

Primer: Authentication of identity (with application to PHRs/PHAs)

Reid Cushman, PhD

Director of Operations, Medical Information Technology, UM-Miller School of Medicine
Project Health Design ELSI Team, University of Miami

Copyright and disclaimer: This content may be freely used for non-commercial educational purposes with appropriate attribution to the source (Project Health Design ELSI Team, University of Miami). This content is offered as educational material only. It is not intended as professional advice.

What is “I&A” and why is it so critical?

Identification-and-authentication (I&A) is a core requirement of all security regimes. "Who are you?" and "Can you prove it?" must be answered in ways that allow legitimate persons in and keep intruders out. And that process must yield acceptable accuracy, at acceptable cost, without undue disruption of legitimate user activities.

One need not use computers to experience I&A. Protocols for proving identity are a ubiquitous feature of every adult's life. Think “Can I see your driver's license?” from the last time you used a credit card or wrote a personal check (at least if you did so in Miami). Even simple devices like mechanical locks can be thought of as performing a kind of primitive I&A – authenticating the right of entry to a physical space based on the possession of a physical key.

Physical proximity makes authentication easier, and not just because it allows the use of simple tokens like identification cards or keys. Consider how often you verify individuals' identities simply because they are familiar to you -- or, failing that, because they just “look right.”

By contrast, information applications require authentication of physically dispersed persons over a network – sometimes referred to as “e-authentication” – and so present greater challenges.

In computer contexts, a user's identification is typically translated as a unique “user-ID.” At least it is unique to that particular system or “name space” if not in the universe beyond. A social security number or employee number is another form of user-ID, unique within its particular context.

Verification that one really is the holder of a particular user-ID rather than an imposter – the A of I&A – is accomplished via three approaches:

- something the person knows, like a password (and the user-ID itself);
- something the person possesses, like a smart card; and/or
- something the person "is," like a fingerprint.

These methods may be used individually or combined.

Knowledge-based authentication

User-ID/password combinations are the classic knowledge-based authentication scheme, and remain the most common means of authentication for information devices and systems. This method is subject to well-understood limitations: Such information can be forgotten by the legitimate user, and can be obtained by theft or guesswork by illegitimate ones.

The simplest passwords are just sequences of alphanumeric (and sometimes symbol) characters. If the user supplies the correct sequence, identity is confirmed. The “password space” is the set of all character sequences that are possible passwords, given the system's limitations (e.g., the maximum length of a “legal” password, the types of characters accepted). The effective password space is much smaller than this theoretical limit. Why? Humans inevitably select character sequences that are easy to memorize, rather than truly random combinations.

Since passwords tend to consist of familiar words, one common method of guessing them is the “dictionary attack.” Repeated trial and error is used until the password is guessed by using a list of common words and word fragments. Limiting the number of incorrect password entries per unit time is one way to counter such trial-and-error probes (particularly those using automated methods). However, setting such error limits low can frustrate legitimate users with bad typing skills and/or bad memories.

Urging users to pick harder-to-guess-but-also-harder-to-memorize passwords tends to lead to another security problem: Users write down their passwords, and the “attack” is simply to find the list. Since users tend to use the same password over and over again for the multiple applications they access, finding a password in one place can lead to vulnerabilities for many systems. (The same problem occurs when users are required to change their passwords often. Typically everything is changed, but to the same new password. And that is written down.)

All in all, passwords are a very poor authentication method. It is widely estimated that the majority of security breaches – as much as 80 percent – are attributable to persons picking “weak” passwords that are easy to guess or to stolen passwords that are compromised because of poor password protection practices. The method survives because it is still generally cheaper than the alternatives.

Token-based authentication

Physical tokens eliminate (or at least reduce) the need to remember things – and with it the security problems that occur when people write down the things they cannot remember. But tokens can be lost by their legitimate holders, and, as with passwords, make their way into the hands of illegitimate ones.

In the physical world, tokens have long been used to authenticate identity and gain access. A key that fits a particular lock authenticates you to gain access to your house, car, etc. An identification badge may be required to authenticate your access to your work site. An ATM card is part of authenticating your access to funds in your bank accounts.

With an old-fashioned key and identification badge, simple possession is enough (unless someone looks at the photograph, a species of biometric authentication). ATM cards generally combine two methods – you must have the physical card, and you must also know a PIN number. Because the PIN is used in combination with the token, the “strength” requirements for PINs are lower. (It would not be acceptable to have 4-digit passwords, but it is ok for PINs.) That reduces the memory stress.

Token cards like identification badges often contain magnetically or optically (2D/3D barcode) recorded information that verifies what is physically inscribed on the card. More sophisticated tokens contain an internal microchip, which can store complex information – e.g., a digital

representation of a fingerprint or retinal scan, offering another form of combination authentication. The embedded microchip memory can also store content, such as elements of a personal health record.

Even more sophisticated tokens may accept input of a “challenge” (a string of letters and digits) provided by the system one is attempting to access, and then display a corresponding string of characters. That string is then input to the system by the user as a response to the challenge. Other varieties of smart tokens display a time-sensitive password, synchronized when the card is created, which the user must enter to gain system access. Both methods are a way of defeating fraudulent tokens, since such algorithmic cards are much harder to falsify.

In addition to the problem of being lost, tokens are expensive (particularly the sophisticated variety). They also tend to require “reader” devices that interface to the computer or other device to which the user is attempting to gain access. This adds to the expense. (Some cards are designed to take advantage of standard interfaces, such as USB. But they still generally require added software on the device to which they are attached. Management of such software is an administrative challenge.)

Biometric-based authentication

The last of the three approaches are generally labeled “biometric” methods. Biometric methods antedate all other methods used “natively” by humans, but represent the newest frontier for devices.

Biometric methods can be divided, somewhat arbitrarily, into two categories. “Physiological” methods use measurement of the face, eye (retina or iris), finger (fingertip, thumb, finger length or pattern), palm (print or topography), and hand geometry. Behavioral methods include measurement of voiceprints and handwritten signatures. There has even been experimentation with the measurement of personal odor (which could fall into either category, depending on your view of personal hygiene).

All of these can be summarized the same way: A “reference” scan is taken of the particular biometric element, such as a fingerprint, and a digital summary of that measurement is created. The same algorithm is used to create a comparative when the user presents for authentication. If the reference measure and the comparative measure are “close enough” the person is authenticated.

Biometric authentication has fallen in price dramatically, and it is now common to see biometrics such as fingerprint readers deployed on a wide variety of computing devices (e.g., laptops, flash drives).

Biometrics cannot be “forgotten” or “lost” like the other two methods, but that does not mean they are without significant failure modes. Biometrics can be “spoofed” to generate a false positive, though the prevalence of such fake-finger-created-with-a-gelatin-mold errors is probably a lot lower in real life than in spy films.

More commonly, the biometric scanner does not render a comparative with sufficient accuracy to match the reference scan, resulting in a mistaken rejection. Anyone who has used a low-grade fingerprint scanner and had to scan their finger multiple times for a single authentication knows the problem. Similar problems occur if the source of the user’s biometric is changed

(e.g., a burned or cut finger). You can always reset a person's password or issue a new token card. Re-issuing a finger, if it's hard for a sensor to read, is a different matter.

Acceptable accuracy

All "solutions" to authentication present tradeoffs among security level achieved, acquisition and maintenance costs, and the implicit costs of user inconvenience. The most important tradeoff is the one just discussed: That is, between acceptance errors (the wrong person is let in) and rejection errors (the right person is kept out). The stricter the authentication test(s), the more errors of the latter kind and the fewer of the former. The relative acceptability of these errors depends on the context.

Multiple methods used in combination (logical AND), such as the card-plus-PIN approach of ATMs, decrease the probability of acceptance errors but increase the probability of rejection errors. Multiple methods used as substitutes (logical OR) do the opposite. For example, many fingerprint-capable laptops and flash drives allow the use of a password if the biometric authentication fails. This increases user convenience, but at an obvious price.

The ideal balance for an information repository like a personal health record (PHR) – or a Personal Health Application (PHA) interacting with such a repository – cannot be specified absent considerations of the situation, and the person's preferences about types of authentication error. Is the information highly sensitive (e.g., genetic information)? Then perhaps erring on the side of rejection errors makes sense, and a stricter I&A method should be chosen. Is it important to get to the information quickly in an emergency (e.g., allergy and medication information)? Then tilting the balance toward acceptance errors makes sense.

While this discussion has focused on humans authenticating to devices, that is not the only interaction of concern for PHR/PHA use-cases. Device-to-device I&A may be important parts of such a system – e.g., a recording sensor "attached" in some way to a person, logging data into a PHR. Is it more important to get the data from the sensor to the log, and not miss anything? Is it more important that the log not contain data from the "wrong" sensor (i.e., wrong person)?

That both are important is obvious. But here as well, how to set the balance between the two is not. Because PHRs/PHAs are by definition to be used in personal settings, rather than the comparatively structured environment of health care organizations, the difficulties are compounded.

References

- National Computer Security Center, "A Guide to Understanding Identification and Authentication in Trusted Systems" (NCSC Technical Guidance TG-017, Sept.1991)

<http://www.fas.org/irp/nsa/rainbow/tg017.htm>

- National Institute of Standards and Technology, "Electronic Authentication Guideline" (NIST Special Publication 800-63, April 2006)

http://csrc.nist.gov/publications/nistpubs/800-63/SP800-63V1_0_2.pdf