Achieving Privacy in a Federated Identity Management System

Susan Landau[1], Hubert Le Van Gong[1], and Robin Wilton[2]

[1] Sun Microsystems
[2] Future Identity
susan.landau@sun.com, hubert.levangong@sun.com,
futureidentity@fastmail.fm

Abstract. Federated identity management allows a user to efficiently authenticate and use identity information from data distributed across multiple domains. The sharing of data across domains blurs security boundaries and potentially creates privacy risks. We examine privacy risks and fundamental privacy protections of federated identity-management systems. The protections include minimal disclosure and providing PII only on a "need-to-know" basis. We then look at the Liberty Alliance system and analyze previous privacy critiques of that system. We show how law and policy provide privacy protections in federated identity-management systems, and that privacy threats are best handled using a combination of technology **and** law/policy tools.

Keywords: federated identity management, privacy, law, policy.

1 Introduction

In solving one problem, federated identity management raises others. For instance, federated identity management can simplify a user's experience through the use of single sign on (SSO) at multiple websites, thus enabling multiple services to be accessed as a unified whole and simplifying the complex process of managing user accounts after systems merge [16, pp. 16-17]. But by enabling the dynamic use of distributed identity information, federated identity management systems blur the divide between security domains, apparently creating a potential risk to privacy. We believe that separating the different contexts of a user's identity through federation creates *privacy in depth* and can substantially enhance user privacy. In this paper we demonstrate how federated systems can share a user's personally identifiable information (PII)[1] and yet *increase* her privacy.

In this introduction, we provide a brief description of federated identity management, then consider the context in which these protocols operate. In section 2

[1] PII is information that can be used to uniquely identify a person. What exactly constitutes PII — e.g., does an IP address do so? — is currently an issue of quite heated public debate.

R. Dingledine and P. Golle (Eds.): FC 2009, LNCS 5628, pp. 51–70, 2009.

we discuss where the privacy pressure points are in federated identity management and the different roles that policy, contracts, and technology have in protecting privacy within a federated identity management system. There have been various papers showing privacy "threats" against the Liberty Alliance system and proposing solutions. In section 3 we examine these and show that, by and large, the threats arise from a preoccupation with technical solutions where the real problems are social. We also show where the potential breaches presented in earlier papers fail (or fail to matter).

1.1 A Brief Introduction to Federated Identity Management

In federated identity management we have three actors:

- The Principal, or user, who has a particular digital identity;
- The Identity Provider (IdP), whose role is to authenticate the Principal once; the IdP then issues authentication assertions to a:
- Service Provider (SP), which provides services (e.g., access to protected resources) to authenticated Principals.

Sometimes the Principal and user agent (typically a browser) are considered as separate entities; in this paper, we will view them as one.

Authentication can occur in various ways: the SP can initiate an authentication request to the IdP the Principal designates when logged onto an SP, or the Principal can first authenticate at an IdP and then access an SP. In either case, the technology enables SSO, in which the IdP authenticates the Principal, thus allowing her access to protected resources at an SP.

There are currently three major variants of federated identity management — SAML (which underlies the Liberty protocols), InfoCard (which underlies Windows Cardspace), and OpenID — as well as some emergent efforts. The technologies were originally designed for different use cases: SAML and Liberty for business-to-business and business-to-consumer; CardSpace, a .NET component, for consumer-oriented activity; OpenID as a lightweight way to do user authentication (e.g., to reduce spam in blog comments). With time, design principles are converging somewhat. For example, OpenID is now being adapted to compensate for a weak security model. We discuss SAML and Liberty further in §3.1; for a more extensive comparison of SAML, InfoCard, and OpenID, see [16].

1.2 The Role of the Social Contract

A user may disclose information about herself for various reasons: she may give information about various accounts to a financial-services provider in order to do online banking; as a result of employment, she may be compelled by contract to reveal her driver's license number in order to use a company car; she may be required by the state to provide her height, weight, and eye color in order to obtain a driver's license. But while the compulsion in each of these examples is different, each of these systems works only if there is "consent of the governed" — a social contract if you will.

Users benefit from the convenience of single sign-on. They no longer have to remember multiple ways of authenticating at each site. Basic information about a user that she is willing to have at each site in the federation — such as work phone number, work address, company ID number — can just "be there," even while other pieces — marital status, ages and names of dependents, passport number — reside in separate SPs. Authentication at the related sites (e.g., the company online travel provider) is seamless. The user benefits from increased simplicity, while federation allows these distributed facts to be securely asserted.

Personal data, its safe storage and appropriate management, is the subject of legislation in many countries, and achieving compliance with data protection principles is seldom cost neutral. Federated systems enable Identity and Service Providers to benefit from increased efficiency. Some — those that choose to hold less PII — also benefit from decreased risk and compliance cost. A provider can't lose information it doesn't have. All providers benefit from the seamlessness of the networked interactions. IdPs and SPs will also find new business opportunities through these. The providers will also protect themselves through business contracts, technology, and through watching the bottom line. Organizations will only participate in federated identity-management systems if they see the benefit (or the cost in leaving).

If the user feels that her employer, or its delegate, the travel agency or health provider, has inappropriately shared PII, she has recourse to the legal system[2] (in extreme cases, she may choose to leave the company). If the user feels that the government is inappropriately sharing her PII, then, depending on the norms within the state, she may choose not to involve herself in the government identity-management system[3]. If the IdP or SP starts sharing user PII in ways that the user objects to, the contract — implicit or not — must be negotiated.

A social contract is a balance: I give you this, I receive something else in return. If the Principal doesn't find value in the system, if the IdP or SP doesn't realize benefits of efficiency or reduced risk, the social contract fails, the system loses participation, and the identity-management system fails. New Zealand provides a good example. New Zealanders have a strong resistance to "Big Brother," so when their government embarked upon an e-government strategy to provide online delivery of citizen services, the focus was on user privacy and security. The architecture was designed accordingly, with the Identity Verification Service uniquely identifying an individual but forwarding only minimal identity attributes to the Government Logon Service [17, p. 53].

[2] Of course, there are instances, such as during law-enforcement investigations, when the question of the Principal's satisfaction is likely to be overridden.

[3] McKenzie et al. observe that social norms vary greatly across the world [17]. New Zealand's identity management system emphasizes user privacy and security, while Scandinavian countries focus on government transparency. The patchwork of U.S. privacy laws can make one wonder if the privacy of video rental data really deserves the same legislative attention as that of banking transactions. In the U.K., a range of public sector policies are being predicated on government data-sharing on a massive scale.

New Zealand's effort points out that all parties must benefit in order for them to participate — and for the system to be viable. Clearly technology is only one part of a complex solution that includes:

- Technology: e.g., mechanisms such as SAML, CardSpace, OpenID
- Business contracts (when applicable, e.g., not always used in OpenID) which, in turn, are part of...
- The legal, regulatory, commercial and technical implementation factors.

What's more, as the legal and regulatory context will change from country to country, a single, uniform technical approach cannot suffice.

Federated ID management systems need the buy-in of all three sets of members: users, identity providers, and service providers. Users are discouraged if participating in the system obviously leads to abuse of their personal data. Identity providers will typically bear liability for the assertions they make, and service providers will seek to reduce their own risk by relying on those assertions. Thus any analysis of the privacy risks of an identity-management system must work by looking at the deployed system holistically.

In fact, the idea of a social contract between the user and other parties reflects a broader principle, namely that each party to a federation must have some sustainable motivation for taking part, and for behaving well in the structure. How that works may well vary from one federation to another — particularly between the public and commercial sectors, where the incentives and penalties can differ enormously. The following table suggests some of the different levels at which there can be an incentive to "behave well" within a federation:

Privacy Driver	Incentive
Best Practices	Improve User Trust
Industry Code of Conduct	Industry Sanctions
Legal and/or regulatory controls	Avoid prosecution and/or liability

2 The Privacy Drivers in Federated Identity Management

When federated identity-management systems were first introduced, there was substantial concern about potential privacy invasiveness. We believe this stemmed from an oversimplified view of the technology. While identity management systems are about sharing information, it is naïve to assume they can succeed if they do so indiscriminately and without regard to context; and this is where user privacy can be enhanced.

As the table illustrates, motivations for good behavior in a federation can be many and varied, and may differ between regulatory contexts. However, there are approaches to the privacy problem that are generally applicable across regulatory contexts; one such is an analysis based on risk. For example, one might identify the following sources of risk in a federated system:

- risk of data disclosure (inappropriate, excessive, without consent);
- risk of metadata disclosure (making it possible to link other pieces of personal data relating to the user, usage patterns, inferred habits and preferences);
- regulatory exposure (if you don't have the data, you don't have to worry about failures of compliance . . .).

These can also provide useful input to processes such as Privacy Impact Assessments, e.g., along the lines already established by the Ontario Privacy Commission [18]. Taken together, approaches based on the "privacy drivers" and the sources of risk shown above offer several ways in which the sharing of data can be analysed, classified, segmented, and strategies devised for managing it appropriately in the context in question.

Consider the following example:

- I have an account with American Express. They have whatever PII is required by law. Most importantly, they know I have an account with them.
- I have an account with UPS (and USPS, Fedex, etc.) They have PII (or access to it) that includes my address.
- I have an account at Privatzon, a book seller. They have no PII — other than purchase history.
- I have a Liberty-enabled Discovery Service from my Liberty-enabled Identity Provider.

When I want to buy a book at Privatzon, I login, select the book and request that Privatzon use my preferred payment and shipping services. Privatzon contacts my discovery service requesting my preferred shipper and payment services, and is told they are, say, UPS and American Express. Separating these pieces of data, for which a federated identity-management system is ideally positioned, protects an individual's privacy. Privatzon contacts the two SPs, which respond, one with a one-time credit card[4] and the other with a one-time shipping label — both are just numbers and/or barcodes. PII remains with those SPs that "have a need to know" and isn't otherwise distributed. Thus American Express has no idea UPS was used as a shipping company and inversely UPS does not know about my account at American Express. None of these companies knows about my book-buying habits either. Indeed, a properly designed federated identity-management system can keep these contexts separate, providing individuals with far more privacy protection than they currently have.

While Privatzon wants to know about my preferences for purchases — including what I buy for my sister, my uncle, and myself — so that it can better predict what products to show me as I browse, Privatzon does not need know where I live or what my credit card number is. It only needs to have authorization from my preferred payment service for the charge and to let my preferred shipping service know to whom ("Uncle Tim") the purchase is being sent. Thus Privatzon keeps the information about my purchase preferences, UPS keeps the

[4] There are, of course, many solutions using one-time credit numbers, starting with [22].

addresses of my friends and family, while American Express keeps my credit-card information. Both Privatzon and UPS have less information about me that could be inappropriately accessed or disclosed, and thus limit their legal liability and their compliance burden. If I work from home, I can have my work mail — copies of *IEEE Security and Privacy, CACM, Science* — delivered to me without listing my home address in the membership database (and thus publishing my home address). Instead of IEEE, ACM, and AAAS each keeping track of two addresses for me: my home (for delivery of journals) and my work location (for publication of membership lists), the organizations tell the mail provider SP that my journals should go to my_home_address. Meanwhile the organizations print my_work_address in their membership listings. The federated system keeps these separate pieces of my PII separate — and private.

2.1 The Principle of Minimal Disclosure

In the end, if Privatzon is to get paid, if the book is to be shipped, if the pharmacist is to fill a prescription, then the data that is my credit-card number, my uncle's address, and my drug dosage has to reside somewhere. Insiders have always posed the greatest security and privacy threat. Identity-management systems should use the principle of minimal disclosure, and should be able to engage where no PII is exchanged. Federation allows information to be distributed with each SP receiving exactly the information needed for its role — though many service providers may have to adjust to the concept (since they will no longer receive PII). To reduce liability, many organizations will choose to limit the PII they hold (and then protect the PII they hold in various ways: protected databases, strict access rules, careful auditing procedures, as well as some PETs, including those described below). Federated systems allow them to do so, and there have been several approaches to this — both theoretical and within deployed systems.

In 2007 Gevers et al. presented a privacy-preserving method for supplying disparate pieces of information to an SP, e.g., revealing that the user is over eighteen and a citizen of < Belgium, France, etc. > without revealing the user's age or citizenship [4]. Their solution is a "claim evaluator" sitting within an Identity Provider that responds to such queries. Sun built such a system in 2005 to satisfy U.S. government requirements about employee contributions to a political fund. Sun's system checked employee citizenship, employee rank, and stockholder status; based on the information gathered, it returned a yes/no eligibility status for company political (PAC) contributions [25]. This conceptually simple approach is readily understood by the user.

Researchers at IBM Zurich have developed a system, Idemix, that uses zero-knowledge credentials to protect user privacy. A Principal presents an encrypted pseudonym to an SP along with credentials that the Principal uses to prove to the SP that it is the owner of the pseudonym [6]. The unlinkability inherent in Idemix means that the IdP will not be able to cancel unused assertions, something unlikely to be attractive to IdPs or SPs. And if Principal "shares" the pseudonym and credentials with another user, the new user is able to impersonate the Principal *everywhere,* rendering such sharing highly unattractive.

It is important to understand when such minimal disclosure is needed. The IBM researchers present a usage case of a Principal seeking to rent a car online and presenting a credential that shows the Principal possesses a valid credit card and driver's license. But while there is no reason for the rental agency to know more about the user's finances than that she has adequate credit coverage for the car rental, there are numerous reasons why the agency should have information about the driver. If a prior interaction gave rise to a dispute (such as unpaid charges or collision damages), the rental agency wants to know about this before agreeing to lease a car to the Principal. It may be sufficient to be able to link past and present records without necessarily using the Principal's name as the link. Observe that the SP wishes to be able to establish the link even if it is not in the Principal's interests for that to happen. However, when a Principal succeeds in concealing the link between the current interaction and previous detrimental behavior, the ultimate protection (for the Service Provider) may be a legal one, rather than a technical one. In other words, if the Service Provider can subsequently prove — for example, through physical forensics — that the Principal acted deceitfully, the mitigation may be that the SP declines to accept liability for the bad behavior of the Principal.

There are situations in which such "cloaked" interactions between an SP and a Principal as provided by Idemix are appropriate, and situations in which cloaking is inappropriate. Often an identity aware approach is more appropriate and use cases should be carefully analyzed to determine which approach fits the situation best. We should also keep in mind that Idemix' zero-knowledge protocols come at a cost: the running time goes up by a factor of approximately five over "normal" identity-management schemes [6].

2.2 Linkages Only When Appropriate

Where more than two SPs are involved, the question of linkability still arises. Again, the mitigation (this time on the Principal's part), might be technical (such as the use of different identifiers for each SP or the opaque identifiers that Liberty automatically provides) or not (such as a reliance on Data Protection principles restricting "purpose of use" to "purpose of collection").

The canonical use case for preventing the IdP from being able to correlate a Principal with her choice of SPs is one often cited by Dick Hardt [5]: presenting a driver's license to a bartender to prove the Principal is over the legal drinking age without the Motor Vehicle Bureau knowing the particular instance of the use of the credential. This use case governs acceptance of e-government identity-management systems. Citizens are uncomfortable if, in using government-based digital credentials to establish certain aspects of their identity in private-sector authorization, they are exposing private activities.

Various zero-knowledge-based specifications, including Idemix and UProve[24], solve this problem. An alternative solution is provided by Microsoft's InfoCard protocol, which defines a message flow eliminating direct communication between the IdP and the SP; the protocol also offers the option of having the identity selector encrypt the SP's identity so as to prevent the IdP from learning

it on receiving a request for a token. Both features together are necessary to ensure that an IdP cannot learn which SPs a Principal visits. The SAML Enhanced Client/Proxy protocol is similar but at present only has the first feature (it could be further profiled to add the second feature).

The question is whether deployers of federated identity systems want such a feature. Although these other identity-management schemes show how to solve the problem of keeping the SP information from the IdP, this may be a problem in search of a solution. The bar-and-driver's-license example is an "off-label" use of the credential — and that aspect is actually the significant aspect of the situation. There is no reason for the Motor Vehicle Bureau to know the Principal has been at a bar, and for this reason we find the IdP's knowing about the "off-label" license usage problematic. On the other hand, the agency has every right to know if a policeman has issued her a ticket for speeding.

Trying to use a non-governmental ID to prove certain properties about the user may not be any better. For example, banks do not underwrite clients' use of bank-based credentials for activities that the banks don't know about[5]. There is a tug-of-war around conflicting needs. A low-trust proof of identity (such as one supplied by OpenID) provides (at risk of coining a truism) a low-quality proof of identity — analogous to a bar accepting hand-written notes that its patrons are over 18. If the user needs a high-quality proof of identity then the identity supplier will likely want some type of control over how that credential is used.

inCommon [7], a federated identity-management program that helps universities share resources, provides an excellent illustration of this type of tradeoff. inCommon uses the SAML-based Shibboleth environment to create a federation whose members are institutions of higher education. Each IdP determines their own authentication mechanism while each SP determines what their access policies are. To obtain resources at an SP, a Principal uses an IdP-issued credential that only says that the Principal is a member of the IdP's institution. Several things make this system work. At the technical level, it is the SAML specifications. At the policy level, Shibboleth depends on trust. At the business level, this is an instance of a low-level access. The Principal doesn't get the keys to the kingdom as a result of their assertion, only access to some online resources at a member institution. Access to more highly valued resources at that institution (e.g., registrar records for cross-registration purposes) would require authorization predicated on attributes other than "is a member of."

2.3 Lowering Exposure

In an analysis of the market failure of P3P, Jane Winn notes that "voluntary, consensus standard-starting processes are more likely to succeed when responding

[5] Though the Scandinavian Bank-ID system shows that this divide can be bridged ... under the right conditions. Under the Bank-ID system, bank-issued credentials can be used to authenticate for access to public-sector services. There is an element of technical interoperability, of course, but just as vital is the legal provision for public-sector bodies to rely on credentials issued for a quite unrelated purpose.

to market demand" [26, p. 16]. We believe that the market driver for privacy in federated identity management systems is the liability threat.

If PII is inappropriately released from an identity-management system, then the organization that allowed this to happen will be liable. The privacy pressure point is the organization with the data, giving the organization strong incentives to protect the PII it has — and *to minimize the amount it collects.* In 2003, California enacted breach notification laws requiring notification if unencrypted data has been inappropriately disclosed[6]; many other states have followed suit.

Both IDPs and SPs have to assess whether the personal data they hold exposes them to disproportionate liability or an excessive burden of compliance. However, in such areas as preventing money laundering, working with vulnerable adults and children, etc., for the IDPs or SPs to hold too little personal data can also be a liability.

2.4 Technical versus Legal/Policy Approaches to Privacy

Technological privacy protections are not only necessary to digital identity management, they are fundamental to its existence. Without encryption (symmetric and public-key), digital signatures and other relevant technologies, there would be no digital identity management, federated or otherwise. Yet even where part of the solution to a given privacy problem is technical (e.g.,[22] [23]), legal and policy protections play critical roles.

Take, for instance, the deletion of PII that is no longer needed. There are several technical options here, ranging from destroying the data, or using strong encryption to protect the data and then destroying the keys on a regular schedule, to more complex privacy-enhancing methods. Which solution is most appropriate depends on non-technical factors such as the value of the data, the risk involved in storing it, and in some cases legal requirements.

We do not argue against the technical solutions — indeed, we expect over time to see various of them adopted — but contend that the privacy solutions for identity management should be a combination of technical and non-technical measures, capable of adjusting to different legal, regulatory and liability contexts. We now discuss the Liberty Alliance model, a solution that relies on both.

3 Liberty's Approach

In 2001 issues of online identity and how to simplify it began appearing. The Liberty Alliance, an industry group consisting of technology companies as well as companies concerned with online identity, was formed to develop open specifications for federated identity management. The SAML/Liberty model is now quite mature; it has been deployed in a number of environments, including governments (e.g., New Zealand[17]), healthcare (e.g., Aetna) financial (e.g., Citi) and communication companies (e.g., Deutsche Telekom AG). We continue our

[6] SB 1386 covers breaches of financial records, AB 1298, breaches of medical and health information.

discussion of the technology (further detailed in Appendix A), following that with a discussion of privacy protections in the Liberty system.

3.1 Liberty Protocols: A Brief Overview

Traditionally, users have needed separate accounts with passwords etc. at each online Service Provider with which they wanted to interact. The *Liberty Identity Federation Framework* (ID-FF) sought to solve that problem by enabling a Principal to establish pairwise federations with a single account at the Identity Provider (IdP). This made Single Sign-On possible within a *Circle of Trust* (CoT): a set of Service Providers and one or more Identity Providers.

ID-FF defined protocols for account linking, single sign-on and global logout, and to accommodate different mechanisms (e.g., browser with redirects, artifacts, etc.). Liberty's ID-FF and OASIS SAML were distinct efforts in developing federated identity management for secure simple sign-on which converged with OASIS's SAML 2.0 specification, the de-facto standard in federated identity management. SAML 2.0 represents a significant improvement in a number of aspects including defederation, metadata definition, and authentication levels.

Liberty's Identity Web Services Framework (ID-WSF) uses federated identity (ID-FF or SAML) as a basis for a web services framework that enables web service consumers (WSC) to invoke services web service providers (WSP). ID-WSF has three key goals: (i) define safe protocols and practices to protect a user's online privacy; (ii) leverage the user's identity to enable personalized services; (iii) allow secure identity-based sharing of resources by a variety of services.

Before a service can be invoked by other web services, it needs to be discoverable and associated with the Principal's Discovery Service (DS). ID-WSF defines a set of protocols for the entire lifecycle of online services, including support for 3 core operations: (i) registration/association; (ii) discovery; and (iii) invocation.

The main reason for registering a service instance at the DS and associating it with a Principal is to allow other web services to discover it. This discovery is the result of a lookup query sent by a requesting WSC to the DS. That request specifies the Principal's identity as well as search parameters like service type.

In addition to a list of services (WSPs) that matched the lookup request, the DS also includes security tokens that will allow the WSC to invoke the service hosted by the WSP(s).

Liberty and SAML use a variety of security technologies: channel security (server-side certificates for identity providers; TLS1.0, SSL 3.0, or other channel security protocols, such as IPSec, with appropriately certified keys), security tokens with a limited lifetime, nonces, and digital signatures for per-transaction integrity. In addition to those security mechanisms, Liberty identifies specific elements of its protocol messages that may present privacy risks (due to the nature of the information they bear) and might need additional security. For instance, in the most privacy-aware environments, it is recommended to employ encrypted Name Identifiers. Also an SP can use the NameID Mapping protocol to add a pseudonym of its choice to an existing federation.

3.2 The Roles of Liberty's Privacy Guardians

The Liberty protocols address only the *exchange* of information. Whatever the parties do to produce or consume the information that was exchanged is outside the scope of these protocols. A Liberty-enabled exchange of information between an IdP and SP starts with a contract specifying partner responsibilities. When engineers discuss the Liberty protocols this step is often omitted; contracts are not omitted in practice. Thus our analysis starts with the *Liberty Privacy and Security Best Practices,* which lays out expectations on handling PII:

> The framework of the Liberty Specifications is built upon the presumption that PII will be shared ("attribute sharing") only in the context of permissions i.e., in line with the Principal's expressed consent and preferences. Such attribute sharing should be predicated upon not only a prior agreement between the Liberty-enabled providers, but also on providing notice to the Principal and obtaining the Principal's consent ... Liberty-enabled providers should take reasonable measures to prevent unauthorized acquisition of a principal's personal information (e.g., by harvesting)[11, p. 9].

Specifically:

- The Identity Provider should safeguard the Principal's credentials and should have some mechanisms in place to require the Authentication Domain to use the credentials in a proper manner.
- Service Providers should inform Principals of their data practices, provide Principals with certain choices regarding secondary uses of the Principal's PII, maintain security of a Principal's PII within their control, and not use or share such information except in accordance with the Service Provider's privacy policy and/or the consent or usage directives of the Principal.
- The Attribute Provider[7] has at least the same responsibilities as Service Providers with respect to clear notice (including notice to the Principal regarding what are the default usage directives and how the Principal can change such usage directives), choice, security, and responsible use and sharing of the Principal's data[11, pp. 9-10].

For PII passing between entities, privacy is achieved through security. Channel security is provided by TLS; protection against replay and main-in-the-middle attacks is provided by digital signing, as is message integrity; correlation of a Principal's PII by multiple SPs is prevented through the use of opaque identifiers (this does not protect against a timing analysis or traffic flow attack).

PII 'at rest' at an entity should be protected through standard mechanisms for data confidentiality and integrity, with audit trails to track data access, etc. Ultimately, though, the protection here is legal. A rogue Service Provider or

[7] The Attribute Provider earns its place in the ecosystem purely by serving up attributes on behalf of the user, as opposed to by providing an actual service (such as car rental, payment transactions etc.).

Identity Provider is in a position to violate a Principal's privacy and technical protections can only reduce, not eliminate this risk. One role that a Liberty-enabled system has, however, is in maintaining "contextual integrity."

Disclosures of PII are made in a specific context: shopping history with a particular retailer, health-care data with a family doctor, etc. When the contextual integrity is broken, and disparate pieces of PII — payment details, age, home address, books bought, health records — are combined, we get a remarkably invasive look at an individual's life. Liberty's federated architecture is designed to keep contexts separate.

Closely related to contextual integrity is data aggregation: using information from a variety of sources to determine aspects of a Principal's PII even if that information has not been specifically provided. Liberty guidelines recommend that the IdP — generally more able than the Principal to understand data-aggregation issues — have default policies that limit the release of PII. It may be appropriate for the Principal to be able to override these policies through an opt-in process[8] [10, pp. 22-23].

3.3 Further Steps in Liberty Privacy Protections: The Identity Governance Framework

Although the Liberty protocols protect the privacy of data exchange, there nonetheless remains a gap: explicit mechanisms for exchanging metadata governing data at rest. With the underpinnings provided by the ID-FF and ID-WSF specifications in place, Liberty is preparing an Identity Governance Framework (IGF) [13] that will provide the Principal with a clear framework for assessing whether her privacy is being protected in the way that she wants.

IGF enables the creation of declarative contracts (or policies) between an Attribute Provider and a Service Provider. To achieve this, IGF defines two declarative syntaxes: Attribute Authority Policy Markup Language (AAPML) and client Attribute Requirement Markup Language (CARML).

The Attribute Provider uses AAPML (a profile of OASIS XACML) to create statements pertaining to the access and use of the protected attributes. For instance, it has the ability to express conditions permitting release of the data (e.g., *any authenticated student can read these teaching notes*), obligations for the SP (e.g., *this document shall not be stored for more than two days*), and the need to obtain the Principal's consent. Meanwhile the SP can specify whether the requested attributes will be automatically discarded after usage. Or the SP could request the possibility of modifying the data or forwarding it to another SP. The CARML document created by the SP can either be created and exchanged ahead of time or it can be created and included in the attribute request.

In addition to these declarative syntaxes, IGF also specifies a certain number of basic privacy constraints such as propagation, usage, retention, storage

[8] It is interesting to note that some electronic health care projects (e.g., the U.K.'s Summary Health Care Record [21]) have opted for implicit consent to allow access to personal health information. Consent must be expressly revoked by the client.

and display of identity data. Although non-exhaustive, these atomic privacy constraints can be combined using WS-Policy [27]. These are viewed as commitments made by the creator of that request.

3.4 Privacy Leaks in Liberty Protocols

In the years since the Liberty federated model was introduced, there have been a number of papers discussing various privacy concerns in the protocols and model. We now consider these.

Birgit Pfitzmann [19] and Pfitzmann and Waidner [20] found various security and privacy risks in the early Liberty protocols. Pfitzmann pointed out technical ambiguities as well as policy issues that ought to be clarified in the Liberty single signon protocol [19]. Pfitzmann and Waidner demonstrated a man-in-the-middle attack against the Liberty-enabled client and proxy profile: a dishonest service provider could interpose itself between a Principal and an honest service provider (or even simply pretend to be the Principal without an initial request from the Principal) and then request authentication to the SP[9]. This attack is possible because Version 1.0 of the Liberty protocols did not require the SP to sign requests. Pfitzmann and Waidner suggested various ways to protect against the attack. Liberty version 1.1 prevents this by determining the SP's URL from the SP's identity and including the URL in its response.

Alsaleh and Adams' paper describes other consumer privacy issues [2]. On one point we agree: Alsaleh and Adams noted that the lack of standard privacy expression languages could lead to inconsistent interpretation of data privacy directives [2, p.72]. This problem was larger than the Liberty specifications and has recently been addressed in the IGF [13] program. We believe that the other "threats" Alsaleh and Adams posit stem from a fundamental misunderstanding of where the privacy pressure points lie in identity management; we provide detailed responses in Appendix B. But we note that design choices occur in technologies. Sometimes these concern security versus usability. Sometimes they relate to whether to solve a problem technically or via policy or regulation. Almost all of the Alsaleh and Adams' concerns are ones for which legal and policy responses are the most appropriate. Indeed, rather than being a problem for the Liberty specifications, this reinforces the view that the Liberty Alliance was right to complement its technical specification work with guidance on where the boundary lies between the technical and policy aspects of an identity-management system. We see no other comparable group that has paid equivalent attention to this aspect.

Jøsang et al. argue that, "SPs are not able to distinguish between a security assertion that reflects a genuine user service request, or one that represents an SP masquerading as a user" [8, p. 121]. This language could be interpreted either as the man-in-the-middle attack raised by [20] — which was resolved in version 1.1 of the Liberty Identity Federation specifications — or that a dishonest SP could claim to be a Principal at another SP. If the latter is what is meant, the dishonest

[9] The problem had been independently discovered by Jonathan Sergent.

SP is not in a position to authenticate as the Principal at the IdP and thus cannot learn anything at the second SP. Jøsang et al. also claim that in a federated model "Different SPs within the same federation domain are technically able to match personal information of the same user because of the mapping between identifiers" [8, p. 1221]. Liberty's use of pairwise, directional opaque identifiers prevents this problem.

Bhargav-Spantzel et al. propose two approaches to protecting the Principal's PII: distribute user identity information amongst several entities and use techniques such as zero-knowledge proofs to prevent identity theft within an IdP or SP [3]. The former is what federated systems do. This paper misunderstands some Liberty protocols. Liberty does not require PKI for Principal authentication [3, p. 25] but rather, allows authentication to be done in a variety of ways [10, p. 10]. Bhargav et al. suggest that a problem with Shibboleth is a single central identity provider. Both Shibboleth and Liberty allow Service Providers to use multiple Identity Providers within a Circle of Trust. The paper correctly observes that Liberty does not account for untrusted SPs or IdPs within the specifications [3, p. 25][10]. The specifications are about data exchanges only; the Liberty Alliance expectation is that individual service and identity providers will develop their own solutions against insider attack.

4 Conclusions

To succeed, digital identity management systems must balance competing sets of needs from users, IDPs, and SPs. Identity-management systems derive both strength and legitimacy from the consent of the individual whose PII is being used. Yet to benefit from the identity and service providers' offerings, the user must disclose PII — and therefore expose it to risk, regardless of the security mechanisms employed. Those organizations that request only minimal PII — and then protect, use and dispose of that PII appropriately — serve the user while minimizing risk and their liability.

Federation allows identity-management systems to be constructed with secure, minimized data exchange and thus to be inherently privacy-protective. The Liberty specifications enable privacy-protecting implementations but cannot mandate them. In the end, implementers must make their risk assessment and decide on the balance between technical and other mitigators.

Laws and policies on the one hand, and technology on the other, form integral parts of effective privacy protection. As the Liberty Alliance observes, "The identity challenge is both technical, business, and policy oriented"[15].

Acknowledgments. Many people have worked on Liberty over the years and developed insights on identity management and its privacy drivers. We have

[10] The Liberty ID-WSF Security and Privacy Overview states that, "[T]hese entities may not adhere to their contracts. In that case, the issue is out of scope for Liberty, which is, after all, a set of technical specifications for data exchange. Instead such a situation is appropriately handled by the legal system."[10, p. 15]

greatly benefited from discussions with Jeff Hodges, Paul Madsen, Eve Maler, and Bill Smith. Paul Syverson helped us put the ideas in this paper into a broader context.

References

1. Acquisti, A.: Identity Management, Privacy, and Price Discrimination. IEEE Security and Privacy 6(2), 46–50 (2008)
2. Alsaleh, M., Adams, C.: Enhancing Consumer Privacy in the Liberty Alliance Identity Federation and Web Services Frameworks. In: Danezis, G., Golle, P. (eds.) PET 2006. LNCS, vol. 4258, pp. 59–77. Springer, Heidelberg (2006)
3. Bhargav-Spantzel, A., Squicciarini, A., Bertino, E.: Establishing and Protecting Digital Identity in Federation Systems. CERIAS Tech Report 2007-18
4. Gevers, S., Verslype, K., De Decker, B.: Enhancing Privacy in Identity Management Systems. In: Workshop on Privacy in the Electronic Society, pp. 60–63 (2007)
5. Hardt, D.: Identity 2.0 Keynote, `http://youtube.com/watch?v=RrpajcAgR1E`
6. idemix for Internet anonymity, `http://www.zurich.ibm.com/security/idemix/ptext.html` (last viewed April 29, 2008)
7. inCommon Federation, `http://www.incommonfederation.org/`
8. Jøsang, A., AlZomai, M., Suriadi, S.: Usability and Privacy in Identity Managements Systems. In: Australasian Information Security Workshop: Privacy Enhancing Technologies 2007 (2007)
9. Wason, T. (ed.): Liberty Alliance Project, Liberty ID-FF Architecture Overview, Version 1.2 (2005)
10. Landau, S. (ed.): Liberty Alliance Project, Liberty ID-WSF Security and Privacy Overview, Version 1.0 (2003)
11. Varney, C. (ed.): Liberty Alliance Project, Privacy and Security Best Practices, Version 2.0, November 12 (2003)
12. Varney, C., Sheckler, V. (eds.): Liberty Alliance Project, Deployment Guidelines for Policy Decision Makers, Version 2.9, September 21 (2005)
13. Liberty Alliance Project, An Overview of the Id Governance Framework, ed (July 2007), `http://projectliberty.org/liberty/content/download/3500/23156/file/overview-id-governance-framework-v1.0.pdf`
14. Hodges, J., Kemp, J., Aarts, R., Whitehead, G., Madsen, P. (eds.): Liberty Alliance Project, Liberty ID-WSF SOAP Binding Specification, Version 2.0 July 7 (2007), `http://www.projectliberty.org/liberty/content/download/897/6267/file/liberty-idwsf-soap-binding-v2.0.pdf`
15. Liberty Alliance Papers, `http://projectliberty.org/liberty/resource_center/papers` (last viewed March 27, 2008)
16. Maler, E., Reed, D.: The Venn of Identity: Options and Issues in Federated Identity Management. IEEE Security and Privacy 6(2), 16–23 (2008)
17. McKenzie, R., Crompton, M., Wallis, C.: Use Cases for Identity Management in E-Government. IEEE Security and Privacy 6(2), 51–57 (March/April)
18. Office of the Chief Information and Privacy Officer, Province of Ontario, Privacy Impact Assessment Guidelines (December 1999) (updated June 2001)

19. Pfitzmann, B.: Privacy in Enterprise Identity Federation –Policies for Liberty 2 Single Signon. Elsevier Information Security Technical Report (ISTR), 9/1, pp. 45–58 (2004); preliminary version appeared as Pfitzmann, B.: Privacy in enterprise identity federation. In: Dingledine, R. (ed.) PET 2003, LNCS. vol. 2760, pp. 189–204. Springer, Heidelberg (2003)
20. Pfitzmann, B., Waidner, M.: Analysis of Liberty Single-Sign-on with Enabled Clients. IEEE Internet Computing, 38–44 (November/December 2003)
21. Ranger, S.: NHS e-record opt-out offered. IT Management News, December 19 (2006), `http://news.zdnet.co.uk/itmanagement/0,1000000308,39285203,00.htm` (last viewed January 18, 2009)
22. Shamir, A.: Secureclick: A web payment system with disposable credit card numbers. In: Syverson, P.F. (ed.) FC 2001. LNCS, vol. 2339, pp. 232–242. Springer, Heidelberg (2002)
23. Stubblebine, S.G., Syverson, P.F.: Authentic Attributes with Fine-Grained Anonymity Protection. In: Frankel, Y. (ed.) FC 2000. LNCS, vol. 1962, pp. 276–294. Springer, Heidelberg (2001)
24. U-Prove SDK Overview, April 16 (2007), `http://www.credentica.com/` (last viewed May 3, 2008)
25. Wilson, Y.: Personal communication
26. Winn, J.: Information Technology Standards as a Form of Consumer Protection Law. In: Winn, J. (ed.) Consumer Protection in the Age of the Information Economy, Ashgate (2006)
27. Web Services Policy 1.5 Framework (October 2007), `http://www.w3.org/TR/2004/REC-xmlschema-1-20041028/`

A Liberty Alliance Protocols

We present here an abbreviated discussion of the Liberty Alliance protocols ID-FF, ID-WSF, and Discovery Service. A fuller treatment appears in [9] and [10].

A.1 Federated Identity Frameworks: ID-FF and SAML

Liberty Alliance's ID-FF and OASIS SAML are frameworks that define protocols for (i) Single Sign-On and Account linking (aka. Federation) (ii) Name Registration (iii) Federation Termination and (iv) Single Logout.

Below are steps of a typical SAML2.0 based establishment of a single-sign-on (browser profile).

1. The Principal browses to an SP site and seeks access a protected resource there.
2. The SP responds with a SAML authentication request in an HTML form which after submission turns into either an HTTP Redirect or a POST sent to the IdP.
3. The Principal, if not already authenticated, is prompted with a login form at the IdP.

4. Upon successful authentication, the IdP sends (HTTP POST) to the Principal's browser to the SP with a <Response> message that contains a signed SAML assertion (possibly several). Among other things, the IdP might add information regarding the Principal default Discovery Service to facilitate bootstrapping an ID-WSF sequence as described in the next section.
5. If the SP is satisfied with the content of the assertion it has obtained, the SP grants the Principal access to the protected resource.

Note that pseudonyms are used between the SP and IdP in assertions to prevent account linking, and the NameID mapping protocol allow SPs to change the pseudonym used in federation.

A.2 Identity-Based Web Services Framework: ID-WSF

ID-WSF is a framework for identity-based web services. Its protocols support three core phases: (i) A WSP associating its service with a Principal's identity; a one-time operation for any given web service that enables the Principal to use this service in the future. (ii) A WSC querying a Principal's Discovery Service to look up a resource hosted at another web service provider (WSP). (iii) A WSC accessing a Principal's resource at that WSP, subject to conditions the resource's owner may place on access.

The ID-WSF protocols are based on a Request/Response design pattern. A SOAP binding document describes how such messages are to be created in a SOAP environment.

Service registration and association: ID-WSF defines a sequence of steps to allow a web service provider to become dynamically discoverable:

1. To registers its service instance at the DS, a WSP sends a SOAP message called <SvcMDRegister> at a known endpoint of the DS. This message contains XML metadata that describe its service (web address and type of service, the framework and security mechanisms supported etc.)
2. The DS returns an identifier (called MDID) for future reference.
3. When the Principal browses to WSP's web site, a typical response of the WSP (acting as a regular SP) will be to authenticate the Principal, possibly through single sign on as previously described. This authentication will provide information about the Principal's Discovery Service. The WSP offers to list its service as one of the Principal's known services at the DS.
4. If the Principal consents to this association, the WSP sends a request to the DS including the previously obtained MDID and some additional metadata.
5. Upon success (or failure) DS responds with a message that contains a <Status> element.

Service Discovery and invocation: Once registered and associated to a Principal, a service instance can be discovered and invoked by other web services. Following are the steps involved in this process:

1. To discover services of a certain type that are associated with a particular Principal, a WSC sends a SOAP message called <Query> at the lookup endpoint of the DS. This request contains a RequestedService XML element (possibly several) containing criteria for the lookup request (e.g. service type, etc.). The identity of the Principal is conveyed by the SOAP headers.
2. The DS returns a list of matching services, represented by endpoint references (EPRs). Each EPR contains a security token, crafted by the DS, to allow the WSC to invoke that WSP.
3. WSC can now invoke WSP, presenting the security token it has obtained during the previous step. ID-WSF defines a Create, read, update and delete interface that is designed for data-oriented services. This is the Data Service Templates, which serves as a base for the definition of many specialized interfaces. Liberty supports the definition of service interfaces for various types of services. The current service interfaces defined are personal profile, employee profile, geolocation, contact book and presence.
4. Upon success (or failure) WSP responds with a message that contains a <Status> element.

Additional Protocols: ID-WSF also defines two additional protocols that are important from a privacy perspective:

- The Interaction Service facilitates communication between a WSP and a Principal in cases where consent must be obtained before the WSP grants a WSC access to the Principal's resource it is hosting.
- People Service improves security and privacy in social networks and defines: (i) A service associated to a Principal, (ii) A flexible, privacy-aware framework to manage the people a Principal interacts with (invitations, identity federation), (iii) A SOAP interface for WSPs to query and manipulate information about a Principal's friends and colleagues.
  The People Service also enables further evolution for online transactions by supporting cross-principal interactions based on identity mapping.

B Detailed Response to Alsaleh and Adams

We have divided the issues raised by Alsaleh and Adams into three categories: those best solved through legal and policy means, those where the Liberty choice was in favor of usability, and those that had already been/are resolved.

1. Privacy Issues Resolved Through Legal/Policy Means:
   (a) Alsaleh and Adams posit a privacy risk during Identity Federation when the IdP introduces the Principal to the CoT [2, pp. 68-69]. That is a policy decision, and is made clear in the Liberty Deployment Guidelines, "The Liberty specifications provide for both access permissions to allow a Principal to specify whether and under what circumstances a Service Provider can obtain given attributes ... Has the Principal consented to all data uses?"[12, p. 8].

(b) Alsaleh and Adams raise concern about user consent for identity federation between IdP and SP; presumably what they mean is account linkage between the Principal's Identity Provider and Service Provider [2, p. 69]. Clearly such account linkage is a Principal's decision[11]; it is not clear why Alsaleh and Adams thought the Liberty framework made it otherwise.

(c) Alsaleh and Adams suggest that if there is a redirect between service providers that have different attribute information about a Principal, then two dishonest service providers could illegally exchange the attribute information, violating the Principal's privacy[2, pp. 70-71]. The key point here is "illegally." Only laws (and contracts, a form of law) can prevent a Service Provider from mishandling Principal information the SP has; technology cannot.

(d) Alsaleh and Adams object to the fact that once the SP has the address of the user attribute resource holder, the SP might retain the address past the current usage [2, pp. 71]. The security token is likely to have expired, but in any case, this is a legal issue; such retention does not conform with the timeliness aspect of the Fair Information Practices [10, p.15].

(e) Alsaleh and Adams state that Interaction Services (IS) hosted by other SPs may have privacy impact on the Principal [2, p. 72]. The IS *must* be trusted by the WSC [10, p. 19]. Vetting is done by the WSC and any liability is at the WSC.

(f) Alsaleh and Adams state that the SP could deny having made a query to the Principal [2, pp. 72-73]. As above, Liberty ID-WSF Security and Privacy Overview observes that such problems are an out-of-band issue, and recommends, "IS providers should make efforts to induce trust in the Principal by offering transaction logs, by employing sufficiently long strong authentication methods, etc." [10, p. 19].

(g) Alsaleh and Adams state that if a Principal deals with two SPs, there is risk of the two sharing PII about the user with each other [2, p. 72]. The opaque handles make such collusion difficult. However, since the SPs each have data about the Principal, such correlation is possible. In this case, the problem is not about the data exchange — the Liberty protocols — but about data at rest.

(h) Alsaleh and Adams observe an SP can amass data about a Principal, thus leading to identification of the Principal [2, p. 72]. This is not about data exchange, but about the fact that once you give data to a provider, they have it. Only law helps here.

(i) Alsaleh and Adams point out that the SP can reuse or share information about the attribute provider that hold the Principal's data [2, p. 72]. This is identical to the issue above.

[11] In the enterprise context, the consent is often part of an employment contract. A company might outsource aspects of its core business functions, e.g., human resources, and then the corporation consents — without an explicit request to the Principal — to have these introductions made.

(j) Alsaleh and Adams argue that privacy policies of the Attribute Provider local access policy and the Principal privacy policy might not match [2, p. 72]. Again there needs to be a business/policy/statutory decision on which policy takes precedence.

2. Privacy Issues Resolved in Favor of Usability:
   (a) Alsaleh and Adams believe that the SP knowing the Principal's preferred IdP is a privacy violation [2, p. 69]. The Liberty design decision was to choose usability with a minimum of privacy loss. After all, the IdP may have millions of users; how does the SP knowing that a particular user does so present a privacy disclosure? The reason the SP and IdP have a relationship is precisely because the SP wishes to rely on information the IdP can provide. If the Principal is genuinely concerned about this knowledge, he has the technical option of turning off federation domain cookies, or the procedural one of simply declining to link accounts.
   (b) Alsaleh and Adams claim that there is a risk that an SP could determine which IdP most recently authenticated the Principal from the federation common domain cookie [2, p. 69]. This is as above, with the same minimal privacy loss.
   (c) Alsaleh and Adams express concern that a federated SP can request that an IdP reauthenticate a Principal whenever the SP chooses and that the SP can query the IdP about the type of authentication method [2, p. 71]. All this reveals is information about the IdP's authentication methods that the SP already knows. There is no privacy risk for the Principal; this is feature (not a bug!).
3. Previously Covered Issues:
   (a) Alsaleh and Adams are concerned that a browser redirect can carry Principal PII unencrypted and is thus subject to eavesdropping [2, p. 70]. The Liberty Alliance previously addressed this concern [9, p. 21].
   (b) Alsaleh and Adams claim that an SP can request a resource-holder address for more than one user attribute from an IdP and that creates a privacy breach since the SP may use only one Usage Directive, which might not apply to all the attributes [2, pp. 71-72]. The UsageDirective can be applied to all conveyed attributes as multiple UDs may be employed [14, pp.46-48][12].
   (c) Alsaleh and Adams express concern that an IdP or SP could fabricate user consent and that the Principal has no way to force the SP to sign a request for consent [2, p. 72]. It is true that the Principal trusts the IdP, but that is an out-of-band issue. While it is true that the IS can be co-hosted with the requesting WSC — creating an "obvious" conflict of interest — the question is could this happen in a real deployment? As the *Liberty ID-WSF Security and Privacy Overview* states, it is simply the case that the IS must be trusted by the Principal [10, p. 19].

[12] The multiple <UsageDirective> header blocks for a SOAP Header was also true for version 1.2, which appeared in May 2005.