ADVISORY COMMITTEE ON ONLINE ACCESS AND SECURITY

SUB-GROUP: ACCESS 4 - IDENTITY, AUTHENTICATION, AUTHORIZATION AND SUPPORTING TECHNOLOGIES

Submitted: 18 February 2000

PURPOSE

The Access 4 subgroup will create an outline of the key issues to be addressed related to implementing appropriate and feasible methods for verifying the identity of and authorization for individuals seeking access to personal information collected from and about online consumers by domestic commercial Web sites. They will consider items discussed at the February 4, 2000 meeting such as: who gets access, what steps are/should be required, privacy implications and ability for consumers to remain anonymous, how access is currently provided, what technology will/may be available, etc.

MEMBERS

James Allen, Lance Hoffman, James Maxson, Frank Torres, Richard Purcell

OVERVIEW

This outline addresses the scope included in the purpose of the subgroup's charter, an illustrative model incorporating the scope, a variety of issues, and a glossary of terms. The basic methods of authentication are something the user knows (like a password), something the user has (like an electronically readable badge), or is (such as a biometric identifier like a retina, handprint, or thumbprint). These are typically used in combination - an ATM card requires a PIN (password) and is also used as an identifying physical and electronic token.

It must be recognized that different populations have differing sensibilities regarding acceptable techniques for identification, authentication, and authorization.

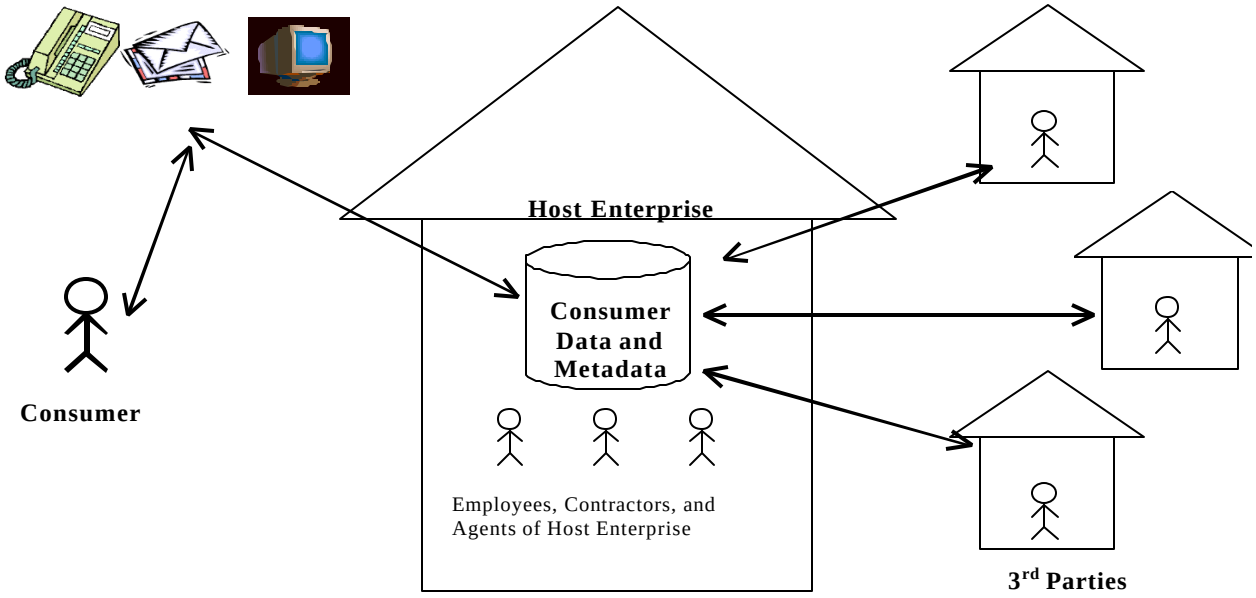
SCOPE

The scope of the Access 4 subgroup focuses on the issues related to identifying, authenticating, and granting authorization privileges to online consumers requesting access to data about themselves from a commercial website.

Other access to consumer data (e.g., by employees of the Host Enterprise and 3rd parties who are granted access by the Host Enterprise) is included in the scope of the Access 4 subgroup, as the issues and technologies are the same or similar.

ILLUSTRATIVE MODEL

We provide here a schematic intended to illustrate the scope.



ISSUES TO CONSIDER

1. Varying strength of access methods with the sensitivity of data

A system for classifying the sensitivity of data is needed so that acceptable access techniques, including identity of the requestor, authentication of requestor identity, and authorization level of the requestor, can be specified, including appropriate security precautions. An example (imprecise) classification system can be demonstrated as follows:

Data Sensitivity	Data Example	Authentication Strength
Publicly Available	Information in local phone books, property tax records, etc.	Weaker ↕ Stronger
Available to the casual observer	What your neighbors can learn about you without talking to you or being in your house; e.g., approximate age, major hobbies, clothing preferences, number of adults and children in household, type of car driven, approximate size of home	
Known to close acquaintances	What your close friends and co-workers know about you: e.g., approximate income, marital status, detail preferences	
Privileged Information	Information that is usually only shared on a need-to-know basis; e.g., exact income, mortgage owned, medical condition, financial account info.	

The problem with such a classification system is one size does not fit all cases – consumers will have different attitudes about the sensitivity of their personal information. Furthermore, the sensitivity may vary with the context in which the data is viewed.

This leads to the question "Should consumers be allowed to specify the sensitivity of data for the purpose of determining the appropriate access technique?" Thus, a negotiation is implied between the consumer and the Host Enterprise in which the consumer may have opportunities to specify sensitivity levels and the Host Enterprise may have opportunities to set ranges of access methods (from permissive to required) depending on the sensitivity classifications.

It may also be desirable to vary the strength of the authentication with the nature of the access (view, modify, add, etc.).

2. Access by consumers with pre-existing account

This issue incorporates authenticating consumers who have an established account with the hosting enterprise, and whose data is unambiguously bound to that account. In this case, the function of authentication is to answer the question "Is the consumer requesting access the person who owns the account?"

This is typically the case when an enterprise is collecting data about existing customers. The data are typically contributed by the customer (form fill out) or is derived from a history of transactions and/or observed behavior. There is no question that the data are about the person who owns this account.

In this case, classic authentication techniques (e.g., user id plus password) can be applied. It is important to note that identification is different than authentication, which is in turn different than authorization.

3. Access by consumers with no pre-existing account

Authenticating consumers who do not have an existing account with the Hosting Enterprise, or when the consumer data is not bound to existing accounts presents different requirements. In this case, the function of authentication is to answer the question "Is the consumer requesting access the person who the data is about?"

This is typically the case when an enterprise is collecting data on prospective customers from a variety of sources and linking it together based on a consumer identity such as a name plus postal address, e-mail address, or GUID in a cookie file.

An obvious, but future, solution involves digital signature technology, which, while available, requires appropriate and reliable legal and administrative solutions in order gain wide acceptance and distribution.

The challenge in this case is matching a consumer requesting access to data about that specific consumer. Financial institutions match consumers requesting online access to existing 'offline' accounts by name, postal address, social security number, mother's maiden name, the extra 3-digits from credit card accounts and other means. The theory is if you know enough information about the account you must be the person who owns the account.

Some Internet voting systems use a closed-looped system that depends on mailing a user-ID and password to the postal address where the voter is registered, and require the voter to return a signed form. Upon receipt of the form the account is released for activation by the voter. This way the Host Enterprise receives a signature to verify and keep on file that is linked to the user-ID and password used to vote.

A variation of the Internet voting system authentication might be used to authenticate consumers requesting access to data that is linked to a name and postal address. But this can be difficult because of the aliases people develop during the course of their lives (e.g., nick names, initials, maiden names), the variety of addresses they use or have used in the recent past, such as current home, previous home, P.O. Box, work address). Such systems have a built-in delay in access because of the dependence on the postal delivery and return of the User-ID and password.

Finally, the Internet presents special circumstances of user and data that are difficult to overcome. Users may have data about their web use stored on servers that track and capture

'clickstream' behavior. This data may not contain any identifying information beyond the GUID stored on the cookie file on the user's machine. Authenticating using only a GUID is suspect.

GLOSSARY

To further our mutual understanding of the concepts and taxonomy of the subject of Online Access and Security, we submit the attached Glossary of terms as an attempt to begin a standardization of the words used in our discussions, papers, and conclusions.

Term	Possible Definition(s)
Access	The mechanism(s) by which individuals can view data specific to themselves
Access/ Participation	The mechanism(s) by which individuals can view data specific to the themselves AND edit and update that data for accuracy and completeness
Access; authorized	The mechanisms by which access to data is granted by challenges to the requesting entity to assure proper authority based on the identity of the individual, level of access to the data, and rights to manipulation of that data.
Access; reasonable	To be defined as part of the advisory committee's work. Generally regarded as meaning that access cannot be constrained by artificial barriers set by interfaces, frequency, or cost of access.
Access; unauthorized	Access by an entity who does not have proper authority to access the information in the manner it is being accessed (view, modify, delete, etc.).
Alias	A name, usually short and easy to remember and type, that is translated into another name or string, usually long and difficult to remember or type. Commonly used as a single name for a list of e-mail addresses or hyper-link re-directs.
Anonymous	Describes an entity whose identity is unknown.
Authentication	Process by which an entity's identity becomes known.
Authentication technique	Method(s) by which an entity's identity can become known. Weak authentication includes weak passwords only. Strong authentication includes complex passwords combined with tokens (either physical, biometric, or electronic)
Authorization	Process by which a known (not anonymous) entity gains specified privileges such as access, read or write rights, system administration rights, etc.
Biometric Identifier	Methods of authentication based on the requester's unique biological traits, such as retinal patterns, handprint, thumbprint, voiceprint, facial details, etc.
Biometrics	The science of determining, storing, comparing, and validating the identity of an entity based on biometric identifiers.
Certificate Authorities	Entities that are empowered to sign digital certificates in order to add credibility to the certificate.
Choice/Consent	One of the five elements of Fair Information Practices, choice indicates that, once provided Notice/Disclosure, have options of choice over the data being requested and the use of that data,

Term	Possible Definition(s)
	including secondary use and sharing with 3rd parties.
Choice: Opt-in/Opt-out	Opt-out –mechanism that states data collection and/or use methods and provides user choice to decline such collection and/or use Opt-in – mechanism that states data collection and/or use methods and provides user choice to accept such collection and/or use
Commercial Web site	A for-profit entity operating an Internet web site.
Consumer	An individual (anonymous or identified) who interacts with commercial entities for personal benefits.
Cookie	A general mechanism which server side connections can use to both store and retrieve information on the client side of the connection. In essence, cookies are small data files written to a computer's hard drive by Web sites when that computer views the site using a Web browser. These data files contain information the site can be extremely simplistic containing only non-identifying data, identifying non-personally identifiable data (GUID's), or such things as passwords, lists of pages you've visited, and the date when you last looked at a certain page. Internet standards require that websites can read only those cookie files that they have issued.
Data Collection	The processes and sources used by a commercial entity to accumulate information about consumers. There are many methods of collection, including automated methods (see cookies), direct or indirect entry by consumers, and 3rd party sources.
Data Practices	The methods by which a commercial entity manages information. Specifically, the policies and methods used by a commercial entity in the collection, storage, access, security and distribution of customer information.
Digital Certificate	A digital certificate is a statement signed by an independent and trusted third party. That statement usually follows a very specific format laid down in a standard called X509, but it doesn't have to. Digital certificates consist of three parts: information about the object being certified (name, etc.), public key of the entity being certified, and the signature of the certifying authority.
Digital Signature	
Encryption	Any procedure used in cryptography to convert plain text into ciphertext in order to prevent any but the intended and/or authorized recipient from reading that data.
Enforcement/Redress	Mechanisms to ensure compliance (enforcement) and appropriate means of recourse by injured parties (redress)
Fair Information Practices	A set of principles designed to guide commercial entities in their data practices for customer and consumer information. See also Notice, Consent, Access, Security, and Enforcement.
GUID (Globally Unique Identifier)	Typically a long sting of alphanumeric characters that are assigned in such a manner that they are guaranteed to be unique within a well-defined context. These numbers are generally, but not always, assigned using a standard protocol called the UUID (Universally Unique Identifiers), however numbers such as social security numbers could also be considered GUID's.

Term	Possible Definition(s)
Host Enterprise	The entity controlling the systems storing personal information to be accessed by consumers
Identity	An identity is the collection of information that uniquely identifies and/or locates an individual. Usually some combination of first and last name, mailing address, email address, phone number, and age can by used to uniquely identify an individual.
Identity, verifying	Process by which an individual's identity if proven. Verification is different from authentication. Jane Doe can enter data identifying herself as Joan Smith and can be authenticated as such. However, Jane Doe's identity can only be verified as being Jane Doe unless she impersonates the identity of Joan Smith.
Internet	A giant network of servers and client computers interconnected through a set of protocols and distributed throughout the world.
Metadata	Data about data, in particular, a description of the data, its history, and authentication data (a digital signature, for example) related to the history and/or description
Notice/Awareness	One of the five Fair Information Practices, notice of an entity's data policies and practices must be provided to consumers prior to collection of personal information.
Password	A set of characters either assigned to or chosen by a consumer to be used to gain access to information and/or services. There are various methods of password construction ranging from weak (alpha only, not case sensitive - johndoe) to strong (alpha, numeric, and special characters required, case required – J0hEn_D0)
Personal information collected online	Information that is overtly collected from an individual via an online mediai (e.g., the world wide web); i.e., the individual contributes the information, and information that is collected from an individual by observation while the individual surfs the web (visits web sites) with or without the consumers knowledge.
personal information; derived	Information that is not directly contributed by an individual or automated collection process, but is calculated as a result of analysis of the collected data. For example, a new attribute of "potential new car buyer" can be derived by seeing that a consumer has recently and frequently visited car-selling sites, auto loan sites, and product-rating sites for automobiles.
Personally Identifiable Information (PII)	Those data elements that enable the identification and/or location of a unique individual. PII can be achieved as a single data point (such as e-mail address) or by a combination of data points (first name, last name, postal address).
Personally Identifiable Information: Sensitive	A classification of data used in the EU data directive that specifies certain information as deserving special treatment due to it's sensitive nature, including financial, health, religious, and sexual data. Sensitive data generally requires higher standards of authentication, authorization, choice, security, and distribution.
PIN	Personal identity number, a code known by an individual that is used to gain access to controlled resources, e.g., a PIN is used in conjunction with a magnetically coded bankcard to gain access to

Term	Possible Definition(s)
	bank accounts via an ATM machine
Security	One of the five Fair Information Practices, security assures that information shall be protected from unauthorized access, use, or distribution and shall not suffer quality degradation or loss.