A Lightweight Biometric Signature Scheme for User Authentication over Networks

Tansu Alpcan
Deutsche Telekom Laboratories
Berlin, Germany
tansu.alpcan@telekom.de

Sinan Kesici
Bogazici University
Istanbul, Turkey
sinan.kesici@gmail.com

Daniel Bicher
Technische Universitat Berlin
Berlin, Germany
bicher@cs.tu-berlin.de

M. Kivanç Mihçak
Bogazici University
Istanbul, Turkey
kivanc.mihcak@boun.edu.tr

Christian Bauckhage
Deutsche Telekom Laboratories
Berlin, Germany
christian.bauckhage@telekom.de

S. Ahmet Çamtepe
DAI-Labor
Technische Universitat Berlin
ahmet.camtepe@dai-labor.de

ABSTRACT

We introduce a lightweight biometric solution for user authentication over networks using online handwritten signatures. The algorithm proposed is based on a modified Hausdorff distance and has favorable characteristics such as low computational cost and minimal training requirements. Furthermore, we investigate an information theoretic model for capacity and performance analysis for biometric authentication which brings additional theoretical insights to the problem. A fully functional proof-of-concept prototype that relies on commonly available off-the-shelf hardware is developed as a client-server system that supports Web services. Initial experimental results show that the algorithm performs well despite its low computational requirements and is resilient against over-the-shoulder attacks.

Categories and Subject Descriptors

D.4.6 [**Security and Protection**]: Authentication

General Terms

Security, Design, Algorithms, Experimentation

Keywords

Biometric authentication, haptics, lightweight algorithms, experimental evaluation, system design

1. INTRODUCTION

The World Wide Web touches today almost all aspects our lives ranging from communication and information access to economic activities such as e-commerce and banking. As an example, the online sales in USA only are expected to reach $259 billion in 2007, with an increase of 18% over the previous year according to reports from National Retail Federation [9]. Authentication mechanisms play a crucial role in enabling all these activities on the Web. However, as the Web grows and its importance increases, the ubiquitous username and password paradigm does not satisfy the need of its users for more secure authentication methods. Biometrics is increasingly getting the attention of the security research community as a usable and secure alternative to classical methods. It offers a variety of alternatives and is widely applicable to a range of scenarios including authentication of users over networks and on the Web.

Biometrics measure individuals' unique physical or behavioral characteristics for identification or authentication purposes. Common physical biometrics include fingerprints; hand or palm geometry; and retina, iris, or facial characteristics. Behavioral characters include signature, voice (which also has a physical component), keystroke pattern, and gait [6]. In this paper, we focus solely on authentication through *online* handwritten signature verification which encompasses features such as speed, velocity, and pressure in addition to the finished signature's static shape (image). In contrast, offline signature verification relies only on the static image of the signature. The signature we consider does not have to match users own private (written) signature that s/he uses on paper documents or credit cards and it can be shorter for added convenience. Hence, we refer to it as ***paraph*** in order to distinguish it from the unique personal "paper signature". Based on this definition, we will use the terms paraph and signature interchangeably for the rest of the paper.

Paraph verification has several favorable characteristics such as user familiarity, ease of use, customization, and revocation. Customization and revocation are especially important for authentication on the Web or over network where multiple use cases, privacy, and identity theft are important factors. Furthermore, paraph is not as intrusive or private as other biometrics, e.g. iris or fingerprints. This is again an advantage for lightweight and Web usage.

1.1 Related Work

Considerable amount of prior work exists on authenticating users via signature verification. Since providing a comprehensive literature survey is outside the scope of this paper, we briefly give an overview of a few selected studies in this area. Feature extraction and distance measures play naturally a significant role in developing classification algorithms for signature verification. Various distance measures have been proposed for offline signature verifi-

SecureComm 2008 September 22 - 25, 2008, Istanbul, Turkey

cation in [8]. A combination of Bayesian classifiers and principle component analysis has been studied and its performace has been analyzed in [5]. In an earlier study [4] graphical passwords have been investigated as an alternative method to signatures and with a focus on mobile devices and PDAs. Again as an alternative, a method to generate biometric hash vectors from statistical features of online signatures has been proposed in [12]. Since it does not require storage of signature reference templates this approach has advantages in terms of "key" management and increases the security of the system. An international competition on online signature verification (SVC 2004) was held in Hong Kong China in 2004 where various machine learning methods were suggested and tested on the given dataset [14].

A statistical approach to off-line signature verification from a Bayesian perspective has been proposed in [7], which relies on Markov chain Monte Carlo (MCMC) algorithms. Threat models against biometric authentication and evaluation methods have been discussed in [2], where the authors have studied a generative attack models and show its effectiveness. In a related study [1], the authors have proposed a measure for biometric information based on the Kullback-Leibler distance or relative entropy between the feature distributions of individuals and entire population in the system. This measure is then used to compute the average information content of biometric authentication systems. More recently, a Hausdorff distance measure has been studied in [10] in the context of iris recognition.

An information theoretic approach to biometric systems has been proposed in [13] and subsequently extended in [11]. While our analysis is parallel to these studies, it differs from them by focusing on additive colored Gaussian noise as the source and channel model and its application to the practical setup of paraph verification.

The main contributions of this paper can be summarized as 1) study and experimental analysis of a scalable and easy-to-configure identity verification algorithm using data obtained from online signatures. The algorithm is based on a modified Hausdorff distance and is computationally cheap. Furthermore, it requires minimal training. 2) development of a lightweight, scalable, and easy-to-use system and a fully functional proof-of-concept prototype for authentication based on the modified Hausdorff algorithm introduced. The client-server system supports identity verification over networks, web services, and relies on commonly available off-the-shelf hardware. 3) A preliminary information theoretic model and capacity analysis of the proposed biometric authentication scheme which brings additional theoretical insights to the problem.

2. MODEL AND ALGORITHM

The nature of the available data and extraction of its features play a significant role in almost all classification problems and the one at hand is no exception. We focus here on a lightweight system for user authentication over networks that does not require any special hardware in most cases. It is interesting to note that most of the laptops shipping today are equipped with a touchpad from the company *Synaptics*, which enjoyed a 70 percent market share within the notebook touchpad segment in 2007. Therefore, we focus in this paper on Synaptics hardware and drivers for obtaining the paraph features. We next provide an overview of the data set collected and then present the specific matching algorithms that utilize them for biometric user authentication.

2.1 Data Collection

Due to the unique properties of the application scenario considered we have collected a new paraph database that contains specific set of features which enables a high level of personalization. Each paraph is captured from the end user via a Synaptics touchpad on a laptop. The user signs her/his paraph directly on the touchpad. Since they were told not to use their paper signatures, the users were given time to select a personal paraph for themselves. Then, the subjects were given enough time to be familiar with their paraphs so that their five paraphs are similar to each other, and hence simulating a normal usage scenario. Subsequently, the client program accesses to the touchpad driver and acquires the paraph data. Five genuine paraphs are collected intially from 59 subjects. The 10 selected physical features that are contained in each paraph data are summarized in Table 1 where *Mickey* refers to one unit of mouse motion as reported by the pointing device.

Table 1: Features

TimeStamp	in milliseconds
XRaw	the raw X coordinate reported by the device
YRaw	the raw Y coordinate reported by the device
ZRaw	the reported Z (pressure) value
W	width and state information reported (e.g. "width" of the finger touching the pad)
Xspeed	the speed computed from the difference between the current and previous X coordinates
Yspeed	the speed computed from the difference between the current and previous Y coordinates
Zspeed	the speed computed from the difference between the current and previous Z (pressure) values
XMickeys	The number of X Mickeys reported
YMickeys	The number of Y Mickeys reported

The data collected is processed afterwards by removing the gaps in user signature when the pressure level on the touchpad falls below a level in order to ensure reliability and normalizing it to the range $[0, 1]$ for each feature point. The resulting data can be represented by a matrix with the number of rows determined by the signature length and columns by the number of physical features selected (in this case 10).

While the authentication scheme proposed has unique properties, it is as vulnerable to usual attack types as any regular password-based scheme, e.g. trojans, viruses, etc. The emphasis here is on the lightweight nature of the system, i.e. a degree of security that at least matches the one provided by traditional passwords in daily usage while improving usability by not requiring the user to memorize random alphanumerical sequences. We conduct a basic experimental security study on the unique "behavioral" property of the system that distinguishes it from password-based authentication. We define for this purpose an "over the shoulder" attack where a malicious person tries to imitate the paraph after watching it from a distance of one meter at various angles to the user while maintaining line of sight to the touchpad without any obstruction. Five such individual forgeries are collected for 15 subjects.

2.2 Hausdorff Distance and Algorithm

The paraph data obtained from the users has unique properties. It can be represented as a matrix, yet in principle, its (row) size is not fixed. Training data collected shows that individual paraphs of even the same user can show significant variation among themselves. Since our objective in this paper is to develop a lightweight system that does not have heavy computational requirements both in terms of training and operation, we resort to a threshold-based scheme and choose a modified Hausdorff algorithm as the distance measure. This enables us to use the almost raw paraph data (with-

out e.g. interpolation) for authentication and develop a training scheme that is dependent only on the data supplied by the user. These favorable properties lead to a scalable and easy-to-deploy solution that is suitable for client-server deployments over networks.

The Hausdorff distance provides a (dis)similarity metric between two sets. One of its important features is that these sets do not have to be of the same size as it is the case in our problem. Let use define x as a sample point from a paraph data such that $x \in [0,\,1]^M$, where M is the number of features (dimensions) including time stamp. Then each paraph can be represented as a compact real set X. In this context we define the Hausdorff distance $d_H(X,Y)$ between two paraphs (non-empty sets) X and Y as

$$d_H(X,Y) := \max\left(\max_{x\in X} d(x,Y),\, \max_{y\in Y} d(y,X)\right), \qquad (1)$$

where $d(x,Y) := \min_{y\in Y} \|x-y\|_2$ and $\|x-y\|_2$ is the simple Euclidean distance on $\mathbb{R}^M$. Hausdorff distance can be described as the "maximum distance of a set to the nearest point in the other set", which we compute here in both directions for each set in order to make it symmetric, and hence a proper distance measure.

During our experiments we have observed that the classical Hausdorff distance leads to substandard results due to its tendency of "punishing" even the simplest variations between two paraphs. Therefore, we define a modified Hausdorff distance as

$$\begin{aligned} d_T(X,Y) := \quad & \max_{X,Y}\left(\frac{L1}{L2}\left[\alpha \max_{x\in X} d(x,Y) + \right.\right. \\ & \left.\left.(1-\alpha)\frac{1}{C(X)}\textstyle\sum_x d(x,Y)\right]\right) \qquad (2) \\ L1 = \quad & \min\left(C(X), C(Y)\right) \\ L2 = \quad & \max\left(C(X), C(Y)\right), \end{aligned}$$

where $\alpha \in [0,1]$ is a scalar, $C(\cdot)$ denotes the cardinality (number of elements) of its argument, and the distance measure is again made symmetric by taking the maximum over both sets, $\max_{X,Y}$. In this version, the mean distance of a set to the nearest point in the other set plays a role in addition to the maximum distance. Hence, we take into account similarities in addition to punishing deviations as well as the length of the signatures. The balance between the two terms in (2) is determined by the weight parameter α.

The training algorithm based on modified Hausdorff distance measure is simple to implement and scalable. Each new user enters personal paraph three times after which the distances between these are calculated. As one possible implementation, the median of them is taken as the user-specific threshold parameter. Depending on the system requirements, a fixed value can be added to or subtracted from this value in order to improve usability or security, respectively. During verification the paraph entered by the user is compared with each of these stored paraph samples via modified Hausdorff distance and a basic threshold scheme is used accordingly to make the authentication decision.

2.3 Experiments

We study the performance of the proposed authentication algorithm and provide standard metrics such as receiver (or relative) operating characteristic (ROC), equal error rate (EER) as well as false accept rate (FAR), and false reject rate (FRR). The ROC plots the values of FAR and FRR with respect to the threshold variable of the verification algorithm where as EER shows the rate (and threshold) at which both accept and reject errors are equal. For each curve, the thresholds are varied between zero and one by an increment of 0.01.

We first plot the ROCs for both the classical and modified Hausdorff distance measure-based verification algorithms in Figure 1.

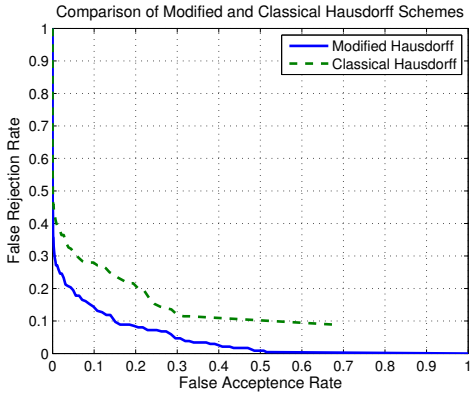

Figure 1: ROCs of the algorithm variants based on the classical and modified Hausdorff distance.

The modified Hausdorff measure leads to a Pareto improvement over the classical one in accordance with the discussion in Section 2. Hence, we use the modified version exclusively for the rest of the paper and choose the parameter α as 0.1.

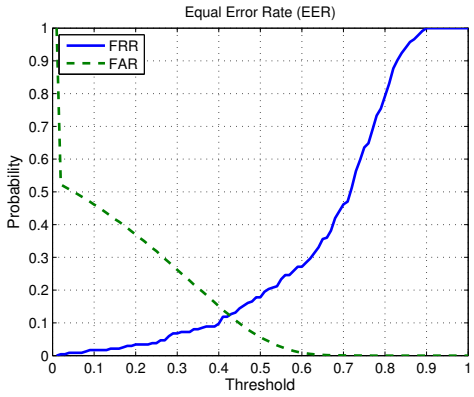

Figure 2: EER of the Hausdorff verification algorithm versus the threshold parameter.

The EER and ROC of the (modified) Hausdorff-based scheme with respect to the threshold parameter are shown in Figures 2 and 3, respectively. The results are promising especially considering the lightweight nature of the algorithm. Notice that, it is quite straightforward to improve upon them through "heavier" machine learning schemes such as support vector machines or boosting-based methods as it is done in the literature. However, our objective in this study is to maintain the focus on low computational requirements, scalability, and easy deployment.

2.4 Attack Analysis

We study the EER and ROC of the algorithm for over-the-shoulder attack data collected. The results are depicted in Figure 4. A comparison with Figures 2 and 3 indicate that our scheme is quite robust with respect to this type of attack. Since over the shoulder attacks are specifically targeting a unique property of our authentication scheme, the results observed are encouraging in the sense that the biometric approach does not have new vulnerabilities over the regular password-based ones . It is also an indicator that collected extra features which allow for personalization (such as pressure, speed, finger width) play a nontrivial role in the verification process.

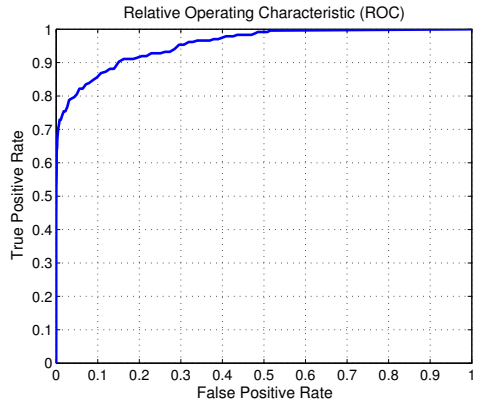

Figure 3: ROC of the Hausdorff verification algorithm versus the threshold parameter.

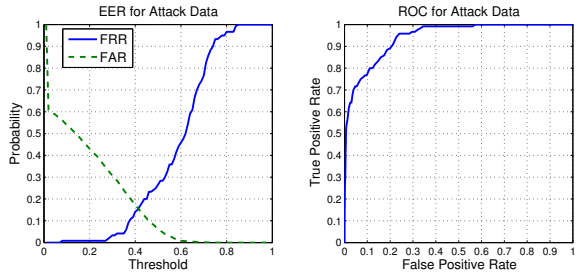

Figure 4: EER (left) and ROC (right) of the Hausdorff algorithm versus the threshold parameter for the attack data.

3. INFORMATION-THEORETIC ANALYSIS

In this section, we investigate an information-theoretic approach to the general problem of user authentication under some mild assumptions. The key point in our setup is to view a "user signature", i.e. the signal that represents the user's identity, as the *codeword* that represents the user whose identity can be viewed as the *message* that the underlying authentication system is trying to transmit to the verifier reliably. One key assumption, which is essential for our information-theoretic model, is that each time the user generates his/her signature, he/she produces a "slightly-corrupted version" of the original signature due to practical circumstances which is obviously the case for most biometric authentication mechanisms, including online signature verification which is the central scope of this paper. Under the aforementioned assumption, our information-theoretic development yields *maximum achievable rate* results and bounds thereof which determine the maximum number of users an authentication system can accommodate at asymptotic optimality. Note that, one remarkable consequence of our approach is that it can be applied to not only online signature verification systems, but also all user authentication mechanisms where the aforementioned assumption is applicable.

3.1 A Communication Theoretic Problem Formulation Of User Authentication

We treat each user's identity as the "message" that we would like to communicate to the receiver side. Each message (user ID) is represented by the original signature of the corresponding user. Thus, a user's signature is viewed as the "codeword" that represents the corresponding message (user ID). This codeword can presumably be a biometric feature or original online signature of the user. Now, in practice, whenever a user is challenged for authentication, he/she produces a slightly different variant of his/her original signature. This slight variation is due to practical circumstances of most user-authentication systems; for example, in a fingerprint-based user authentication system, each time a user produces a fingerprint, there will be a slight variation due to smearing effects, scanning imperfections, synchronization problems and alike; similarly, in an online signature-based authentication system, whenever a user tries to produce his/her original online paraph, the movement and the speed of the fingertip and the applied pressure are not going to be exactly the same as the ones that were used in the generation of the original online paraph. This naturally results in slight variations. In our communication-theoretic framework, such variations are modeled by a "lossy channel" which introduces disturbances to the original codeword. As a result, a slightly-corrupted codeword is obtained by the "receiver" (also termed as the "decoder" throughout the paper) which models the act of the authenticator or verifying mechanism. Given the corrupted version of an original codeword, the receiver makes a decision as to the origins of the codeword and produces an estimate of the original message (user ID).

Notation: In our developments, we follow a standard information-theoretic formulation notation. Boldface letters denote vectors; regular letters with subscripts denote individual elements of vectors. Furthermore, capital letters represent random variables and lowercase letters denote individual realizations of the corresponding random variable. The sequence of $\{a_1, a_2, \ldots, a_N\}$ is compactly represented by $\mathbf{a}_1^N$ (or equivalently by $\mathbf{a}^N$). The abbreviations "i.i.d." and "p.d.f." are shorthands for the terms "independent identically distributed" and "probability density function", respectively. Given a continuous random variable X, the corresponding p.d.f. is denoted $p_X(\cdot)$. Given a Gaussian vector $\mathbf{X}_1^N$ (i.e., $\{X_i\}_{i=1}^N$ have a multivariate normal distribution of dimension-N) with mean μ and covariance matrix Σ, we use $\mathbf{X}_1^N \sim \mathcal{N}(\mu, \Sigma)$ to denote the distribution. In this paper, we carry out our developments for continuous random variables; however all of our results can be extended to the discrete case with little or no difficulty.

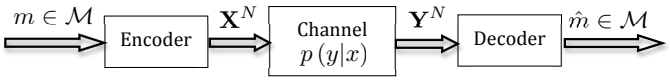

Figure 5: Block diagram representation of our communication-theoretic approach to the user authentication problem.

Our communication-theoretic model of the user authentication system is depicted in Fig. 5. We use $\mathcal{M} \triangleq \{1, 2, \ldots, M\}$ as the set of message indices, which represents the set of users of a biometric authentication system. Each element of $\mathcal{M}$ represents the identity of the corresponding user. The original signature of each user $m \in \mathcal{M}$ (also termed as the "codeword" for user m) is denoted by $\mathbf{X}^N$ (and hence assumed to be of length-N) [1]. In practice, the decoder receives the length-N signal $\mathbf{Y}^N$, which is a degraded version of the codeword $\mathbf{X}^N$. Such degradations are modeled by a channel, which is represented by the conditional probability density function $p\left(\mathbf{y}^N \mid \mathbf{x}^N\right)$. Given $\mathbf{Y}^N$, the receiver acts as a decoder and finds out an estimate $\hat{m} \in \mathcal{M}$ of m. For the biometric authentication application, $\hat{m}$ represents the detected user identity given a biometric signature. Accordingly, the corresponding error probability is given by $P_e = \Pr(\hat{m} \neq m)$, which should ideally be 0 (in practice as small as possible). Note that, unlike the clas-

[1]The dependence of each X^N on the corresponding m is omitted for the sake of notational convenience.

sical communication-theoretic setups, in our biometric authentication scenario the system designer has no freedom in designing the encoder as the users themselves implicitly construct the encoder by choosing their biometric signatures. Thus, in our setup the system designer aims to design the receiver so as to enhance the performance as much as possible, which is equivalent to minimizing P_e.

The *differential entropy* of the random vector $\mathbf{X}_1^N$ with p.d.f. $p_{\mathbf{x}^N}$ is denoted by $h\left(\mathbf{X}_1^N\right)$. Given the random vectors $\mathbf{X}_1^N$ and $\mathbf{Y}_1^N$, with the joint p.d.f. $p_{\mathbf{x}^N,\mathbf{y}^N}\left(\mathbf{x}^N,\mathbf{y}^N\right)$, the *conditional differential entropy* is $h\left(\mathbf{X}^N \mid \mathbf{Y}^N\right)$. The mutual information between two random vectors $\mathbf{X}^N$ and $\mathbf{Y}^N$ with joint p.d.f. $p_{\mathbf{x}^N,\mathbf{y}^N}\left(\mathbf{x}^N,\mathbf{y}^N\right)$ denoted by $I\left(\mathbf{X}^N;\mathbf{Y}^N\right)$. For further standard definitions and results from information theory we refer the interested reader to [3].

3.2 Application to Signature-based Biometric Authentication

For a multivariate-Gaussian channel with fixed input distribution, $I\left(\mathbf{X}^N;\mathbf{Y}^N\right)$ is the maximum number of bits that the encoder can reliably send to the decoder per input block $\mathbf{X}^N$. This, in turn, is equivalent to having a message alphabet $\mathcal{M}$ of maximal cardinality $2^{I\left(\mathbf{X}^N;\mathbf{Y}^N\right)}$ in case of reliable transmission for a single input block $\mathbf{X}^N$. In the biometric authentication scenario, the quantity of $2^{I\left(\mathbf{X}^N;\mathbf{Y}^N\right)}$ indicates the *maximum number of users* that the authentication system can identify reliably for a single input block $\mathbf{X}^N$. Accordingly, the quantity of $2^{I\left(\bar{\mathbf{X}}^N;\mathbf{Y}^N\right)} = \left(\frac{|\Sigma_X+\Sigma_Z|}{|\Sigma_Z|}\right)^{1/2}$ indicates an upper bound on the maximum number of users the authentication system can identify reliably for a single input block $\mathbf{X}^N$ [3].

In our experimental setup, we use the dataset described in Section 2.1 which contains paraphs from a total of 59 users each consisting of 10 features. To facilitate theoretical analysis, we adjust each paraph to a fixed length of 100. The features of these signatures are then decomposed into blocks of length-25, resulting in a total of 40 blocks. Each block-i is treated as a codeword $\left(\mathbf{X}^N\right)_i$, $N = 25$, $1 \leq i \leq 40$; the covariance matrix $\left(\Sigma_X\right)_i$ of each block-i is then found via empirical estimation utilizing all 59 realizations. In the next set of experiments, we collect 4 different realizations of each signature from every user, which results in a total of $59 \times 4 = 236$ realizations of the channel noise. The channel noise data is also adjusted to be of length-100 for each feature and subsequently decomposed into blocks of length 25. Then, the covariance matrix $\left(\Sigma_Z\right)_i$ of the channel noise for block-i is estimated empirically utilizing all 236 realizations. The empirical distribution of the "normalized signature data" (achieved via multiplying each block-i by $\left(\Sigma_X\right)_i^{-1/2}$) and the "normalized noise data" (achieved via multiplying each block-i by $\left(\Sigma_Z\right)_i^{-1/2}$) are shown in the left and right panels of Figure 6 (solid lines), respectively.

Note that, the resulting normalized data is both zero-mean with identity covariance matrix. We compare the resulting distributions with the corresponding zero-mean, unit variance i.i.d. Gaussian p.d.f. Figure 6 (dashed lines). The figures show that the noise distribution is sufficiently close to the Gaussian p.d.f. (which justifies the multivariate-Gaussian channel model); the signature distribution, on the other hand, slightly deviates from the jointly-Gaussian p.d.f.

In the last step of our experiments, we compute the quantity $\bar{R}_i \stackrel{\triangle}{=} I\left(\left(\bar{\mathbf{X}}^N\right)_i;\left(\mathbf{Y}^N\right)_i\right)$ for each block-i, $1 \leq i \leq 40$. The maximum number of users our system can reliably accommodate is bounded above by $2^{\sum_{i=1}^{40}\bar{R}_i}$, which is found to be 2^{322}, which is quite high. On one hand, this large number is a good indicator

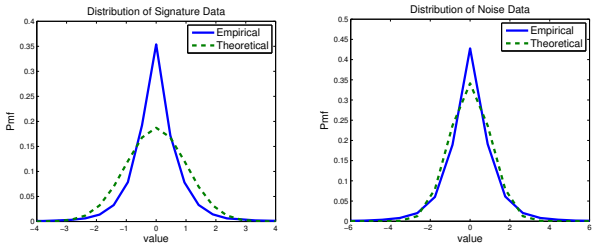

Figure 6: Empirical distributions of the normalized signature data (left) and the normalized noise data (right). Respective Gaussian distributions are shown with dashed lines. We observe that noise data closely resembles the Gaussian distribution.

of the optimum reliability of the proposed online-paraph system; on the other hand, the tightness of the bound needs to be quantified for the true distribution $p_{\mathbf{x}^N}\left(\mathbf{x}^N\right)$), which constitutes part of our future research.

4. SYSTEM DESIGN AND PROTOTYPE

A proof-of-concept prototype, which supports authentication over networks or the Internet, is implemented for testing the algorithms and demonstration purposes. The prototype is based on a client-server architecture and programmed in Java language. The client is a Java swing application. On the server-side Apache Tomcat 6 is used as the application server and servlet container. Since it relies on the touchpad driver to obtain paraph data, the client-side component cannot be embedded into the browser and has to be a stand alone application. We address this problem and simplify deployment by resorting to a unique solution within the Java framework called *Java Web Start*. Java Web Start allows application software for the Java Platform to be started directly from the Internet using a web browser. Unlike Java applets, Web Start applications do not run inside the browser but in a special sandbox. One chief advantage of Web Start over applets is that they overcome many compatibility problems with browsers' Java plugins and different JVM versions. Hence, it simplifies client-side software, deployment and updates through direct downloading and seamless execution of the client program from the server. By relying on the Java Network Launching Protocol (JNLP), Java Web Start thus enables large scale distribution, installation, and update of client-side software. These properties match perfectly with our objective of developing a lightweight system.

The training and verification algorithms are implemented on the server-side for additional security. In addition, Java Web Start runs in a sandboxed environment that can be secured by requiring the JAR (Java archive files that contain code) files to be signed using public key encryption. The communication between clients and the server can be encrypted for added security through a combination of public-key cryptography and symmetric encryption utilizing session keys. A detailed overview of the architecture is depicted in Figure 7.

5. CONCLUSION

We have investigated a lightweight biometric solution for identity verification over networks utilizing online signatures. The algorithm proposed is based on a modified Hausdorff distance and has favorable characteristics such as low computational cost and minimal training requirements. In addition, we have studied a information theoretic model for capacity and performance analysis

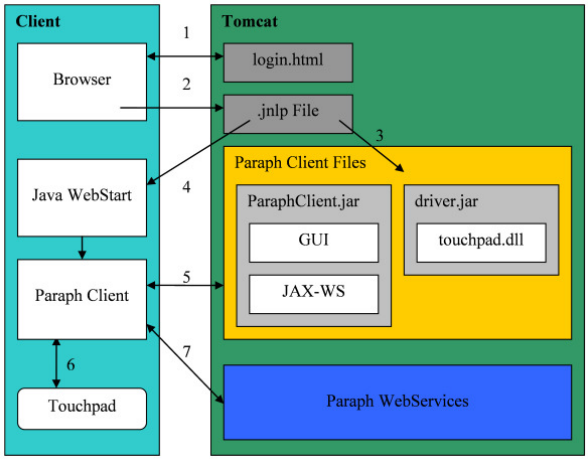

Figure 7: The client-server system architecture for authentication over network. The login page (1) contains a link to the *.jnlp* file (2), which refers to the necessary *.jar* files on the server (3). The server initiates Java Web Start on the client (4) which downloads the program (5). This can access the touchpad (6) and communicates with the paraph Web Services (7).

of this biometric authentication scheme which brings theoretical insights to the problem at hand.

A fully functional proof-of-concept client-server system that relies on commonly available off-the-shelf hardware has been developed. Since it relies on the touchpad driver to obtain paraph data, the client-side component cannot be embedded into the browser and has to be a stand alone application. We have addressed this problem and deployment-related issues by resorting to a unique solution within the Java framework called *Java Web Start*, which allows application software for the Java Platform to be started directly from the Internet using a web browser. The training and verification algorithms have been implemented on the server-side for additional security. Furthermore, the identity verification can also be offered as a web service.

Initial experimental results utilizing a dataset generated for this purpose are promising and show that the algorithm performs well despite its low computational and training requirements. They also indicate that our scheme is quite robust with respect to over the shoulder attacks where attackers, with a direct view of the users touchpad from a short distance, forge the paraphs they observe.

The information theoretic analysis and models introduced constitute a first step towards an analytical framework for biometric authentication systems. We believe that our interdisciplinary research efforts bringing together machine learning, security, and information theory communities, open a new research avenue. In another direction, the lightweight algorithms proposed can be improved upon in several ways while maintaining the focus on commonly available off-the-shelf hardware. For example, we are currently working on a biometric verification scheme that combines face recognition (utilizing ubiquitously available laptop cameras) with the methods in this paper to improve verification accuracy and robustness.

Acknowledgments

The research of Tansu Alpcan, Christian Bauckhage, and Daniel Bicher is supported by the Deutsche Telekom AG. The research of M. Kivanc Mihcak was partially supported by TÜBİTAK Career Award No 106E117 and TÜBA-GEBIP Award.

6. REFERENCES

[1] A. Adler, R. Youmaran, and S. Loyka. Towards a measure of biometric information. In *Proc. of Canadian Conf. on Electrical and Computer Eng. CCECE '06*, pages 210–213, Ottawa, Canada, May 2006.

[2] L. Ballard, F. Monrose, and D. Lopresti. Biometric authentication revisited: understanding the impact of wolves in sheep's clothing. In *Proc. of 15th USENIX Security Symposium*, pages 3–3, Berkeley, CA, USA, 2006. USENIX Association.

[3] T. M. Cover and J. A. Thomas. *Elements of Information Theory*. Wiley, 1991.

[4] I. Jermyn, A. Mayer, F. Monrose, M. K. Reiter, and A. D. Rubin. The design and analysis of graphical passwords. In *Proc. of 8th USENIX Security Symposium*, volume 8, pages 1–1, Berkeley, CA, USA, 1999. USENIX Association.

[5] A. Kholmatov and B. Yanikoglu. Biometric authentication using online signatures. In *Proc. of 19th Int. Symp. on Computer and Information Sciences, ISCIS 2004*, Lecture Notes in Computer Science, pages 373–380. Springer Berlin / Heidelberg, Antalya, Turkey, October 2004.

[6] S. Liu and M. Silverman. A practical guide to biometric security technology. *IT Professional*, 3(1):27–32, 2001.

[7] I. W. McKeague. A statistical model for signature verification. *Journal of the American Statistical Association*, 100(469):231–241, March 2005.

[8] J. R. Parker. Simple distances between handwritten signatures. In *Proc. of 15th Int. Conf. on Vision Interface VI'2002*, pages 218–222, Calgary, Canada, May 2002.

[9] F. Research. State of retailing online: a shop.org study. `http://www.nrf.com`, 2007.

[10] N. Sudha and Y. Wong. Hausdorff distance for iris recognition. In *Proc. of 22nd IEEE Int. Symp. on Intelligent Control ISIC 2007*, pages 614–619, Singapore, October 2007.

[11] P. Tuyls and J. Goseling. Capacity and examples of template-protecting biometric authentication systems. In *Proc. of ECCV 2004, Int. Workshop BioAW 2004*, volume 3087/2004 of *Lecture Notes in Computer Science*, pages 158–170. Springer Berlin / Heidelberg, Prague, Czech Republic, May 2004.

[12] C. Vielhauer, R. Steinmetz, and A. Mayerhöfer. Biometric hash based on statistical features of online signatures. In *Proc. of 16th Int. Conf. on Pattern Recognition (ICPR'02)*, volume 1, page 10123, Washington, DC, USA, 2002. IEEE Computer Society.

[13] F. Willems, T. Kalker, J. Goseling, and J.-P. Linnartz. On the capacity of a biometrical identification system. In *Proc. of IEEE Int. Symp. on Information Theory, ISIT 2003*, page 82, Yokohama, Japan, July 2003.

[14] D. Yeung, H. Chang, Y. Xiong, S. George, R. Kashi, T. Matsumoto, and G. Rigoll. SVC2004: First international signature verification competition. In *Proc. of 1st Int. Conf. on Biometric Authentication, ICBA 2004*, volume 3072/2004 of *Lecture Notes in Computer Science*, pages 16–22. Springer Berlin / Heidelberg, Hong Kong, China, July 2004.