Measuring Strength of Authentication

Workshop: Applying Measurement Science in the Identity Ecosystem
Version: 1, December 16, 2015

Information Technology Laboratory, NIST

1 Introduction

This document serves as a primer for discussions to be held at the "Advanced Identity Workshop" at the National Institute of Standards and Technology (NIST) in Gaithersburg, Maryland, on January 12 and 13, 2016. The workshop will convene federal agencies, commercial relying parties, and identity solution providers to collaborate on improving standards, guidance, and practices related to identity management.

1.1 Purpose

NIST seeks to draft a framework that will support a greater understanding of the strength of authentication technologies used in identity management systems. Such a framework will provide a methodology to compare authenticators and allow a determination to be made on the selection of appropriate authenticators that are commensurate with assessed risk. In addition to providing a clearer understanding of an individual authenticator's ability to mitigate risk, the framework will also present a common basis to understand the strength of combinations of authenticators within multi-factor systems as well as comparability among authenticators. We anticipate that this framework will be developed to accommodate state-of-the-art authentication practices, solutions, and technologies, so that their strength can be understood and evaluated through standardized testing methodologies and reporting metrics.

To develop such a framework, NIST has chosen to initially focus on biometric technologies due to their expanding use in the consumer market as a primary authentication factor used to access remote, online services, but for which measurement science has not reached the same degree of maturity as other authentication factors, such as cryptographic systems. This lack of parity between measuring the strength of biometric and cryptographic solutions has consequences, including the exclusion of biometrics as a single or primary authentication factor in NIST guidance for accessing remote federal systems. Certain system characteristics that are unique to biometric technologies have challenged the development of standardized evaluation processes to date, although several organizations have examined this problem over the years.[1] Examples of characteristics unique to biometric technologies include:

[1] See, for example, (1) http://www.cesg.gov.uk/publications/Documents/bem_10.pdf; (2)"Security considerations for the implementation of biometric systems", Chapter 22, pp 415-431 in *Automatic Fingerprint Recognition Systems*, edited by Nalini Ratha and Ruud Bolle, Springer (2004), http://dx.doi.org/10.1007/b97425; and (3) http://www.biometrics.org/bc2011/presentations/International/0927_1445_Mr13_Fernandez-Saavedra_v2.pdf.

- Biometric samples are different each time they are captured, which means no direct matching can occur in the cryptographic space.[2]
- Biometric samples are not secrets, which means that template protection schemes must create application-specific templates to allow revocation of the templates if compromised.[3]

This document will further detail these characteristics and discuss options for building a measurement framework given their constraints.

1.2 Scope

This work will explore creating a framework for a set of measurable or testable criteria used to come up with an overall score representing the security of authenticators. This concept is similar to the "strength of function" defined in the Common Criteria Scheme and elsewhere, and has previously been discussed in the context of biometric technologies.[4, 5] Strength of function relates to the amount of effort required to defeat a security component.

As stated above, the initial focus for the framework is biometric authenticators; however, this framework should be structured in a way that common considerations among other types of authenticators share common evaluation techniques.

This framework will provide the means to (1) assess the strength of authentication for biometric technologies as a component in identity management systems, and (2) compare the use of biometric technologies with other authentication factors, thus allowing entities to make decisions on the choice of authenticator(s) appropriate to their organizational risk profile.

2 Workshop Focus Areas

As described above, the content that follows introduces a set of considerations for measuring the functional strength of biometric authentication mechanisms. The intent is to address these concepts in a manner that produces a broadly applicable and highly extensible framework capable of satisfying the needs of diverse communities and sectors.

To this end, NIST requests that readers consider the following questions as they review this document and prepare for January's workshop:

- Is the development of an authentication scoring framework feasible? Could it be used to determine how strong the authentication technology is? If not, are there other techniques that could be developed to determine the relative strength of authentication factors?
- Is a vulnerability-based approach the most effective method for establishing a scoring model? If not, are there alternative concepts around which a scoring framework could be built?

[2] Juels, A., and M. Sudan, "A fuzzy vault scheme," in *Proceedings. 2002 IEEE International Symposium on Information Theory*, p.408, 2002. http://dx.doi.org/10.1109/ISIT.2002.1023680.

[3] http://biometrics.nist.gov/cs_links/ibpc2010/workII/4buschB_IBPC-ISO-24745-100305-2p.pdf

[4] "Common Criteria for Information Technology Security Evaluations," Version 3.1 Revision 4, CCRA, September 2012. https://www.commoncriteriaportal.org/cc/.

[5] *Supra,* note 1.

- Is the proposed framework sufficiently flexible to accommodate other types of authenticators? If not, what adjustments could result in a cross-authenticator model for determining strength?
- What existing work that can be leveraged to develop a strength of authentication framework?
- What arrangement of laboratory process should be established to test the strength of authentication? Should third-party assessors, laboratory accreditation, and certification bodies be considered? Would relying parties trust self-assertion by individual entities?
- How would the testing procedures for evolving mitigation strategies be developed and maintained? Should there be a central repository of knowledge? If so, under what structure or type of organization(s) should this repository exist?
- Are there other considerations for strength of biometric authentication not considered here?
- In the proposed framework for strength of biometric authentication, what mitigation strategies exist for each area of potential attack, and how should the scores be calculated?
- By what method should scoring assign weight to system authenticator functions? Should the weights differ for various architectures or scenarios?
- Should scoring aggregate function scores or represent them individually? Should a minimum score be specified for functions prior to aggregation?

3 Strength of Biometric Authenticators

Figure 1 depicts a generic biometric system and identifies the points at which an adversary may attack a biometric authenticator. The elements of this system could be self-contained in a mobile device, where the biometric is never released, or the system can be distributed among multiple corroborating entities. NIST’s proposed approach is to develop a framework that considers potential vulnerabilities and their respective mitigation strategies as the primary method of evaluating biometric authenticators. Based on these evaluations, each mitigation strategy would be assigned a score, the aggregate of which creates an overall score representing the strength of authentication of the biometric authenticator. Defining this framework must avoid aggregation of scores in a manner that obfuscates the mitigations applied across the appropriate threat vectors. That is, the framework must account for efforts to achieve a higher score by mitigating a significant number of vulnerabilities in only portions of the overall system, while leaving others vulnerable.

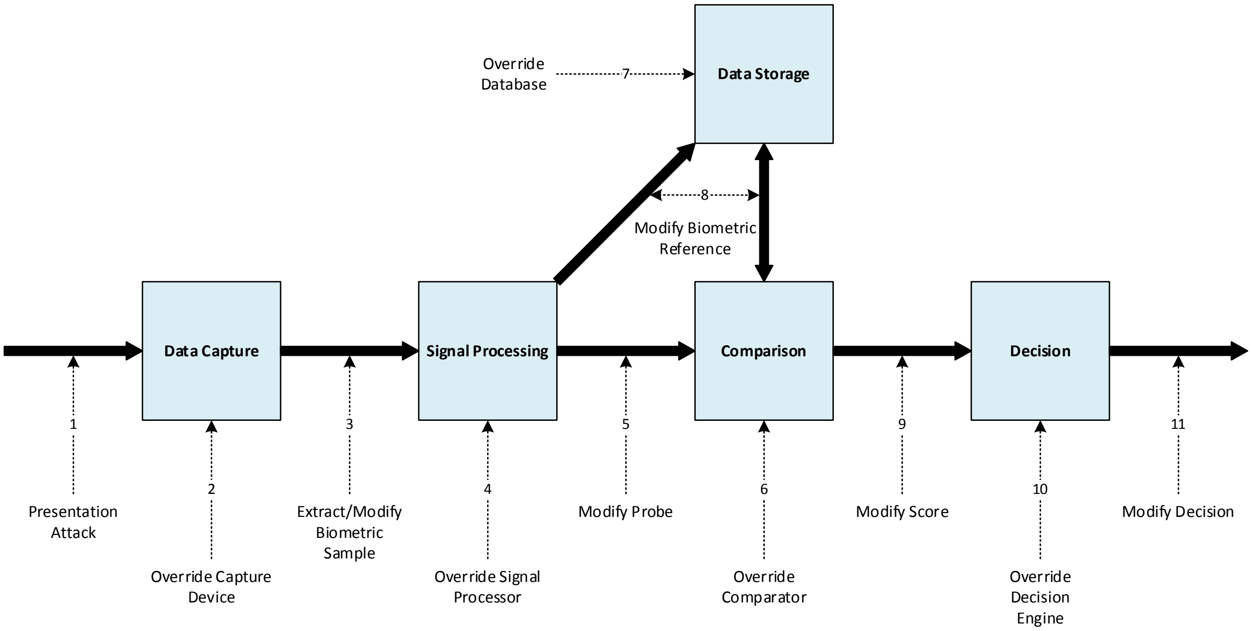

Figure 1 – Biometric system attack diagram[6]

The remainder of this section details each area of attack in a biometric system and explores potential mitigation strategies and methods of expressing strength. Parenthetical references in headings correspond to the numbers that appear in Figure 1.

3.1 PRESENTATION ATTACKS (#1)

A person's biometric patterns are not secret; some modalities are regularly shared publicly and available online (e.g., facial images on social media and company websites). Some modalities are easier to spoof than others. Biometric spoof attacks are a threat that needs to be mitigated either through presentation attack detection at the sensor—often referred to as liveness detection—or through other strategies to mitigate risk. Biometric spoof attacks are especially challenging in uncontrolled environments (i.e., remote authentication) where there is not an operator monitoring the placement or capture of a biometric.

Presentation attack detection (PAD) are methods created to directly counter spoof attempts at the biometric sensor. Artifact detection and liveness detection are types of PAD. Artifact detection attempts to answer the question: “Is the biometric sample at the sensor artificial?” Liveness detection attempts to answer the question: “Is the biometric sample at the sensor from a living human presenting a sample to be captured?”

Assessment of a biometric authenticator’s ability to detect presentation attacks could include the following criteria:

- Was the authenticator tested by the product vendor or integrator by simulating presentation attacks? Was it tested by any third parties? If so, how many?

[6] Inspired by N.K. Ratha, J.H. Connell, R.M. Bolle, “Enhancing security and privacy in biometrics-based authentication systems,” *IBM Systems Journal* 40(3), 2001, pp. 614-634, http://dx.doi.org/10.1147/sj.403.0614; and by ISO/IEC JTC 1 SC 37 in the development of DIS 30107-1.

- What was the number of types/recipes of spoofs used in testing?
- What system or subsystem was evaluated: the PAD subsystem only, the biometric data capture system, the biometric system, or the full authentication system (for multi-factor systems)?

Testers should implement methods to mitigate biometric spoofing that do not affect the matching performance or distinctiveness of the authenticator.

3.2 BIOMETRIC CAPTURE DEVICE $(\#2, 4)$

With the recent increase in the popularity of biometric technologies on mobile devices, consumer authentication events will be unsupervised. An unsupervised situation gives a malicious user the ability to modify or override the capture device itself. While this vulnerability is not unique to unsupervised situations, supervision makes such attacks more difficult.

Signal processing often occurs immediately following a capture. The signal processor, usually an algorithm or application, extracts important information from the captured data, which is then used to construct a template. Protecting this process from a malicious user helps ensure the integrity of the overall system.

Assessment of the protection of a biometric capture device could include the following criteria:

- Is the housing tamper-proof?
- What is the extent of integration of the algorithms with a Trusted Platform Module (TPM) or other secure element?

3.3 BIOMETRIC SAMPLES $(\#3, 5, 7, 8)$

The biometric and related data must be protected throughout all aspects of the process of authentication. The following subsections describe some aspects of protecting the biometric sample and related data.

3.3.1 Template Protection

Traditional cryptography cannot protect biometric data in the same way it does passwords or keys. Due to the variability of captured data between each session, authentication cannot rely on comparisons of encrypted or hashed biometrics as is done with passwords. Passwords and keys are deterministic data; authentication relies on an exact match each time they are typed in or presented. Biometric data is probabilistic, meaning a user's biometric sample is considered a match when it is sufficiently similar to an enrolled sample. These differences notwithstanding, secure templates can be used to encapsulate the biometric data along with additional information in such a way that a leaked template from one location cannot be used elsewhere.[7]

Such a secure biometric template must be cryptographically or computationally difficult to prevent reconstruction of the original biometric template. If the original biometric template is compromised, there is little to stop an adversary from using that information to create spoofed biometric artifacts. Unlike passwords, tokens, and certificates, biometric samples cannot be revoked or reissued and are not

[7] *Supra*, note 3.

considered secrets. A biometric template protection scheme must address ease and effectiveness of a revocability process. In addition to being able to issue a new template, the template should be diverse enough so that it can be application-specific and cannot be used to search against other databases. Providing enough data, in addition to the biometric, can increase diversity such that it differs greatly from other secure templates using the same biometric modality and sub-modality, thus preserving user privacy by limiting the reuse of templates across multiple applications and limiting the impact of a compromised template.

Finally, this template protection scheme should have little to no impact on the accuracy, or distinctiveness, of the authenticator.

Assessment of a biometric template protection scheme could include the following properties:

- Diversity: Is the template unique enough that it cannot be used across databases?
- Revocability: Does the scheme allow for revocation of a compromised template and reissuance of a new one?
- Security: Is it computationally or cryptographically difficult to obtain original biometric data?
- Performance: What is the recognition performance (i.e., False Match Rate and False Non-match Rate)?

Measures taken to address security, revocability, diversity, and performance of the template protection scheme could directly influence the overall score.

3.3.2 In-Transit (#3, 5, 7)

Data flows are subject to potential vulnerabilities and biometric systems must implement mitigation strategies when transmitting data from one location to another—even if within a device, organization, or network.

Data in transit is particularly susceptible to man-in-the-middle attacks of both extraction and modification. Modification of transmitted data directly affects the process's result, potentially allowing an adversary unauthorized access to information or systems. If extracting the transmitted data, malicious individuals have an opportunity to reverse engineer the template or data and to reconstruct the biometric template, thus compromising the integrity of that victim's specific biometric and allowing fraudulent authentication to occur. Furthermore, such fraudulent authentication may be undetectable as the authentication appears valid to the biometric system.

In addition to biometric template protection schemes, assessment of protecting data in transit may include the following criteria:

- Does the system use a symmetric or asymmetric encryption algorithm?
- What is the key size of this algorithm?
- Does the system use transport-layer security (TLS)?
- Does the system employ hashing algorithms to verify the integrity of the data?

3.3.3 At Rest (#8)

Data at rest are equally susceptible to attack as when in transit. Without appropriate mitigation strategies implemented, an attacker may successfully extract and modify stored information.

In addition to, or as part of, secure biometric template protection schemes, assessment of protecting data at rest may include the following criteria:

- Does the system employ encrypted storage to protect data at rest? If so, what are the encryption algorithm and key length?
- Is there physical storage isolation for data at rest within a TPM or other secure element?

3.4 Comparison & Decision (#6, 9, 10, 11)

The comparison of biometric data and decision of a match is another vital aspect of a biometric authenticator. The distinctiveness of a biometric modality is important—the higher the distinctiveness, the greater chance a comparison algorithm across the same modality or submodality will result in higher scores when used against biometric samples from the same user.

The use of multiple biometrics (e.g., two irises versus one) may affect the overall strength and warrants further exploration.

3.4.1 Distinctiveness

A suitable biometric modality (e.g., fingerprint, face, iris) must consistently distinguish an individual's biometric pattern from attempted impersonations of the individual. Said another way, intra-class variability must be sufficiently small, and the inter-class variability must be sufficiently large.[8] Intra-class variability refers to the variations of a biometric pattern of the same modality from the same user while inter-class variability refers to the similarity of a biometric pattern between different individuals.

Assessment of the natural distribution of a biometric pattern across the population should leverage empirical testing to determine if a biometric modality is distinct enough for a particular risk level. This requires a large sample of data to assure statistical significance of sufficiently low error rates. [9]

Although the inherent distinctiveness of a biometric modality is a necessary component of system performance, the actual implementation of the signal processing and matching algorithms, as well as any lack of fidelity caused by the capture device, may significantly impact the matching performance. Thus, a key component of biometric strength of function must include an assessment of the actual matching performance of the system that accounts for an implementation's capture device and level of processing capability.

3.4.2 Comparator, Score, & Decision

The comparator, the algorithm or application used in generating match scores between sets of biometric samples, is essential to a successful biometric authenticator. The integrity of the overall system depends on protecting this from a malicious user.

As match scores are sent to the decision engine, that channel must prevent the modification of any data. The decision engine relies on untainted match score data when calculating the final decision.

[8] The false non-match rate is more a concern for vendors and ease of use to their customers, but these error rates are not independent and must be reported together.

[9] For more information, see the Biometric Evaluations Homepage: http://www.nist.gov/itl/iad/ig/biometric_evaluations.cfm.

Like the comparator, the decision engine is an important component of a biometric authenticator. Modifying the decision engine gives a malicious user a quick way to shift the results: Accept this attempt or reject this attempt.

Finally, a biometric authenticator must protect the transmission of the decision to the requesting entity.

Assessment of protecting the comparator, score and decision may include the following criteria:

- Does the system use secure processing elements and memory?
- Does the system employ secure communication channels?

4 SCORE CALCULATION

A weighted combination of individual scores for the aspects described above can result in a final strength score. The final score will be comprised of weighted scores from analysis of strategies implemented to mitigate risk and attacks and requires evaluation to determine the appropriate weights to determine the most effective resulting score.

The following equation details how the score may be calculated under a weighting system C_n and scoring mitigation strategies for each vulnerability point V_n.

$$\text{Strength of authentication score} = \sum_{n=1}^{m} (C_n V_n)$$

Equation 1 – Final score calculation

In order for relying parties to understand the integrity of the entire system, score aggregation must be possible among distributed entities. For example, an organization that is interested in accepting a biometric authentication from a system deployed wholly on a mobile device will obtain the entire score as one package, likely from the handset vendor. However, there may be use cases where a local device sensor is used, but data storage, comparison, and decisions occur in one or several central locations. The service provider performing the central functions may interoperate with many capture sensors, each with their own strength of function score. The entire strength of function should factor in both the local sensor and the central system, each of which may be evaluated and tested individually. Relying Parties will want to know the individual and total scores of distributed systems to effectively compare them with localized approaches.

5 Conclusion

NIST intends to further the field of measurement science related to the strength of authenticators. We consider biometric authentication technologies a first step towards a generic framework for authenticator strength. Analysis of potential vulnerabilities for biometric systems across the stages of presentation, capture, enrollment, comparison, and decision is proposed as the basis for calculating the strength of biometric authentication, and an overall score based on combining the weighted scores of individual mitigation strategies.

This whitepaper represents an early proposal of such a candidate framework and NIST requests input from a range of experts to refine its direction and contribute to its development.