Chapter 9
Secure Robust Hash Functions and Their Applications in Non-Interactive Communications

Qiming Li
Institute for Infocomm Research, Singapore

Sujoy Roy
Institute for Infocomm Research, Singapore

ABSTRACT

A robust hash function allows different parties to extract a consistent key from a common fuzzy source, e.g., an image gone through noisy channels, which can then be used to establish a cryptographic session key among the parties without the need for interactions. These functions are useful in various communication scenarios, where the security notions are different. The authors study these different security notions in this paper and focus on forgery attacks, where the objective of the attack is to compute the extracted key (hash value) of a given message. This paper will examine information-theoretical security against forgery under chosen message attacks. The authors prove that it is not possible due to the entropy of the hash value of a given message can be reduced arbitrarily when sufficient message/hash pairs have been observed. In this regard, the authors give a computationally secure scheme, where it is computationally infeasible to compute the hash value even when its entropy may not be high.

A robust hash H is a function that maps an input message $X \in U$ to a binary string $b \in \{0,1\}^*$ such that, when given another message $X' \in M$ where X' is close to X, the hash of X' remains the same as b with high probability. In this regard, a robust hash function is different from a cryptographic hash function, which does not tolerate even a single bit of error. Furthermore, the domain M can be real-valued, e.g., M can be feature vectors extracted from images.

Robust hash functions are very useful in secure non-interactive communications, where two or

DOI: 10.4018/978-1-4666-1758-2.ch009

more parties wish to derive a session key from a common fuzzy source without interaction. Such a session key can then be used, for example, in identity verification or encryption.

A typical application scenario of robust hash functions is the protection against copying attacks, where attackers attempt to copy a legitimate watermark from a marked multimedia object to an unmarked object (Kutter et al., 2000; Craver et al., 1998). In such scenarios, we could use a watermark that is dependent on the content of the multimedia object. To achieve this, a robust hash function could be employed to extract a key from the given multimedia object, and then a watermark could be generated from the extracted key. In this case, the communication parties would be the watermark embedder and detector, where the multimedia object serves both as a communication channel and the common fuzzy source to generate the watermarking key.

In this scenario, we would require that the hash function should be robust against the noise expected in the actual watermarking application, yet it should be difficult (if possible at all) to estimate this key generation process for an unmarked object.

We note that the central part of the above security application is the extraction of the session key from the common fuzzy source. Therefore, in this paper, we are concerned with the more abstract key extraction scenario as illustrated in Figure 1. Suppose two parties A and B have access to some correlated random sources X and X' respectively (e.g., X and X' could be the picture of the same scene taken at different times of the day), and they wish to agree on a common (secret) session key based on their own random source without communication. In this case, a keyed robust hash function $H(\cdot)$ can be applied to allow both parties to generate the same hash b using a shared key K. This allows both to decide upon a session key that they can use to do various tasks without directly using their shared secret key or exchanging any information as required by common key agreement protocols.

As we can see from Figure 1, if X is an original multimedia object, and X' is a watermarked object obtained by embedding a digital watermark into X, then the hash b that can be consistently extracted can be used to validate the authenticity of the multimedia object. Nevertheless, such a consistent string b can be used in many other scenarios, where it is desirable to extract a consistent key from noisy data.

Despite the potentials of robust hash functions, it is often not easy to analyze the security. This is perhaps partly due to the complexity of the interactions among many different parameters, which affect the robustness and security (such as collision and forgery resistance), and partly due to the lack of clear threat and attack models.

Roughly speaking, robustness of a robust hash function measures its tolerance to permissible noise, and collision resistance measures the difficulty of an attacker finding two dissimilar messages that yield the same hash value (more precise definitions will be given in later sections).

In this paper, we study *forgery resistance* of robust hash functions (Swaminathan et al., 2006), which measures the difficulty for attackers to compute the hash value of a given message without knowing the secret key. Similar to settings used by Swaminathan et al. (2006), we first investigate information theoretical security measured by conditional entropy. However, instead of considering just one message X_1 and its hash

Figure 1. Session key extraction