party does little more than issue you with a syntactically-correct identifier which it will confirm when asked;
- “behavioural” identities, where service providers collect enough data about an individual to tell that the same person is visiting multiple times.

It's worth noting that the third kind of identification (behavioural) may require no explicit action on the part of the user. If a website sets a browser cookie, or makes a note of your IP address, that's enough to form the basis of a behavioural identity – though there are much more sophisticated ways, too.

Identity providers (IDPs) have also evolved, and continue to do so. The first IDP most users encounter is probably one of three kinds:

- the government (especially in the case of non-digital credentials like passports, driving licences and so on);
- an educational establishment (issuing a student ID for online access to student resources, libraries, networks, etc);
- an employer (issuing logins for email, business applications and so on).

To generalise/simplify slightly: all these IDPs issue credentials that are siloed. A credential issued by the government probably isn't usable for logging in to your employer's systems (unless they are one and the same, of course, but that's a special case). The student ID you used all through university probably won't work for your employer's systems, either. As I say, this is a simplification – federated identity schemes allow the gaps between silos to be bridged; but it's more common for federations to limit their scope to, say, the higher education context, or the government context, or single sign-on across commercial organisations[1].

Choice

Another notable feature about all three of those examples is that you, as the user, have little or no choice in the matter. If you enrol at a university or take a job, you probably have no option but to sign up for the authentication service offered to you, or face the prospect of trying to work without access to online resources (which, these days, is often simply impossible).

There are two ways in which choice enters the equation. First, with the emergence of schemes like OpenID and OAuth, users gained the option of self-asserting their ownership of a resource (such as an email address or an online account), and thereby, by implication, their identity.

Second, schemes such as Thawte's "Web of Trust" (now sadly discontinued) or the emerging UnitedID[2], seek to offer users a persistent credential associated with a self-asserted identifier (such as an email address or an authentication token). UnitedID is particularly interesting, in that it seeks to provide a trustworthy online identity to the mass-market consumer, without having to resort to the common, ad-funded commercial model of online service provision.

That credential may or may not be inherently trustworthy. Whether it is depends, initially, on how reliable the enrolment process is. If it's too easy for me to get a syntactically-valid credential that says I am Grace Kelly, the utility of the system is undermined. However, if the credential is sufficiently persistent, it may not matter that it appears to say that I am Grace Kelly: over time, a persistent identifier can 'accrete' trustworthiness in exactly the same way as a human being does, namely, by building up a record of consistently trustworthy behaviour.

But hold on: isn't there another category of IDPs that offer users choice and allow them to authenticate to many different service providers? In a way, yes. If you have ever accepted the offer to log in to service X using your Google, Facebook or Twitter ID (to name a few of the usual options), then you have "acquired" an IDP by default, without ever having consciously chosen one. In a way. It might be more accurate to say that you have opted to use an IDP that has been pre-selected for you. For the moment, I'm going to refer to these as "social IDPs". They have privacy implications you should carefully consider...

Implications

One of the choices I mentioned above was the self-asserted option (OpenID, OAuth, UnitedID and others), where the IDP is just that. Identity assertion (directly via credentials, or indirectly via implied access) is all it does. Schemes like this live or die by their ability to attract a critical mass of relying parties (RPs). If they can't attract enough, or can't attract RPs that are indispensable to the end user's online life, self-asserted schemes have a hard time adding value, and are likely to atrophy for lack of use. As a

[1] There are notable exceptions, such as the Scandinavian BankID system, which crosses the commercial/public sector divide, and federations in the defence/aerospace industry, where there is a high degree of interaction between government agencies and their contractors.

[2] United ID home page: http://unitedid.org/about/

side-note: even government schemes that are practically mandatory-to-use (such as the UK's authentication system for online tax returns) can suffer from this problem; an ID that you can only use in one place, once a year, adds little value, and is hard to remember and easy to ignore.

From that perspective, the big advantage of “social IDPs” is that they have what is, effectively, automatic critical mass, in both frequency of interaction and number of subscribers. It's quite possible that you interact with your social IDP more frequently than you interact with any other online service – even your work email. And a service like Facebook, which is reckoned to have around 750 million daily active users[3], offers RPs the prospect of access to (and by) a huge user group with one-click authentication. In one sense, this is the grail of IDPs. You aren't going there to authenticate: you're going there to do stuff, and authentication is a convenient side-effect. If visiting your IDP means you can also do stuff elsewhere, without having to re-authenticate, the convenience grows.

So, what's the down-side? In a word: panopticality. The ability of your IDP to keep track of everywhere you authenticate.

You may say that this isn't a new problem: panopticality was a criticism levelled at the first mature wave of federated IDPs[4], but there is a slight difference. In that wave of deployments, the IDP was part of a “circle of trust”, with pre-established contractual relationships between IDPs and RPs, and corresponding terms of service with users. The IDP's raison d'**ê**tre was the relationship of trust between itself and the user. If you trusted your IDP to verify your identity, you would probably trust them not to abuse your data.

But in the “social IDP” model, as we have established, authentication is essentially a side-effect. Their primary business model is the monetisation of personal data (through aggregation and re-sale to advertisers). It is in the social IDP's interest to collect as much data as possible about your online activities, and to monetize that data. It is also in the social IDP's interest to build as comprehensive picture as it can of your social graph. That includes subsets of your acquaintances that you might otherwise wish to segregate (for instance, work contacts and family/friends).

Consider this: if you have a Google+ ID that you use for personal online activity only, then by default, Google will only “see” your personal social graph. However, if your employer then decides to outsource calendaring to Google Calendar, and you have to authenticate using your personal Google+ ID, all of a sudden Google can draw the connection between your personal and work graphs. The result, for Google, is a richer, more comprehensive and more monetizable view of your social graph. The result, for you, is the erosion of a contextual boundary between your personal and work-related online data, and that is bad for personal privacy and online self-determination.

Remember, that contextual erosion is not something you explicitly signed up for – it's a side-effect of opting for one of the “social IDPs”. In fact, this social graph and context erosion is such an important factor that I prefer to refer to these as “social graph IDPs”.

This also reflects the fact that a social graph is extremely hard to fake. It is one of the most stable forms of behavioural identifier (and that brings us back to our opening

[3] http://expandedramblings.com/index.php/by-the-numbers-17-amazing-facebookstats/

[4] The most prominent examples being SAML-based federations designed to Liberty Alliance/OASIS specifications

contextual observation – that digital identity now includes behavioural identities as well as the more traditional kind).

Conclusions

First, 3rd-party imposed credentials are not going away. Governments will continue to need to issue credentials under their own control and for their own purposes (frontier controls, vehicle licensing and so on). Some of those credentials may be issued in forms that can be presented in commercial-sector authentication contexts, but the appetite for that still seems limited, even after a couple of decades of technical viability.

Second, users will continue to face a decision about whether to accept the convenience of "social graph IDPs", even if they become more conscious of the drawbacks of doing so, in terms of panopticality, privacy and self-determination. Remember: it is in the interests of the social graph IDP that you perceive only the convenience, and not the privacy down-side. That way, they get more data about you to monetize.

Third, in the authentication ecosystem of IDPs, RPs and users, there is a niche for self-asserted, persistent identifiers that allow an individual to (i) maintain control over the identity or persona they choose to assert, and (ii) build up a record of trustworthy behaviour, reliably associated with them rather than with anyone else. But that niche is a new and fragile one: it depends on RPs perceiving the value of those assertions and congregating around the IDPs in question.

This third model also offers the potential for an IDP service that is not a by-product of the monetization of personal data – but if that isn't the commercial model that sustains it, what will be?

cc BY SA