Abstract

Tools and techniques for digital identity management represent an important technology for enabling transactions and interactions across the Internet. Because identity information is often privacy sensitive, it is important that suitable privacy and security techniques be adopted for its protection. In this paper we discuss relevant concepts and issues and survey an approach based on the notion of multi-factor verification. Such approach, developed for federated digital identity management systems, is based on privacy-preserving cryptographic protocols and thus achieves high assurance privacy. In the paper we also discuss relevant open research issues, including interoperability, and protocols to support sophisticated policies for identity verification.

1. Introduction

Today a global information infrastructure connects remote parties worldwide through the use of large scale networks, relying on application level protocols and services, such as recent web service technology. Execution of activities in various domains, such as shopping, entertainment, business and scientific collaboration, and at various levels within those contexts, is increasingly based on the use of remote resources and services. The interaction between different remotely-located parties may be (and sometimes should be) based on little knowledge about each other. Thus, as the richness of our cyberspace lives begins to parallel our physical world experience, more convenient IT (Information Technology) infrastructures and systems are expected. We expect, for example, that personal preferences and profiles of users be readily available when shopping over the Web, without requiring the users to repeatedly enter them. In such a scenario, digital identity management (DIM) technology is fundamental in customizing user experience, protecting privacy, underpinning account-ability in business transactions, and in complying with regulatory controls. Digital identity can be defined as the digital representation of the information known about a specific individual or organization. Such information can be used for different purposes, ranging from allowing one to prove his/her claim to identity (very much like the use of a birth certificate or passport) to establishing permissions (like the use of a driver's license to establish the right to operate a vehicle). It may include not only “attributive information” about an individual, such as social security number or passport number, but also “biometric information”, such as iris or fingerprint features. For this technology to fully deploy its potential, it is crucial that strong protection of digital identity be achieved. Identity management (IdM) systems must assure that such information is not misused and individuals privacy is guaranteed.

In this paper we focus on a solution for the privacy-preserving verification of digital identity information, based on multi-factor verification strategy. By multi-factor verification we mean that whenever some identity information, referred to as identity attribute, about an individual needs to be verified by a party, for example a service provider, such party may verify the identity by requiring several identity proofs. The specification of which identity attributes have to be presented is stated by verification policies. Different parties in a distributed system may specify different policies. To assure that such an approach does not undermine privacy, we have developed a cryptographic protocol, referred to as aggregate zero knowledge proof protocol. Such a protocol allows a user to prove the knowledge of multiple secrets to a party, that is, a verifier, without having to reveal them to this party. In the paper in addition to illustrating the basic elements of the approach, we discuss important research issues, including interoperability and more sophisticated cryptographic protocols to support complex policies for identity verification.
2. Digital Identity

In this section we first introduce basic concepts in digital identity, such as strong and weak identifiers, and identifier ownerships. We then discuss relevant research and current trends.

2.1. Basic Concepts

Digital identity can be defined as the digital representation of the information known about a specific individual or organization. More specifically, the notion of digital identity refers to two different, not necessarily disjoint, concepts: nyms and partial identities. A nym essentially gives a user an identity under which to operate when interacting with other parties. Nyms can be strongly bound to a physical identity (that is, there exists a party or a combination of different parties that can link the nym to an individual) or be only meaningful in the context of a specific application domain or even a single transaction. Weakly bound or unbound nyms are useful in various contexts, including chat rooms and on-line games. Partial identities encompass a set of properties, such as name, birth date, credit-card-numbers, biometrics, transaction histories, referred to as identity attributes or identifiers, which are associated with individuals. Each subset of identifiers represents a partial identity of the individual. Partial identities may or may not be related to the human identity of one or more actual individuals. An important notion is represented by the notion of owner of an identifier, by which we mean the individual to whom this identifier is issued by a trusted authority or an individual who is authoritative with respect to the claiming of the identifier. In the former case, the trusted issuer of the identifier is also responsible for providing information about the validity of the identifier. Validity of an identifier encompasses several notions, including: correctness, that is, the identifier is correct (possibly with respect to the real-world), and timeliness, that is, the identifier is up to date.

When talking about identifiers, it is important to distinguish between weak and strong identifiers. A strong identifier uniquely identifies an individual in a population, whereas a weak identifier can be applied to many individuals in a population. Whether an identifier is strong or weak depends upon the size of the population and the uniqueness of the identifier. The combination of multiple weak identifiers may lead to a unique identification. Examples of strong identifiers are a passport number or social security number. Weak identifiers are attributes like country of origin, age and gender. This distinction is significant because misuse of strong identifiers can have more serious consequences, such as identity theft, compared to misuse of weak identifiers.

The notion of identity verification on which we focus in this paper deals with verifying that the identifiers claimed by an individual are actually owned by that individual. Identity verification is coupled with the notion of identity assurance, that is, the confidence about the truth of the claims related with the identity of an individual. High assurance verification of an identifier claimed by an individual means that the identifier is considered valid and the verifier has high confidence that the identifier is owned by that individual.

Strong and weak identity assurance exist regardless of the linkability of the identifier to the human identity of the actual individual. Note also that in some cases the knowledge about the actual human identity of an individual is not relevant for carrying out transactions and interactions. For example, consider the case of a scientific equipment covered by export control regulations, such a supercomputer. In such case for an individual to use the equipment the only identity attribute that is relevant is the country of origin. Additionally linkability among identifiers may exist with or without the identifiers being linked to the actual individual. Various cases exist that are summarized in Table 1 and illustrated by the following example.

Example 2.1 Consider an individual whose real name is Mary Brown and having a digital pseudonym Helen07. In a digital interaction when Helen07 claims to have a SSN equal to 123456789 and the verifier has strong assurance that the claim is correct (i.e. the SSN is valid and owned by user Helen07) and linked to the real world individual Mary Brown, then this corresponds to Case 1 in Table 1. Consider now another scenario in which Helen07 claims to have Italy as country of origin and the verifier does not know which real world individual does the claim belong to, but at the same time, is confident that the claim is correct. Such scenario corresponds to Case 2. Cases 3 and 4 correspond to the situation in which the verifier is not confident that the claim is correct; the difference is that in Case 3 the verifier knows who is the real individual presenting the claim, whereas in Case 4 the verifier is not aware of who this individual is.

2.2 Related Research and Current Trends

Several areas in computer science research have resulted in techniques and tools potentially relevant for DIM, rang-
ing from operating systems, security and distributed systems to human-computer interaction, artificial intelligence and cognitive science. DIM is indeed related not only to technology but also to people and organizations use of technology. However, there are only a few specific research efforts directly dealing with DIM in these arenas. The area that perhaps is most closely related to DIM is security. Security research has resulted in technologies that are important building blocks for any comprehensive DIM solution. Most relevant technologies include: digital certificates and certification authorities; PKI [3]; anonymity and pseudonymity techniques [13, 15, 29]; access control models and techniques, such as RBAC [1] and credential-based access; trust-negotiation systems [33]; privacy-preserving techniques, such as private information retrieval and zero-knowledge computational techniques; and threshold-based identity recovery schemes. Additionally, the standardization world has several relevant standards of interest, such as OASIS XACML and SAML dealing with access control policies, or even more important the P3P and APPEL standards dealing with privacy handling and preferences. It is important to note, however, that the integration and interoperability of these technologies towards an effective platform for DIM is an issue that has not been actively investigated by the academic research community. Interestingly, the main efforts towards more comprehensive DIM are underway in industry. One of the first approaches has been Microsoft Passport [19], based on the so-called centralized model. Under such an approach users perform single sign-on (SSO) and therefore they can authenticate once and gain access to protected resources across multiple systems. Here a central Identity Provider (IdP for brevity) is responsible for collecting and provisioning individuals identity information. A major drawback of such approach is that the IdP is potentially a single point of failure and is often not trusted by all participating parties. More advanced systems have adopted a decentralized approach whereby the responsibility of the IdP is distributed to multiple IdPs and users can select with which IdPs to interact. Such systems are often coupled with the notion of federation [21, 32]. The goal of a federation is to provide individuals with protected environments to share identities among organizations by managing individuals identity attributes. The members of a federation have trust relationships amongst themselves to share and use individuals identity attributes. Federations are usually composed of two main types of entity: IdPs managing identities of individuals, and service providers (SPs) offering services to registered individuals. In a typical federated IdM, the individual registers with his/her local IdP and is assigned a login name. Based on this information a registered individual can submit additional identifiers and corresponding identifier release policies that are stored at the IdP. From then on, the IdP is contacted whenever the individual interacts with any SP in the federation and additional identifier is needed. The IdP is then in charge of sending the SP the identifiers of the individual in accordance with the identifier release policies. In such federated systems, multiple IdPs are distributed and can store partial identity information of individuals, if required. Federations typically do not have the problem of single point of failure, but an IdP must be chosen that is also trusted by other entities. In most of these systems individuals have thus to depend on an online IdP to provide the required credentials and hence these systems are referred to as provider centric [16, 23, 27]. In some cases, such systems do not provide user control on his/her identity information, which is one of the key drawbacks of such systems. As a result, an emerging paradigm in federated IdM systems is that of user centricity, that is, the notion that individuals must be in full control of transactions involving their identity information. There are several terms closely associated with the concept of user centricity, including “user control”, “user consent”, and “user in the middle”. Achieving a good user control also implies strong security properties such as non-repudiation and stealing prevention. Thus, the new federated IdM systems need to incorporate the advantages of previous approaches, for example SSO and decentralization of IdP and at the same time provide further user control.

3. Identity Theft

The management of identifiers raises a number of challenges because of conflicting requirements. Although identifiers need to be shared to speed up and facilitate authentication of an individual and access control, they also need to be protected as they may convey sensitive information about an individual and can be targets of attacks such as identity theft. By identity theft we mean the act of impersonating others identities by presenting stolen identifiers or proofs of identity. More precisely, the act of identity theft occurs when an individual successfully uses an identifier or proof of an identity which he/she does not own. Usually, identity theft in the digital word occurs to obtain credit or perform other crimes, such as accessing classified information without having the appropriate authorization. People are increasingly concerned about identity theft as it is a serious economic crime. In 2005, the Consumer Sentinel, a Federal Trade Commission (TFC) complaint database, received over 685,000 consumer fraud and identity theft complaints [10]. There is also federal and state legislation regarding identity theft that has brought a heightened awareness to identity theft in general. For instance, the Identity Theft and Assumption Deterrence Act of 1998 makes identity theft a federal crime (18 U.S.C. 1028 (2003)). The purpose of this statute is to criminalize the act of identity theft itself, before other crimes are committed.
Through attacks such as password cracking, pharming, phishing [34], and database attacks, malicious parties can collect sensitive identity attributes of (targeted) individuals and use them to impersonate these individuals or to sell the identity attributes. Even though technical solutions are available that mitigate such attacks [11], a comprehensive approach to the problem of identity theft cannot rely solely on these techniques and must be able to offer protection from the threat of identity theft also when these solutions fail.

4 Federated DIM Approaches - Shortcomings and Requirements

Several initiatives are investigating federated systems for IdM [21, 24, 25, 32, 36]. However all those systems suffer from one or more shortcomings, which we discuss in the next subsection. Based on such discussion, we then identify some relevant requirements with focus on the problem of strong protection of identity attributes.

4.1 Shortcomings

In what follows we discuss shortcomings based on a simple notion of identity lifecycle.

- **Identity Issuance.** A first limitation is that no information is provided about whether the strong and weak identifiers being enrolled and stored at the IdPs have been verified to be correct with respect to validity and ownership, and the strength of this verification. If an IdP has such information then the SPs are in a position to make a more accurate judgment concerning the trustworthiness of such identity information. Second, most IdM systems lack flexible enrollment mechanisms for the individuals who want to enroll in their systems. Enrollment can be in-person at a physical location of an IdP or online. Current systems however, do not provide for alternative mechanisms for individuals to enroll. Moreover, the types of identifier that can be enrolled in most systems are also restricted based on the nature of the IdP organization [21].

- **Identity Usage.** A major drawback is that no specific techniques are provided to protect against the misuse of identifiers stored at the IdPs and SPs. Even the notion of misuse of such identifiers has not been thoroughly investigated yet. By misuse we refer to the case when dishonest individuals register fake identifiers or impersonate other individuals of the federation, leading to the threat of identity theft. To mitigate this threat, an upcoming trend is represented by strong authentication. Strong authentication often refers to systems that require multiple factors -possibly issued by different sources- to identify users when they access services and applications. However current approaches to strong authentication (such as those deployed by banks, enterprises, and governmental institutions) are neither flexible nor fine grained. In many cases strong authentication simply requires two forms of identity tokens, for example password with biometric. Through prior knowledge of these token requirements, an adversary can steal the required identity information to compromise such authentication [26]. Moreover if the same tokens are repeatedly used for strong authentication at various SPs, then the possibility that these tokens be compromised increases. Thus the implemented strong authentication does not meet the stronger protection requirements of identities in a federation. Individuals should be able to choose any combination of identifiers to perform strong authentication provided that the authentication policies defined by the verifying party are satisfied. Another drawback in the identity usage phase is the inability of the individuals to disclose minimal identity information about themselves to the SPs and IdPs as per the need of the service requested. There are several security and privacy concerns related to the extraneous identity information of the individuals that are stored at the SPs and IdPs. Moreover, such data may be aggregated or used in a manner that could potentially violate the privacy requirements of the individuals on their data. Current approaches also do not address how biometric data can be used in their system; in that digital identities are defined by digital attributes and certificates. The use of biometrics as an integral part of individual identity is gaining importance. At the same time, because of the nature of the biometric data, it is not trivial to use such data in a way similar to the traditional identifiers. It should be possible to use biometric data together with other identifiers to provide protection against identity attribute misuse. Another type of identity data that is not supported in current systems is the one related to individuals histories of online activities. If this information can be verified and used for evaluating properties about an individual, for example reputation, then this information becomes a part of the individuals identity. For example, consider a scenario where an individual frequently buys books from an online store. This history based information can be encoded as an identity attribute of that individual, which in turn can be used to evaluate the reputation of this individual as a buyer.

- **Identity Modification.** Most approaches do not provide flexible mechanisms to update or modify enrolled identifiers. As the information is shared within the
federation, the updates performed on one system do not ensure consistency of the individuals information within the federation. Additionally, these systems fail to prevent malicious updates by attackers that impersonate the honest individual.

- **Identity Revocation.** Finally, current federated IdM systems lack practical and effective revocation mechanisms. To enable consistency and maintain correctness of an individual identity information revocation should be feasible. Revocation in provider centric systems, in which the IdP provides the required credential to the user each time, is relatively simple to solve. Such credentials are typically short term, and cannot be used without consulting the issuer again. If, however, the credentials are stored with the user, such as a long-term credential issued by the appropriate authority, then building a revocation system becomes more challenging and critical.

## 4.2 Requirements

Current federated approaches to DIM [21, 24, 25, 32, 36] do not adequately protect individuals from identity theft. Dishonest individuals can register stolen identifiers or impersonate other individuals. Protection from identity theft should be one of the main desiderata in all IdM solutions. Even if an identifier of an individual is stolen, the system should make it hard for an adversary to use it successfully. Verification of identifiers is a key component of any solution to the problem of identity attribute misuse. Other important requirements for a secure and privacy preserving IdM system are as follows:

1. IdM systems main resource is represented by identifiers, thus security of such information should always be guaranteed. Security includes a comprehensive set of properties, such as integrity, confidentiality, revocability, and non-repudiation of ownership of identifiers.

2. Identity verification protocols should preserve individuals privacy, and enforce a “need to know principle” [30] when requiring identifiers. Privacy refers to the concept of giving an individual control over the release and use of his/her identifiers. In this context data minimization is required, in that only the attributes actually required to access a service should be submitted to the SP. Data minimization can be achieved by a combination of appropriate policies, and data release mechanisms supporting selective release of information.

3. A federated IdM system should ensure consistency of the identity data shared within the federation. Although validity of identifiers can only be verified by checking with actual identifier issuers, which could be outside the federation, the system should be able to detect misuse of identifiers based on the information available within the federation.

4. The verification methods should be efficient and require a limited number of message rounds between the SP and the individual. This would be one way to ensure usability of the system, as it is one of the main aims of federations.

5. The system should be able to support a variety of identifiers, including biometric data and individual usage history data referring to his/her online activity.

## 5. An Overview of the VeryIDX System

The VeryIDX system [26] has been designed as an approach to achieve strong verification of identifiers in federated environments. Such an approach is based on three key elements. The first element is the notion of multi-factor verification of identifiers, which consists of verifying that an individual owns an identifier by requiring from this individual other associated proofs of identity, that is, other strong identifiers. Our approach is based on the concept of proof of identity, which consists of a cryptographic token bound to an individual, versus the actual value of the individuals attribute. A proof is created in such a way that only the individual to whom the proof is bound can properly use it. The identifier proofs are built using Zero Knowledge Proof of Knowledge (ZKPK for brevity) techniques. The second key element is the notion of identity assurance level, that is, a level associated with an identifier that indicates the degree of confidence that the federation has in a certain identifier. Thus, the level indicates how strong the verification is for a given identifier. Such level is important for SPs in the federation when making decisions about granting access to services or resources. The third key element is the notion of authentication policies. Such a policy is a high-level statement by a SP specifying the types of identifier required by the SP when performing identity verification. An important feature of our approach is that each SP can require any combination of identifiers and different SPs may require different combinations. Also the number of identifiers to be combined is variable and each SP may specify a different number, according to its own authentication requirements. Such combinations can also be changed at any time by the SPs.

### 5.1 Aggregated ZKPK

The multi-factor identifier verification protocols adopted in VeryIDX are supported by efficient cryptographic prim-
itves. Our mechanism allows a party to prove the knowledge of multiple strong identifiers stored as cryptographic commitments using an aggregated, ZKPK protocol. The commitments are signed by a special federation entity, referred to as registrar, and the corresponding signature can be verified in an aggregated fashion at the time of use. To achieve aggregate signature we have developed a novel technique based on the Pedersen commitment scheme and integrated it with aggregate signature scheme to establish a new cryptographic primitive for aggregate proof of knowledge on those commitments. Once a client receives SP’s authentication policy, it retrieves from the registrar or the commitments $M_i$ satisfying the policy and the corresponding signatures $\sigma_i$. The client aggregates the commitments by computing $M = \prod_{i=1}^n M_i = g^{m_1 + m_2 + \ldots + m_i} h^{r_1 + r_2 + \ldots + r_i}$, and the signatures into $\sigma = \prod_{i=1}^n \sigma_i$, where $\sigma_i$ is the registrar’s signature on the committed value $M_i = g^{m_i} h^{r_i}$. According to the ZPK protocol, the client randomly picks $y, s \in [1, \ldots, q]$, computes $d = g^y h^s \pmod{p}$, and sends $d, \sigma, M, M_i, 1 \leq i \leq t$, to the SP. The SP sends back a random challenge $e \in [1, \ldots, q]$ to the client. Then the client computes $u = y + em \pmod{q}$ and $v = s + er \pmod{q}$ where $m = m_1 + \ldots + m_i$ and $r = r_1 + \ldots + r_i$ and sends $u$ and $v$ to the SP. The accepts the aggregated zero knowledge proof if $g^u h^v = d^e \sigma$. If this is the case, the SP checks that $\sigma = \prod_{i=1}^n \sigma_i$.

We have carried out an experimental evaluation of this approach that has shown that the time required to build a proof is constant and independent from the number of identifiers used in the proof [37].

5.2 Extensions

As the goal of VeryIDX is to provide a comprehensive approach to the problem of identity theft, we have also investigated the use of biometrics in the context of IDM systems. Today a large number of biometric devices and techniques are available and biometric-based solutions are increasingly being deployed [17, 28]. It is thus important that our framework be able to incorporate identifiers encoding information about physical features of individuals, in addition to “attributive” identifiers (such as SSN). The introduction of biometrics poses several non-trivial security challenges because of the inherent features of the biometric data. Biometric matching is probabilistic in nature, which implies that two samples from the same individual are never exactly the same. To preserve privacy and achieve interoperability between biometric identifiers and other identifiers, we have developed a biometric key generation algorithm. We have built on mechanisms from image hashing [9] and data classification techniques [14]. We have used Singular Vector Decomposition (SVD) on biometric images to derive a hash vector representing the biometric. Biometric images of the same individual would result in similar hash vectors. The similarity is evaluated using a Support Vector Machine (SVM) that classifies the hash vectors. We use the classification information to generate the final biometric keys. Such keys are used to generate ZKPK similar to the other strong attributes of the individual. Our algorithms capture generic biometric features to ensure unique and repeatable biometric keys. We have carried out an experimental evaluation of the proposed techniques using 2569 images of 488 different individuals for three types of biometric images, namely fingerprint image, iris image and face image. The experimental evaluation has shown that, based on the biometric type and the classification models, our approach can generate keys ranging from 64 bits up to 214 bits.

Finally we have extended our protocols to support the notion of history based identifiers. Such notion is motivated by the fact that such history can provide reliable information about the individual characteristics and behavior based on the online activities of the individual. We have thus extended our approach to support history based identity information in the context of e-commerce transactions and history-based trust management systems in which information about past transactions of the individual is used to make trust-based decisions concerning current transactions. It is important that these decisions be based on reliable transaction history information and that misuse of this information be prevented. A prototype of the receipt verification protocol has been developed for use also on cellular phones based on near field communication technology [23].

6. Open Research Issues

The development of policy-based DIM with strong protection from identity theft and user privacy and control entails addressing many issues. In what follows we focus on three such issues, by first discussing issues related to interoperability arising from the use of different name spaces and vocabularies by different parties in a federated system, and then discussing issues related to more complex protocols for verification of identity attributes.

6.1. Interoperability

Because digital identities are very often exchanged, federated and negotiated across different organizations interoperability is a significant problem. Interoperability issues may occur at different levels within a digital identity management infrastructure; they range from the use of different identity tokens and different identity negotiation protocols, such as the client-centric protocols and the identity-providers negotiation protocols, to the use of different names for identity attributes. The use of different names for identity attributes, that is referred to as naming
heterogeneity, typically occurs because the various parties involved in managing digital identities, that is, clients, SPs and IdPs, very often belong to different domains each using a different vocabulary to denote identity attribute names.

Therefore, SPs and clients are not able to have “meaningful” interactions because they do not understand each other. It is thus necessary to devise an approach to match the identity attribute names of SPs and clients vocabularies. A first issue that needs to be investigated to develop such an approach is the matching technique to use, which in turn depends from the types of variation in identity attribute names.

The variations that can occur in identity attribute names can be classified into syntactic, terminological and semantic variations. Syntactic variations arise because of the use of different character combinations to denote the same term. An example is the use of “CreditCard” and “Credit Card” to denote a client’s credit card. Terminological variations refer to the use of different terms to denote the same concept. An example of terminological variation is the use of the synonyms “Credit Card” and “Charge Card” to refer to a client’s credit card. Semantic variations are related to the use of two different concepts in different knowledge domains to denote the same term. Syntactic variations can be identified by using look up tables. Look up tables enumerate the possible ways in which the same term can be written by using different character combinations. Instead, to detect terminological variations, dictionaries or thesaurus such as WordNet [38] can be exploited. Dictionaries are used to retrieve all the synonyms of a given term. Semantic variations can be determined by using ontology matching techniques. An ontology is a formal representation of a domain in terms of concepts and properties with which those concepts are related. It is used to define the domain and reason about its features. Ontology mapping is the process whereby two ontologies are semantically related at conceptual level; source ontology concepts are mapped onto the target ontology concepts according to those semantic relations [16]. Typically an ontology matching algorithm takes in input two ontologies $O_i$ and $O_j$, and returns a set of triples of the form $(c_i, c_j, s)$, where $c_i$ is a concept belonging to ontology $O_i$, $c_j$ is a concept belonging to ontology $O_j$ that matches concept $c_i$, and $s$ is a confidence score, that is, a value between 0 and 1, indicating the similarity between the matched concepts. Therefore, an approach to match the identity attributes of clients and SPs should be based on the combined use of look up tables, dictionaries and ontology mapping in order to detect all the possible variations in identity attribute names.

A second issue is related to the matching protocol to use, that is, by which party the matching has to be performed and whether the fact that a client has already performed a matching with a SP may help in a subsequent matching. In digital management systems, the matching can be executed by the client or the SP. Performing the matching at the client has the obvious drawback that the client may lie and asserts that an identity attribute referred to in the SP policy matches one of its attribute, whereas this is not the case. The client may want to lie in order to get access to services or resources for which the client does not have the proper identity credentials. Therefore, SPs should perform the matching.

Finally, it is important to take advantage of previous interactions that the clients have performed with other SPs. It should be avoided that clients have to prove several times the possession of a same set of identity attributes. A possible solution is that SPs issue certificates to clients. Such certificates should assert that an identity attribute by a client matches an identity attribute by the SP and that the SP has verified that the client owns the attribute. Clients can use these certificates to prove that they own a set of identity attributes without going through the authentication process or to prove there is a mapping between a set of client’s attributes and a SP’s ontology.

### 6.2. Content-based Verification of Identity Attributes

Protection of identity attribute values in content-based access control poses another problem for DIM. For example, in certain cases a SP needs to verify that a user is older than 21 in order to grant the user access to its service. Revealing the user’s age in clear text allows easy access control on the SP’s side, but compromises the user’s privacy and makes it vulnerable to attacks like identity theft, therefore should be avoided whenever possible in digital identity management.

A zero-knowledge proof of knowledge (ZKPK) protocol, based on Pedersen commitment scheme [27] and Schnorr’s zero-knowledge proof protocol [31], are used in the VeryIDX system to allow a user to prove to the SP the ownership of an identity attribute without showing its value in clear. However, after the execution of the protocol, the SP only learns if the user knows how to open the commitment associated to the identity attribute value, but nothing about the value itself. Therefore, this limits its direct use for content-based verification of identity attributes, in which a condition needs to be checked. For example, it is not possible to use this ZKPK protocol to prove that a user’s committed value (age) is greater than 21.

Several zero-knowledge proof protocols [22, 5, 8, 4] are potentially useful for verification of conditions on attribute values in digital identity management. These protocols make it possible for a user to prove that a committed integral value lies in an interval $I = [a, b]$. Efficiency will be a major concern for employment of such protocols in DIM. Costly mathematics operations are performed in all these protocols. The soundness of these protocols are probabili-
tic: to achieve a desired level of assurance, multiple runs of basic operations/interactions must be executed. Although they sometimes can be optimized (e.g., by converting an interactive proof into a non-interactive one), the actual cost of communication and computation can still be heavy for applications (e.g., in a mobile environment). Furthermore, the protocols proposed in [5, 8] are able to prove that a committed integer \( x \in I \) belongs to an interval \( J \subseteq I \) with \( \#J/\#I > 1 \). This restriction makes applications of the protocols less straightforward in condition verification: depending on the real-world scenario, it may require a relaxed policy definition, or more strict rules on the encoding of committed values, and sometimes even declares these protocols unsuitable for application. An important feature of these protocols is that the verifier knows the outcome of the proof. This means that a SP learns via executions of the protocols whether a user’s identity attribute value is in the range specified in the policy or not. It marks a major difference between the zero-knowledge proof protocols and the OCBE protocols which we shall discuss in the text that follows.

Another candidate of content-based verification scheme for DIM is the Oblivious Commitment-Based Envelope (OCBE) protocols, proposed in [20]. The OCBE protocols allows a sender/SP to send a receiver/user an encrypted message (called an “envelope”), computed based on the receiver/user’s commitment, such that the receiver/user can decrypt the message (“open the envelope”) if and only if the committed value of the identity attribute satisfies a predicate specified by the sender/SP. The predicates supported by OCBE are the comparison predicates, i.e., \( =, \neq, >, <, \geq, \text{and} \leq \). By default, the sender does not know the result of the protocol, i.e., it does not know whether the receiver’s identity attribute satisfies the predicate or not. It automatically suits the need for privacy-preserving attribute-based policy enforcement in the case that the SP does not need to learn the verification result. Otherwise, if the SP has to know the result, it may ask the user to show the decrypted message, which can be a random bit string, then verify its integrity to continue communicating with the user. Since the decrypted message contains no information about the requested service, a user must own a committed identity attribute that satisfies the specified predicate to be able to correctly encrypt the message, thus to stay in the communication. However, the SP still only knows if the user’s identity attribute satisfies the condition, but not its actual value. As in the case of various zero-knowledge proof protocols discussed above, mathematics operations consist of the most computationally heavy part of the OCBE protocols. However, the OCBE protocols are deterministic: policy enforcement/condition verification can be ensured through one run of the basic protocol, for one predicate. Therefore, content verification can be more efficient with OCBE protocols, compared to other zero-knowledge proof protocols. Note that to verify that a committed value is in an interval, two OCBE protocols need to be executed, for the two predicates corresponding to the end points of the interval.

It is worth remarking that no protocol, including zero-knowledge proof and OCBE protocols, is able to allow a user to prove to a SP an attribute is of a certain value, in a privacy-preserving way, if the SP needs to learn the result of the protocol execution. In this situation, the SP automatically learns the user’s attribute value after knowing the value satisfies the condition.

As indicated in the discussion, mathematics computation involved in the zero-knowledge proof and OCBE protocols can make content-based verification quite expensive as the number of involved identity attributes increases, because multiple rounds of protocols need to run, one for each attribute. It is a challenge in practice to handle such a situation in a more efficient way. It is desirable to have methods that can simultaneously verify multiple conditions on attribute values, without significantly increasing the communicational and computational cost.

### 7. Concluding Remarks

In this paper we have discussed basic concepts in digital identity management. The discussion has focused on the notion of identifiers, that is, properties characterizing the digital identity of an individual. We have discussed the main dimensions in the verification of such identifiers and presented a brief overview of a system being developed based on such concepts.

As future work, we plan to extend the VeryIDX protocols to support different interaction models among the client, the service provider and the identity providers. We will also investigate how our protocols can be integrated into current DIM platforms and whether different verification protocols and credentials, such as anonymous credentials, can be integrated in VeryIDX.

**Acknowledgements.** This material is based in part upon work supported by the U.S. Department of Homeland Security under Grant Award Number 2006-CS-001-000001, under the auspices of the Institute for Information Infrastructure Protection (I3P) research program. The I3P is managed by Dartmouth College. The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the U.S. Department of Homeland Security, the I3P, or Dartmouth College.

**References**


