



Entropy of Fingerprints

Matthew R. Young

*Purdue University
(previous)
West Lafayette, IN USA*

Stephen J. Elliott

*Purdue University
West Lafayette, IN USA*

Catherine J. Tilton

*Daon Inc
Reston, VA USA*

James E. Goldman

*Purdue University
(previous)
West Lafayette, IN USA*

Abstract— The inherent differences between secret-based authentication (such as passwords and PINs) and biometric authentication have left gaps in the credibility of biometrics. These gaps are due, in large part, to the inability to adequately cross-compare the two types of authentication. This paper provides a comparison between the two types of authentication by equating biometric entropy in the same way entropy of secrets are represented. Similar to the method used by *Ratha, Connell, and Bolle* [1], the *x* and *y* dimensions of the fingerprints were examined to determine all possible locations of minutiae. These locations were then examined based on the observed probability of minutiae occurring in each of the designated locations. The results of this work show statistically significant differences in the frequencies and probabilities of occurrence for minutiae location, type, and angle, across all possible minutiae locations. These components were applied to Shannon's Information Theory [2] to determine the entropy of fingerprint biometrics, which was estimated to be equivalent to an 8.3-character, randomly chosen password.

Keywords — Fingerprint recognition; authentication methods; entropy

I. INTRODUCTION

Secret-based authentication mechanisms are some of the most widely used authentication mechanisms because of their very low cost to initially implement and the ability to be easily transported across multiple computing platforms. The established nature of secret-based authentication (such as passwords and PINs) has propelled these mechanisms to be the “de facto” standard for user authentication. Secret-based authentication ultimately translates into strings of information represented as input characters entered into the authentication system. The process of entering alphanumeric characters can be easily accomplished using a keyboard or keypad. Such devices have a definitive number of possible values that are available to the user. In the case of passwords, this is a 94-character alphabet, and, for PINs, this is a 10-character alphabet.

In biometric-based authentication, there is a tight bond between the authentication credential and the identity of the individual user. The inherent properties of this tight bond are described by O’Gorman [3] who states, “A biometric purports to inextricably link the authenticator to its owner, something passwords and tokens cannot do, since they can be lent or stolen” (p.4). The degree of binding between the user’s identity and the biometric sample presented by that user varies depending upon which biometric modality is being evaluated. However, one common theme is that the biometric sample is a digital string of information derived from the biometric characteristic provided by the user.

The inability to quantify the number of possible values in the biometric sample is currently one of the major

criticisms of biometrics, when trying to make cross-comparisons to secret-based authentication. According to O’Gorman [3], “A biometric doesn’t have a fixed number of possible values. Theoretically, the key space of biometrics such as fingerprints is unlimited because if you could measure the continuous signal with infinite precision, no two would be the same” (p. 15).

Entropy and strength of function are two concepts that have been the focal points in much discussion in recent years with regards to authentication mechanisms and, ultimately, to security systems. According to Shannon [2] entropy is defined to be the uncertainty of randomness of a discrete message space or string of information. The starting point for determining entropy is establishing the key space or the total number of possible values of keys in a cryptographic algorithm or other security measure, such as a password [4].

Key space, for both passwords and biometrics, is limited to the parameters of the respective systems. For instance, the length of a password credential could be as short as one or two characters or as long as hundreds of characters. In reality, the password length lies somewhere in between those values, normally around six or eight characters depending on the constraints of the system. Similarly, a biometric sample is not measured in a continuous space; rather, it is measured discretely. An example of a discrete measurement is a fingerprint image.

Strength of function is a term that is used to quantify the strength of an authentication mechanism (i.e., how difficult is a successful attack) taking into consideration many environmental factors about the system such as confidentiality (secrecy), integrity, and availability of the authentication credential.

According to O’Gorman [3], “For a biometric, the probability of falsely matching is analogous to the probability of succeeding in a brute force password guessing attack. One has to be careful in comparing *kp* (key space for passwords) and *kb* (key space for biometrics) The *kb* is based on an experimentally determined value of FMR” (p.15).

Statham [5] uses and expands the concept of entropy in breaking down the strength of authentication mechanisms. He suggests the concept of “real” entropy in determining relative strengths of function and its relationship to binding strength, which is the confidence that a person presenting the authentication credential is actually who they claim to be. Therefore, real entropy consists of three components: raw entropy, technical strength, and human/procedural strength. Raw entropy is the ability of the mechanism to distinguish between individuals. This is the exploitation avenue most used for casual (low or zero-effort) attacks.

Technical strength relates to exhaustion attacks against an authentication mechanism that exploit the vulnerabilities of that mechanism, as well as indirect attacks against the supporting infrastructure such as transmission paths or databases. Human or procedural strength attacks concern social engineering, “easy” secrets, failure to guard secrets, and corrupt users or administrators. This type of attack reduces effective entropy, sometimes to zero. While O’Gorman [3] and Statham [5] both address the entropy of authentication mechanisms from slightly different perspectives, it should be noted that in all cases, the discussion of entropy is limited to a functional error rate of the system. However, until now, a thorough examination into the entropy of the authentication information from Bishop’s model [6] for biometrics has not been conducted.

In Bishop’s model, the authentication process contains five sets of information (A, C, F, L, S); where the set A of authentication information is the set of specific information with which the entities prove their identities. Set C is complementary information that the system stores and uses to validate the authentication information. The complementation functions of Set F generate the complementary information from the authentication information. Set L of authentication functions verifies the identity, and set S of selection functions enables an entity to create or alter the authentication and complementary information in A or C.

One of the most well-known works that addressed biometric entropy from the perspective of key space was done by Ratha [1]. Their approach focused on a hypothetical brute force attack against minutiae based fingerprints. The first step in this process was to define the total number of possible sites at which minutiae could lie in a fingerprint image, taking into consideration the following:

- The dimensions of the image in pixels
- The number of pixels standard minutiae would consume
- The number of orientations allowed for ridge angle of the minutiae points
- The number of minutiae points in a reference template that is attempting to be matched against a new fingerprint image

Using the parameters noted above, a linear relationship was made between the amount of minutiae required to be matched and the amount of information represented in bits contained in the fingerprint. Based on this method, a fingerprint requiring 25 minutiae points to be matched would have 82 bits of information. According to that study, this equates to a 16-character randomly generated password. Our study builds on that work by determining the probabilities of occurrence for individual potential minutiae locations. To equate entropy of biometrics to those secrets based on probability of potential values, the principles of Shannon’s Information Theory were applied to minutiae points in fingerprints. Shannon’s equation for determining entropy as values of H in bits based on a single analyzed variable (X) with a probability value (P) is shown below [2].

$$H(X) = \sum_x P(x) \log_x \left(\frac{1}{P(x)} \right)$$

In the case that there are two variables being observed, the base equation is expanded to represent the concept of Joint Entropy. The probability value (P) is the probability of both variables occurring jointly [2].

$$H(X, Y) = \sum_x \sum_y P(x,y) \log_2 \left(\frac{1}{P(x,y)} \right)$$

The joint entropy equation is used for this study to allow for each possible minutiae location to be examined, as well as the combination of minutiae angle and minutiae type at a given location.

II. METHODOLOGY

A. Data Collection

The variables of fingerprint minutiae were analyzed on the basis of location, type, and orientation angle for every minutia in all usable fingerprint images collected. The location was represented as a coordinate (x, y) on a two dimensional pixel plane of the fingerprint image.

The database comprised 251 subjects, who each placed eight fingers (left index, left middle, left ring, left little, right index, right middle, right ring, right little) three times onto the fingerprint sensor to produce 6024 images (251 x 8 x 3 = 6024). Prior to the data analysis, the dataset was reviewed for image acceptability. This process revealed that 308 images were either completely white or completely black, and consequently, these images were manually removed from the dataset. These problematic images could be attributed to the automated timing window between each image captured and to the fact that the subject did not receive any feedback about each individual placement. In total, 5716 images were used for the analysis. These images were processed through Neurotechnology VeriFinger 4.2 SDK in conjunction with Microsoft Visual Studio .NET 2005 Application Development Suite. VeriFinger 4.2 minutiae extraction complies with ANSI-INCITS 378-2004 Finger Minutiae Format for Data Interchange [7].

The dimensions of all images in the database were 248 by 292 pixels, which created a total of 72,416 total possible pixel locations for a single minutiae point. Minutiae type was represented as one of two possibilities, either a ridge ending or ridge bifurcation. The minutiae orientation angles for a single point can range from 0 to 360 degrees, with 0° being the direction of a minutiae pointing directly to the top of the image, and 180° pointing directly to the bottom of the image. The use of four angle allocations was based on the same allocations in the previous work by Ratha [1]. The four possible quadrant values starting counterclockwise (0°-89°, 90°-179°, 280°-269°, and 270°-359°) are shown below.

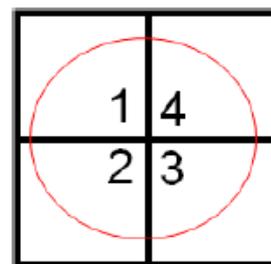


Fig 1: Minutiae Orientation Angle Quadrant Assignments

For the purposes of statistical analysis, the two-dimensional pixel plane was segmented into a matrix of 25 groups. These groups were created by dividing both the x and y dimensions into five segments. The numbering of the groups originated in the upper left corner of the image plane, which is the origin of the x and y coordinates used by VeriFinger. The center group (13) is shown in dark font, adjacent groups are highlighted in italics, and the remaining groups depicting the outer edges have no special highlighting. The following figure shows the group numbering from top left to bottom right, the x dimension in pixels and the y dimension in pixels for each group.

1	6	11	16	21		49	50	50	50	49		58	58	58	58	58
2	7	12	17	22		49	50	50	50	49		58	58	58	58	58
3	8	13	18	23		49	50	50	50	49		58	58	58	58	58
4	9	14	19	24		49	50	50	50	49		58	58	58	58	58
5	10	15	20	25		49	50	50	50	49		60	60	60	60	60

Fig 2: Group Numbering, x dimension, y dimension.

Each group has a calculated surface area which was designed to be as similar as possible so that groups could be evaluated on the frequency of minutiae presence. The surface area equation is shown below:

$$x \text{ dimension} \times y \text{ dimension} = \text{surface area}$$

The surface area (in pixels) for each group is shown below:

2842	2900	2900	2900	2842
2842	2900	2900	2900	2842
2842	2900	2900	2900	2842
2842	2900	2900	2900	2842
2940	3000	3000	3000	2940

Fig 3: Surface area (in pixels) for each group

The first hypothesis examined frequencies of occurrence for minutiae points to determine whether they are equally likely to appear in the center of the images as opposed to the outer edges and corners. The test results show that all groups are statistically significant compared to the center group (13) based on the minutiae frequency of occurrence. This finding is important when considering how difficult it would be to predict where minutiae would occur. Related to entropy, minutiae in locations that are less common provide more entropy than minutiae in locations that are more common across all samples.

The second hypothesis examined minutiae types (either ridge ending or ridge bifurcation) to determine if they are equally likely to appear in the center of the images as opposed to the outer edges and corners.

The results for this hypothesis show a statistically significant difference in all but the last group, which is adjacent to the center group (13). The numbering and location of the groups, as defined previously, and their statistical significance are shown in Table I.

TABLE I
HYPOTHESIS 2 EST AND P VALUES BY GROUP

		Ending		Bifurcation	
Ending		Est	P	Est	P
Intercept		-7.8216	<.0001	-8.1191	<.0001
Group	21	-1.9627	<.0001	-2.1211	<.0001
Group	25	-1.8495	<.0001	-1.9584	<.0001
Group	1	-1.6628	<.0001	-1.9784	<.0001
Group	5	-1.6400	<.0001	-1.9388	<.0001
Group	22	-1.0527	<.0001	-1.2428	<.0001
Group	24	-0.9015	<.0001	-1.2166	<.0001
Group	2	-0.8849	<.0001	-1.1517	<.0001
Group	23	-0.8738	<.0001	-1.0593	<.0001
Group	4	-0.8278	<.0001	-1.0534	<.0001
Group	3	-0.7533	<.0001	-0.9392	<.0001
Group	16	-0.6811	<.0001	-1.0849	<.0001
Group	6	-0.6258	<.0001	-0.9547	<.0001
Group	11	-0.4703	<.0001	-0.8432	<.0001
Group	20	-0.4905	<.0001	-0.6814	<.0001
Group	10	-0.4977	<.0001	-0.6431	<.0001
Group	15	-0.3412	<.0001	-0.5331	<.0001
Group	17	-0.2717	<.0001	-0.4129	<.0001
Group	7	-0.2110	<.0001	-0.3494	<.0001
Group	18	-0.1728	<.0001	-0.2380	<.0001
Group	8	-0.1461	<.0001	-0.2066	<.0001
Group	19	-0.1169	<.0001	-0.2447	<.0001
Group	12	-0.1270	<.0001	-0.2031	<.0001
Group	9	-0.1055	<.0001	-0.2054	<.0001
Group	14	-0.0041	0.8146	-0.0612	0.0028

An important finding from hypothesis 2 suggests that the ridge endings and bifurcations occur at different frequencies depending on the location the image. In particular, ridge endings are more prevalent towards the outer edges of the fingerprint image, whereas ridge bifurcations are more prevalent in the center of the image. A possible rationale for this could be that all the ridges start from a common point in the center of the image and branch out from the center, which results in a higher rate of bifurcations. As the multiple ridges created from these bifurcations approach the outer edges of the image, the finger starts to lose contact with the sensor. Subsequently, a ridge ending may appear in the fingerprint image even through the actual ridge is not exhibiting a true ridge ending.

The third hypothesis examined minutiae orientation angles (separated into four equal quadrants between 0° and 360°) to determine if the minutiae orientation angles are equally likely to appear in the center of the image as opposed to the outer edges and corners. The results showed statistically significant differences for all but four groups. The groups not showing statistically significant differences were group 18 for angle 1, group 16 for angle 2, and groups 2 and 3 for angle 4. The numbering and location of the groups are shown below:

1	6	11	16	21
2	7	12	17	22
3	8	13	18	23
4	9	14	19	24
5	10	15	20	25

Fig 4: Numbering and location of groups

Angles one and three are nearly identical in the frequencies of occurrences, and the same can be seen for angles two and four. This is because when considering a ridge as a line in a 360° space, the direction forward and direction backwards is a 180° degree difference. This reflects these similarities because angles one and three have 180 degrees of separation, as do angles two and four.

B. Entropy Calculations

The first step in calculating the entropy was to define the key space, which is the total number of possible combinations of minutiae occurrence, type, and orientation angle. To determine this total number of possible values, the space was categorized into two main parameters. The first was the possible pixel locations on the two-dimensional plane and is denoted by L, which is the surface area of the image with dimensions x and y. The second parameter was the possible characteristics about a minutiae point and is denoted by C, which comprises the minutiae type and minutiae orientation angle. L and C together are the basis for determining the key space for fingerprint images. Based on the definition of key space, the total number of possible values of keys in a cryptographic algorithm or other security measure such as a password, the generic equation for key space used in this study is:

$$\Sigma_L = \Sigma_X \Sigma_Y$$

$$\Sigma_C = \Sigma_T \Sigma_D$$

$$Keyspace = \Sigma_L \Sigma_C$$

These parameters of L (pixel locations) and C (minutiae characteristics) are the same two parameters used to calculate joint entropy. The x dimension of 248 pixels and the y dimension of 292 pixels provide 72,416 possible values for L. For an individual point, type is one of two values (either a ridge ending or ridge bifurcation); and angle is one of four options (either angle 1, 2, 3, or 4). These options provide eight possible values within C. Therefore, the key space for a single image is calculated using the following values below, which take into account all possible values for the x dimension (x), y dimension (y), type (t), and angle (d).

$$\Sigma_L^{74216} = \Sigma_X^{248} \Sigma_Y^{292}$$

$$\Sigma_C^8 = \Sigma_T^2 \Sigma_D^4$$

$$579,328 = \Sigma_L^{74216} \Sigma_C^8$$

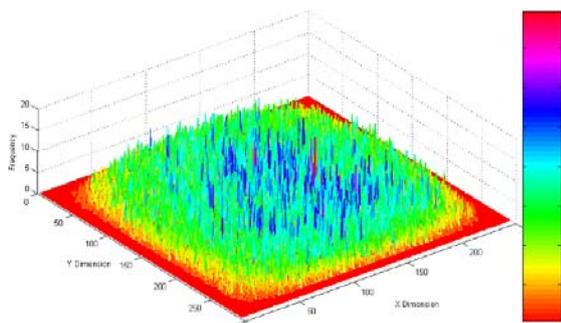


Fig 5: differences in frequency of occurrence for all of the 72,416 pixels in a two-dimensional image plane

Figure 5 shows the differences in frequency of occurrence for all of the 72,416 pixels in a two-dimensional image plane. Using the color-coded scale to the right of the diagram, the frequency of occurrence for minutiae is higher in the pixels towards the center of the image as opposed to the corner and outer edges. The combined probabilities for angles and types are shown below.

The joint entropy equation is shown below:

$$H(L, C) = \Sigma_L^{74216} \Sigma_C^8 P(l,c) \log_2 \left(\frac{1}{P(l,c)} \right)$$

The result of the joint entropy equation is 55.02 bits of entropy for the two dimensional plane across the entire dataset. The average number of minutiae present in each image was 28.02. Therefore, dividing the entropy of 55.02 bits by the average minutiae of 28.02, each minutia provides 1.963 bits of entropy. Table II illustrates the calculated entropy in bits and the conversion to the estimated length of a secret-based authentication credential. This is analogous to table A.1 from as shown in NIST SP800-63 - Electronic Guidelines for Federal Agencies [8] which used character length to estimate entropy. Figure 6 represents a similar linear relationship to the one presented by Ratha [1], however, the table presented above goes a step further to convert the bit strength to an estimated character length for a randomly chosen password, user-chosen password, or user-chosen PIN using the NIST methodology.

TABLE II
ESTIMATED SECRET LENGTH

		User Chosen			Randomly Chosen		
		94 char Alphabet			10 char Alpha	94 char Alpha	
Minutiae Present	Bits	No checks	Dict. Checks	Dict & Complexity Checks			
15	29	15	13	8	24	4.5	
20	39	23	23	17	34	5.9	
25	49	33	33	27	44	7.4	
28.02	55	39	39	33	50	8.3	
30	59	43	43	37	54	8.9	
35	69	53	53	47	64	10.4	

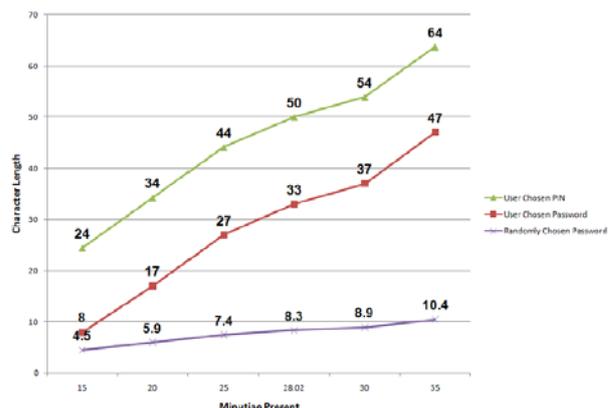


Fig 6: minutiae matched and character length relationship

The bottom line represents the randomly chosen password; the middle line represents the user-chosen password, and the top line is the user-chosen PIN. A randomly chosen password was used based on the rationale that while users can choose the nature and order of the characters in their password, users have no discretionary ability to choose the nature and order of the minutiae on their fingerprint. Using the average number of 28.02 minutiae present in a single image from the dataset, an equivalent randomly chosen password would be 8.3 characters in length. A user-chosen password using the same number of minutiae would be 33 characters in length. A user-chosen PIN would be 50 characters in length. Figure 6 shows that the randomly generated password does not increase in estimated character length at the same rate as the user-chosen password and PIN. The rationale behind this is that users are not likely to choose all of the possible values equally in the key space (94-character alphabet for passwords, 10-character alphabet for PINs). Therefore, each additional character in the estimated length for the password or PIN accounts for less entropy than if the values were chosen at random. A greater number of estimated characters are needed to achieve the same amount of entropy in a user-chosen password or PIN compared to a randomly chosen password.

The results for hypothesis four showed a lower calculated value of entropy in this study compared to that of previous work. However, it can be said that the entropy value calculated in this study is more precise because it is based on observed frequencies of occurrences of minutiae. Furthermore, the methodology for estimating character lengths for passwords and PINs developed by NIST was used to cross-compare the number of minutiae present and the estimated password and PIN character length. Due to their relation with the U.S. National Standards and government initiatives such as e-government, the NIST methodology provides further credibility to the results of this study.

C. Summary of Findings

Almost all of the groups tested had statistically significant results, which led to the rejection of the null hypothesis for hypotheses one through three. In total, 170 of the groups tested were statistically significant. These results show that, overall, the variables of minutiae occurrence, type, and angles are statistically significant 97.14% of the time based on location in the two-dimensional image plane.

III. CONCLUSIONS

As interest in biometrics for user authentication in both physical and logical access control environments grows, it is necessary to understand how biometrics relates to traditional authentication principles. This is important for not only biometric system developers and integrators, but also the end user of the system. Correlating Bishop's Authentication Model to that of the Components of the General Biometric System is one instance of accomplishing this task. The considerations for implementing biometric authentication are often driven by cost. Additionally, a comparative analysis of authentication mechanism effectiveness must also be contemplated. To do this, a straightforward comparison of secrets and biometrics is needed. Currently, means of comparing authentication mechanisms based on fundamentals are scarce, which has even led to the exclusion of biometrics for some environments, as shown in NIST SP800-63 [8], which stated, "Biometrics do not constitute secrets suitable for use in the conventional remote authentication protocols addressed in this document" (p.2). This work observed the frequency of occurrence of minutiae within a two-dimensional plane and utilized the concepts of Information Theory [2]. The result was a representation of biometric entropy in the same way entropy of secrets is represented. This allowed for analysis of the relationship between fingerprint entropy and password and PIN entropy according to NIST SP800-63 Appendix A.1.

REFERENCES

- [1] Ratha, N. K., Connell, J. H., & Bolle, R. M. (2001). Enhancing security and privacy in biometrics-based authentication systems. *IBM SYSTEMS JOURNAL*, 40(3).
- [2] Shannon, C. E. (1949). Communication Theory of Secrecy Systems. *Bell Systems Technical Journal*, 28, 656-715.
- [3] O'Gorman, L. (2003). Comparing Passwords, Tokens, and Biometrics for User Authentication. *Proceedings of the IEEE*, 91(12).
- [4] The Computing Dictionary. (2006). Definition of Key Space. Retrieved November 9th, 2006, from [HTTP://COMPUTING-DICTIONARY.THEFREEDICTIONARY.COM/KEYSPACE](http://COMPUTING-DICTIONARY.THEFREEDICTIONARY.COM/KEYSPACE)
- [5] Statham, P. (2005, September 21, 2005). *Threat Analysis, How Can We Compare Different Authentication Methods?* Biometric Consortium Conference, Arlington, VA.
- [6] Bishop, M. (2003). *Computer Security: Art and Science*: Pearson Education.
- [7] INCITS M1. (2004). ANSI-INCITS 378-2004 - Finger Minutiae Format for Data Interchange (American National Standard ed.).
- [8] Burr, W., Dodson, D., & Polk, W. T. (2004). *NIST SP800-63 - Electric Authentication Guidelines for Federal Agencies*: National Institute of Standards and Technology.