

## Introduction to cryptographic hash functions

Krystian Matusiewicz

Centre for Advanced Computing Algorithms and Cryptography,  
Department of Computing, Macquarie University

A *hash function* is a function that maps strings of arbitrary length to strings of fixed length, i.e. any function

$$h : \{0, 1\}^* \rightarrow \{0, 1\}^n$$

for some fixed  $n$  called the hash length.

Hash functions (ordinary) are ubiquitous tools of computer science and engineering.

- hash tables
- pattern matching
- checksum algorithms

In all those applications the essential property is that the hash value is a short “tag” or “fingerprint” of possibly long data.

$$h(M_1) \neq h(M_2) \xrightarrow{\text{always}} M_1 \neq M_2$$

$$h(M_1) = h(M_2) \xrightarrow{\text{with high Prob.}} M_1 = M_2$$

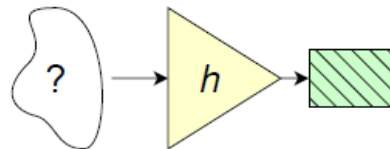
A *collision* for the hash function  $h$  is a pair of different messages  $M, M'$  that have the same hash value  $h(M) = h(M')$ .

- Since the domain is bigger than the range, collisions necessarily exist.
- Quite often we don't care as long as the probability of a collision is low.
- However, low probability of a collision for random inputs does not mean they cannot be found easily.

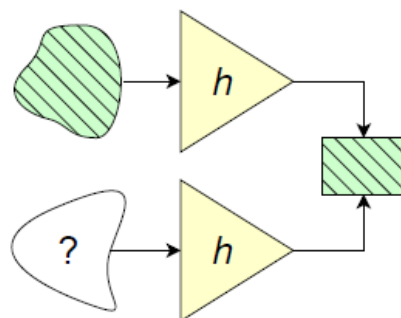
This ability to construct different messages that collide under the hash function may be exploited by malicious persons to attack a system.

This naturally leads to the definition of a **cryptographic hash function**, i.e. a hash function that satisfies the following security requirements:

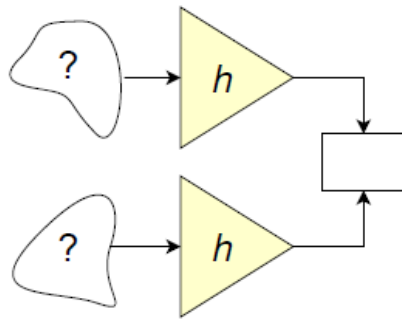
- preimage resistance
- second-preimage resistance
- collision resistance



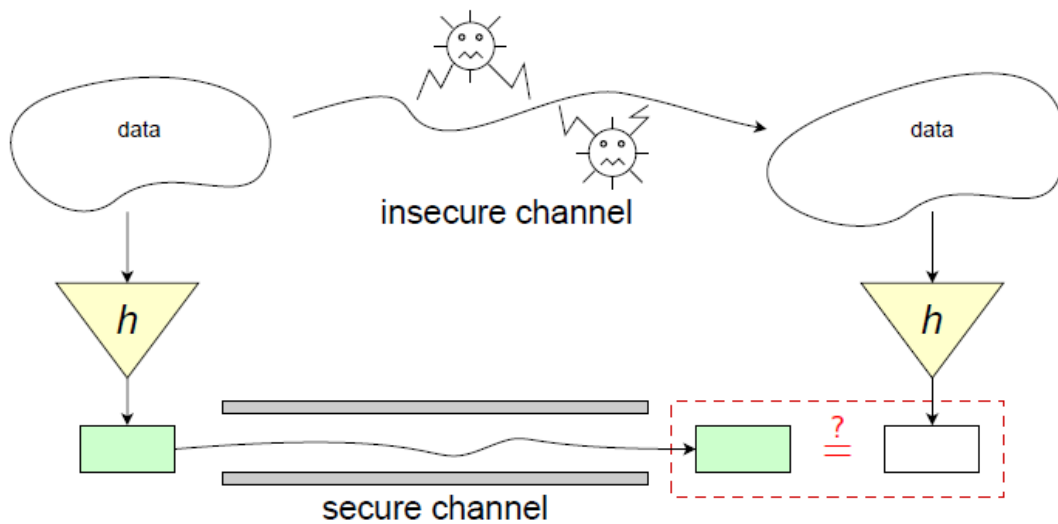
**Preimage resistant** : Given an output  $Y$  of the hash function it is difficult to find any *preimage* - an input  $X$  such that  $h(X) = Y$ .



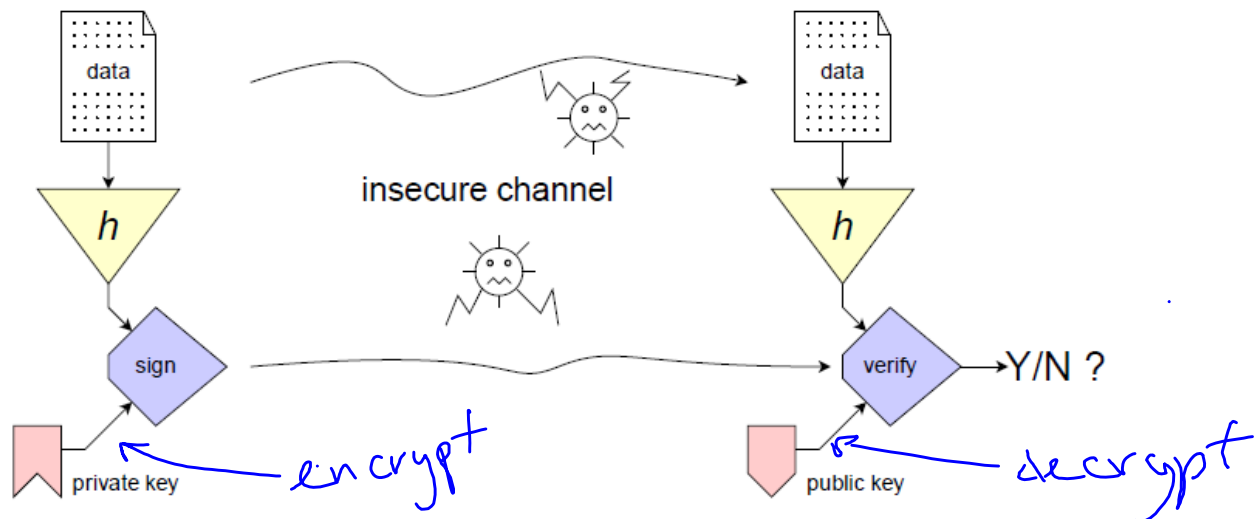
**Second preimage resistant** : Given a fixed input  $X$  to the hash function and the corresponding output  $h(X)$  it is difficult to find a *second preimage* - another input  $X'$ ,  $X' \neq X$  such that  $h(X) = h(X')$ .



**Collision resistant** : It is hard to find any pair of distinct messages  $(X, X')$ ,  $X \neq X'$  such that  $h(X) = h(X')$ .



Comparing the digest of data sent over insecure communication channel with securely obtained original digest allows to verify integrity of the data.



Digital signature schemes with appendix use cryptographic hash function  $h$  to obtain the digest of the data which is later signed.

---

## Password Authentication with Insecure Communication

Leslie Lamport  
SRI International

The first weakness can be eliminated by using a *one-way function* to encode the password. A one-way function is a mapping  $F$  from some set of words into itself such that:

- (1) Given a word  $x$ , it is easy to compute  $F(x)$ .
- (2) Given a word  $y$ , it is not feasible to compute a word  $x$  such that  $y = F(x)$ .

Instead of storing the user's password  $x$ , the system stores only the value  $y = F(x)$ . The user identifies himself by sending  $x$  to the system; the system authenticates his identity by computing  $F(x)$  and checking that it equals the stored value  $y$ . Authentication is easy, since our first assumption about  $F$  is that it is easy to compute  $F(x)$  from  $x$ . Anyone examining the system's permanently stored information can discover only  $y$ , and by the second assumption about  $F$  it will be infeasible for him to compute a value  $x$  such that  $y = F(x)$ . This is a widely used scheme, and is described in [2] and [3].

Our solution is to let the  $i$ th password  $x_i$  equal  $F^{1000-i}(x)$  for some fixed word  $x$ , where  $F^n$  denotes  $n$  successive applications of  $F$ . Thus, the sequence of 1000 passwords is

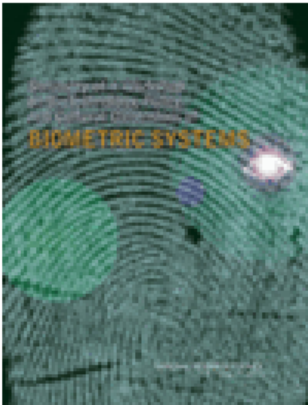
$$F^{999}(x), \dots, F(F(F(x))), F(F(x)), F(x), x.$$

The sequence of  $y_i$  needed by the system to authenticate these passwords is

$$F^{1000}(x), \dots, F(F(F(x))), F(F(x)), F(x).$$

Since it is feasible to compute  $F^n$  for  $n \leq 1000$ , property 2 of the one-way function implies that these  $y_i$  are distinct. For example, if  $F^{987}(x) = F^{123}(x)$ , then given  $y' = F^{123}(x)$ , one can compute  $x' = F^{986}(x)$  where  $y' = F(x')$ .

It follows from our definition of the  $x_i$  that  $y_i = x_{i-1}$  for  $i > 1$ . In other words, each user password is the value needed by the system to authenticate the next password. Hence, the system must initially be given the value  $y_1 = F^{1000}(x)$  and need subsequently remember only the last password sent by the user.



### **Summary of a Workshop on the Technology, Policy, and Cultural Dimensions of Biometric Systems**

Kristen Batch, Lynette I. Millett, Joseph N. Pato, Editors,  
Whither Biometrics Committee, National Research  
Council

ISBN: 0-309-65787-3, 62 pages, 8 1/2 x 11, (2006)

**This free PDF was downloaded from:**

**<http://www.nap.edu/catalog/11573.html>**

In Session 1, participants from industry, government, and academic research centers discussed the state of the art of biometric systems, the current bottlenecks, and areas where performance could be improved. Among the different types of biometrics, three were highlighted by the panelists—fingerprint, iris, and face—as being those accepted by the International Civil Aviation Organization for use in border-crossing documents. All panelists agreed that biometric systems cannot be made perfect—that is, the focus should be on how to evaluate and reduce, rather than eliminate, error rates. The challenges relevant in varying degrees to all biometric systems were grouped into three categories by the panelists, with primary emphasis during this discussion given to the first category of challenges.

- Improving the accuracy of biometric technologies and related performance evaluations through research on sensor resolution and ergonomics, algorithms and techniques for biometric fusion, characteristics of biometric feature spaces, and scientific methods to better quantify biometric systems' performance under realistic conditions.
- Systematically and thoughtfully integrating biometric systems with other security systems.
- Promoting interoperability of biometric systems, especially internationally, through a framework of standards, test methodologies, and independent evaluations.

Session 2 explored issues surrounding the measurement, statistics, testing, and evaluation of biometrics and biometric systems. It should be noted that statistical analysis in the context of biometric systems is and can be employed for a range of purposes, including assessments of the underlying technology, analysis of user behavior, data mining, and so on. Indeed, such issues were discussed throughout the workshop in several different contexts. Questions raised for this panel included these: Do biometric systems work? What is meant by “work” in the context of a biometrics system? *What* is being measured, tested, and evaluated, and how can confidence in the experiments be created? The panelists presented a range of perspectives on these issues, from broad explorations of the nature of experimentation and representative populations to discussions of specific evaluation regimens and real-world deployment at a major international airport. Several overarching themes arose:

- Evaluating biometric systems serves three purposes: to guide and support research and development, to assess the readiness of a system for deployment, and to monitor performance of a system in the field.
- As in many other domains, appropriate experimental design and solid statistical underpinnings are needed to produce effective testing and evaluation regimes. There is no one-size-fits-all solution, given the many types of systems that are deployed.
- Data and data selection choices, which include understanding the reference and expected user populations, can have a large impact on the accuracy and effectiveness of testing and evaluation.

In Session 3 panelists were asked to address the legal, policy, social, and cultural aspects of biometric systems, as well as the broad implications for society of the collection and use of biometric data in different contexts at both national and international levels. Major threads of presentations and discussions at this session included the following:

- Three different modes of identification evidence—mitochondrial DNA, facial recognition, and latent fingerprints—were discussed in relation to “general acceptance” and “scientific validity”—two legal standards for the admissibility of evidence in a court of law.
- Lessons for biometric system security were drawn from the current uses of Social Security numbers (SSNs) and the growing incidence of identity fraud. The proposition of a new law restricting the sale and disclosure of biometric identifiers was actively debated.
- The meaning of “privacy” in relation to the use of biometrics technologies was discussed in terms of legal principles and some preliminary public opinion survey research.
- Issues related to the collection and use of data generated by biometric technologies and associated fair information practices were discussed in relation to an earlier study on the use of radio-frequency identification tags (RFIDs) and access cards in the private sector.
- The international legal and cultural dimensions of privacy were discussed, including their implications for the use of biometrics.