Understanding Identity and Identification

U.S. Public Policy Council of the Association for Computing Machinery
January 2007

Professionals who work with issues of security and control use some terms to precisely describe access to resources and naming. These same terms have usage in general language, but the words frequently are used imprecisely and even misleadingly. When describing how security in information systems operate, and when formulating regulations or laws, it is important that these terms are understood and used precisely.

The purpose of this short document is to describe these important terms for readers who are not familiar with the more formal definitions. These related terms are *identification*, *authentication*, and *authorization*. Related concepts include *uniqueness* and *biometrics*.

Terms

Identification is associating a distinguishing label (*identifier*) with something within a specific group or context. You can identify someone by getting both their label and the context of that label. An ID card can provide both the name (e.g. “John Smith”) and the context (e.g., “licensed driver”). Identification can also occur by providing only the context or group name, such as identifying oneself as a police officer, a student, a graduate of West Point, or a member of Congress by wearing an appropriate badge, uniform, or class ring. The reliability of an identification depends on the confidence that the distinguishing label and context actually apply to the individual in question.

Note that even when identification is reliable – and it often is not – it does not imply anything beyond being able to distinguish among items or people. Identification can be used to determine if someone is a member of a group or not, or among members of the group. If someone were to identify herself as “Snow White,” that is an identification if she uses it consistently. In the context of a Halloween party or an Internet chat room, that may be a logical label to adopt.

A key concept is that identification does not need to be a standard name. It can be a nickname, a login, or a simple description, such as “I am the tallest one here” or “I am the one with red hair.” Those are means to distinguish one person from another in a particular group context.

People are most often identified in social situations by their names. In the United States, these names are usually composed of a first (given) name, one or more middle names (usually), and a last (family) name. In other countries, names may be a single word, or everyone may have a common family or middle name.

Uniqueness is when multiple items do not have the same identifier. Human names are seldom unique across a large enough population. For instance, there are many, many people named “John Smith” in the USA. If we also consider ancestors, then there may be even more individuals who have been associated with the same identifier (name). We can further qualify an identifier to make it more specific and less likely to be a duplicate of another identifier. For instance, someone could be “John Smith who was born April 1, 1952 in Boise and whose mother was named Matilda.” However, we cannot always be certain this is unique, and it is unwieldy to use in formal documents. Thus, we commonly use an artificial identifier that is generated and assigned in a manner that ensures that it is unique within context. For instance, Social Security numbers are supposed to be assigned without reuse, making them theoretically unique. Other identifiers (e.g., driver’s license numbers) are similarly generated to provide uniqueness.

Authentication is the process of verifying – to some desired level of confidence – that a claimed identifier is valid and actually associated with a particular item or person. Often, this validation is performed by one or more persons inspecting the identification and authenticator(s). The authenticators can also be examined by some technical means, such as a login program or a badge reader connected to a computer.

Authenticators of people are typically some combination of “something known,” “something possessed,” and “something about (structural)” the person. These items have been previously registered with the persons or organizations performing the authentication. Additional factors can also be used, such as physical location, recognition by human or canine guards, and so on.

- *Something known* is a secret or a fact that is unlikely to be known to an impostor. Passwords, when properly chosen and protected, are this form of authenticator. In many old combat movies, the spy is exposed because he doesn’t know which team won the World Series the previous year – this is another form of “something known” as a group authenticator. Many companies use items such as “mother’s maiden name,” “birth date” or “social security number” as authenticators, but this is bad practice as those items are often easily discovered facts: Many of these items are public information as a matter of law or custom.

- *Something possessed* is a distinguishable token or a key that matches a counterpart. A license issued by a government agency is a form of token. Another example from an old movie is the dollar bill or playing card that is ripped raggedly in half – the two halves are kept and joined together to *mutually authenticate* two parties.

- *Something about* (structure) the object or person being authenticated. We can examine something physical about the person we wish to identify, such as a fingerprint, or the pattern of blood vessels inside the eye. If the comparison of a person's distinguished characteristic is automated, then it is known as a *biometric*. A current location may also be used for authentication, such as GPS coordinates, telephone caller-id or computer network address.

Using a combination of authenticators is known as *multi-factor authentication*.

Authorization is the granting of rights (verb) or the grant itself (noun). Generally, one authorizes an authenticated party. *Permission* is used by some people as a synonym for authorization.

An example

Consider a scenario involving a person who wishes to enter a guarded building. When the person approaches the building to enter, a guard stops him to verify that he can enter. The person produces an *identification* card (something possessed) issued by a trusted authority (the context). The guard compares the picture on the ID with the face of the person, and causes him to put his fingers on a scanner (a biometric). These checks confirm that the person is the one identified by the card. She has been instructed that anyone with a valid blue card is allowed to enter, but without a cell phone, so she allows the person to pass after determining that he does not have a cell phone.
Note that this is use of multi-factor authentication, and the identification is based on group membership ("people with a valid blue badge") – no specific name or ID number is required. Permission to enter is the authorization involved. A further element of access control that is not based on identity or authentication is also involved: there is no authorization to carry a cell phone in.

There are many potential weaknesses in this system as described. The system can be redesigned to prevent the weaknesses, but defensive measures may be too expensive or cumbersome to be worth the effort given both the likelihood of the threats occurring and the value of what is behind the door. Examples of weaknesses include:

- The picture on the card may be old and the guard makes a false negative authentication: she refuses to allow the authorized person to pass.

- The guard may be overpowered or bribed so that unauthorized people enter.

- The card has been altered from a valid card — the color has been changed, or the original holder's photograph and fingerprints have been replaced by this impostor.

- The cards are made to published standards without adequate safeguards: this is a forged card made by a well-informed and sophisticated attacker.

- The attacker has stolen the card, disguised himself as the cardholder, and donned fingerprint caps that fool the scanning machinery.

- The guard is unable to recognize a disguised cell phone.

- Someone pretending to be a law enforcement officer, in uniform, orders the guard to let him pass or he will arrest her for obstructing justice. She complies.

- If too many people arrive in a short time, the guard may not be able to process them in a timely fashion, and someone is either denied access incorrectly or slips in unnoticed.

- The guard may fall ill and leave her post, leaving the door locked or unlocked for subsequent visitors.

- A first-time visitor has no way of knowing that this is really a legitimate guard and the right door.

Additional Notes

1. As illustrated by the last point in the previous example, the problem of authentication is bidirectional — all parties in the transaction need some level of assurance that they know the identities of the other parties. This is one reason why *phishing* succeeds: the customers enter their authenticating information, but the other party (the purported merchant) is not strongly authenticated to the customer.

2. It is possible to have authentication and authorization without specific identification. For instance, producing an *authentic* $20 bill provides authorization to make a purchase for something up to $20 in cost. It is not a requirement to *identify* the purchaser beyond being a member of the group who has cash.

3. Knowing precise, authentic identity **does not disclose intent.** Knowing the name of everyone who enters a building or boards a plane does not mean that they will be well-behaved. Mohamed Atta's Florida driver's license and picture were legitimate and examined when he passed through airport security on 9/11/2001. Most identification checks instituted in the wake of 9/11 perform at most a weak security function because there is poor (or no) authentication, and even when the identity is known it does not prove anything about intent.

4. Social security numbers are not supposed to be reused. However, numerous recorded cases of SSN duplication make the use of these numbers as unique IDs problematic.

5. Most biometrics have been developed and tested for authentication of a claimed identity, not for performing the identification itself; fingerprints are a notable exception. Insufficient experience has been gained with both physical features and biometrics to know error rates over large populations. By example, given the data that John Smith is 6'1" tall, has brown hair and green eyes, we can determine with some confidence whether a person in the room claiming to be John is actually John. However, given that same information and a crowd of people in a football stadium, we cannot be certain that we can uniquely identify John if he is present. Almost certainly, we will also make many false positive identifications. The same problems may exist with automated biometrics such as measuring facial features or hand geometry.

6. We know that every potential biometric has deficiencies. Not everyone has valid fingerprints over their entire lives, twins and triplets have the same DNA, and so on. People with special interests in some technologies have made unsupported claims about the performance of certain biometrics.

7. Most organizations use weak authenticators. In part, this is because most people are poor at remembering items such as long passwords and multiple ID numbers. As noted, use of authenticators such as mother's maiden name, social security number, or other such items is poor practice because those items can be easily found for many people.

8. Every instance where identifiers and authenticators are to be used should be carefully analyzed to determine strengths and weaknesses. This includes the value of what is being protected, and the consequences of false positives (authenticating an incorrect identity) and false negatives (failing to authenticate a valid identity).

9. As noted, identification and authentication mechanisms depend on context. Any security protocol is only as strong as the weakest element.

Association for Computing Machinery (ACM)

With over 90,000 members worldwide, the Association for Computing Machinery is the world's largest educational and scientific computing society, uniting computing educators, researchers and professionals to inspire dialogue, share resources and address the field's challenges. ACM strengthens the computing profession's collective voice through strong leadership, promotion of the highest standards, and recognition of technical excellence. ACM supports the professional growth of its members by providing opportunities for life-long learning, career development, and professional networking.

About the ACM U.S. Public Policy Council

The ACM U.S. Public Policy Council (USACM) serves as the focal point for ACM's involvement with U.S. government organizations, the computing community and the U.S. public in all matters of U.S. public policy related to information technology. Supported by ACM's Washington, D.C., Office of Public Policy, USACM responds to requests for information and technical expertise from U.S. government agencies and departments, seeks to influence relevant U.S. government policies on behalf of the computing community and the public, and provides information to ACM on relevant U.S. government activities. USACM also identifies potentially significant technical and public policy issues and brings them to the attention of ACM and the community. USACM publishes a monthly newsletter, the ACM Washington Update, which reports on activities in Washington that may be of interest to those in the computing and information policy communities, and highlights USACM's involvement in many of these issues. USACM is actively engaged in number of public policy issues of critical importance to the computing community.

For more information about USACM, please contact the ACM Office of Public Policy at (212) 626-0542 or see http://www.acm.org/usacm/.