

# Extreme Value Analysis for Mobile Active User Authentication

Pramuditha Perera and Vishal M. Patel  
Department of Electrical and Computer Engineering  
Rutgers, The State University of New Jersey  
94 Brett Road, Piscataway, NJ 08854

pramuditha.perera@rutgers.edu, vishal.m.patel@rutgers.edu

**Abstract**—In this paper, we propose to improve the performance of mobile Active Authentication (AA) systems in the low false alarm region using the statistical Extreme Value Theory (EVT). The problem is studied under a Bayesian framework where extremal observations that contribute to mis-verification are given more prominence. We propose modeling the tail of the match distribution using a Generalized Pareto Distribution (GPD) in order to make better inferences about the extremal observations. A method based on the mean excess function is introduced for parameter estimation of the GPD. Effectiveness of the proposed framework is demonstrated using publicly available unconstrained mobile active authentication datasets. It is shown that the proposed EVT-based method can significantly enhance the performance of traditional AA systems in the low false alarm rate region.

## I. INTRODUCTION

Mobile devices have become an integral part of human life at an increasing rate in the recent years. Largely owing to the expeditious growth in the fields of semi-conductor electronics and communications, modern mobile devices are well equipped with higher data rates, faster clock speeds and a colossal amount of internal memory. As a consequence, it has transformed its role from being a simple communication tool to become the perfect personal assistant that manages multiple personal needs including communication, networking, finance and entertainment. Subsequently, modern mobile devices contain a range of sensitive personal information including personal photographs, messages, contacts, bank account numbers, passwords, etc. Therefore, mobile device theft in the present context could directly lead to information theft which has a huge intangible cost associated.

Industry surveys have shown that 10% of phone theft victims have claimed to lost confidential company data, 9% of the victims have experienced identity theft, and 12% of the victims have experienced fraudulent charges on their accounts [18]. The total cost associated with information theft is substantial considering that 2.1 million cases of phone thefts were reported in 2015 in US alone [8]. In this context, mobile user authentication is paramount in safeguarding information security of the user. Traditional mobile authentication systems are essentially based on a one-time user verification methods such as a PIN number, a swipe

This work was supported by US Office of Naval Research (ONR) Grant YIP N00014-16-1-3134.

978-1-5090-4023-0/17/\$31.00 ©2017 IEEE

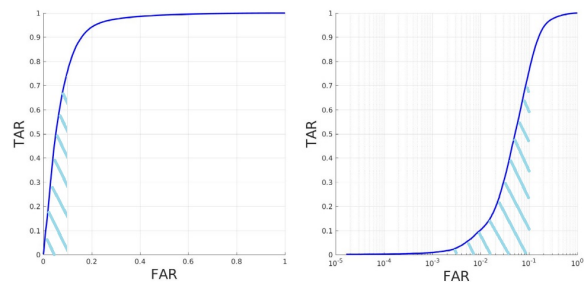


Fig. 1. Operable region of an AA system. To comply with security needs, an AA system should operate with a very low FAR rate. This operable region is shaded in the standard ROC curve (Left). Performance in this region is more observable in the log scaled plot (Right). In this paper we discuss how performance of this region can be improved using the statistical EVT.

pattern or a password. However, recent studies reveal that such one-time verification schemes (also know as explicit authentication) have failed to prevent fraudulent access to devices [1], [5].

As an alternative, both biometrics and security research communities have developed techniques for active authentication (AA) on mobile devices [21]. These methods essentially make use of physiological and behavioral biometrics using built-in sensors and accessories such as gyroscope, touchscreen, accelerometer, orientation sensor, and pressure sensor to continuously monitor the user identity. For instance, physiological biometrics such as face can be captured using the front-facing camera of a mobile device and can be used to continuously authenticate a mobile device user [13], [19], [7], [9], [24], [22]. On the other hand, sensors such as gyroscope, touchscreen and accelerometer can be used to measure behavioral biometric traits such as gait [34], [14], touch gestures [10], [28] and hand movement [29] transparently<sup>1</sup>.

In principle, the primary goal of an authentication system is to ensure information security through intruder prevention. In order to prevent intrusions, an authentication mechanism should operate with a very low degree of false alarms as shown in Figure 1. However, how well previously proposed AA systems perform in the low false alarm region has not extensively studied in the literature. Nevertheless, for AA to

<sup>1</sup>Note that the terms continuous authentication, active authentication, implicit authentication, and transparent authentication have been used interchangeably in the literature [21].

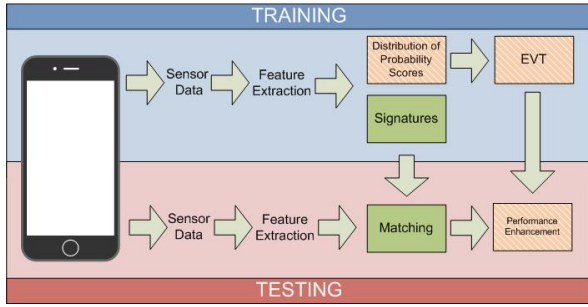


Fig. 2. Overview of the proposed system. Non shaded blocks represent typical components of an AA system. Shaded components are the proposed additions for performance enhancement.

be a viable alternative for traditional explicit authentication, this remain an important criterion. In this paper, we focus on the performance of AA systems at the low false alarm region (from 0.001 to 0.1) and we present a performance enhancement mechanism in this region for unimodal mobile AA systems based on the statistical Extreme Value Theory (EVT).

Figure 2 gives an overview of the proposed EVT-based AA system. A typical AA system extracts features of a probe and compares them against the enrolled features. In the proposed system, distribution of the probability scores are also obtained in the enrollment phase. Tail of the probability score distribution is modeled using the EVT and is used together with the similarity score generated in the standard AA system to enhance the performance of the standard system. It should be noted that the proposed mechanism is independent of sensors and features used in the underline AA system. Therefore, any existing AA system can be extended by incorporating the proposed performance enhancement scheme.

The statistical EVT has been previously used to modal the occurrence of extreme events in finance [15], hydrology [30] and novelty detection [23], [6], [17], [12]. In particular, several recent works have proposed using EVT for computer vision applications such as score-level fusion [32], open-set recognition [16], and visual attributes [26]. However, to the best of our knowledge this is the first attempt of using EVT to enhance the performance of an AA system.

This paper is organized as follows. Traditional and AA user verification systems are described in Section II. Section III briefly introduces the statistical EVT. EVT-based performance enhancement methods for unimodal AA systems are presented in Section IV. Experimental evaluations are described in Section V. Finally, concluding remarks are presented in Section VI with a brief summary and discussion.

## II. USER VERIFICATION IN THE AA SYSTEMS

### A. Traditional Verification Problem

Taking face as an example, in the traditional face verification problem, the verification task determines whether the given probe matches the enrolled faces or not. Mathematically, based on a set of face images  $X$  (the gallery) belonging to a person, the verifier generates a distance figure and uses

hard thresholding to assign a label  $L(x)$  to a given image  $x$  (the probe) based on whether it belongs to the same person or not as in,

$$L(x) = \begin{cases} \text{Non-match,} & \text{if } d_X(x) \geq \delta \\ \text{Match,} & \text{if } d_X(x) < \delta, \end{cases} \quad (1)$$

where,  $d_X(\cdot)$  is the distance with respect to the distribution of  $X$  and  $\delta$  is a predetermined threshold. Performance of the verification system solely depends on the produced distance distributions for positive and negative probes. Distributions of positive and negative probes are henceforth referred to as match distribution and non-match distribution, respectively. Placement and shape of match and non-match distributions of a subject in the UMDAA-01 dataset [9] is shown in Figure 3(a). An ideal verification system is characterized by a zero overlap between its match and non-match distributions. However, in practice these distributions often overlap.

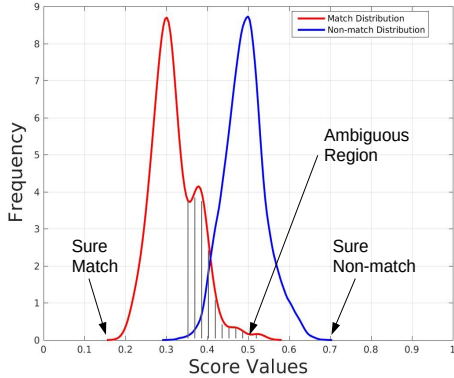
### B. Verification in the AA Systems

By definition, AA falls under the problem of user verification where the requirement is to verify the claimed identity of the current user. A typical AA system would operate on a set of obtained features and a specific comparison rule to evaluate the similarity between gallery and probe images. In what follows, we use face biometric to describe concepts, challenges and modifications in AA. It should be noted that applicability of the discussion is valid for all AA systems based on other sensor data as well.

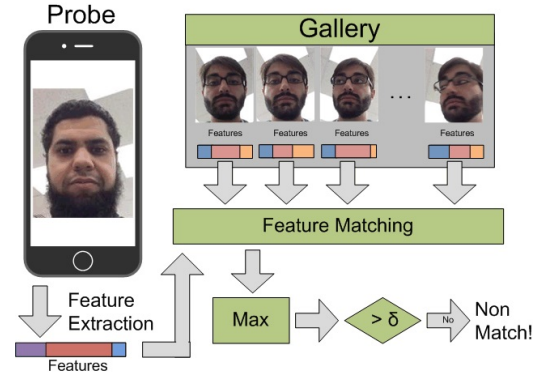
An overview of a typical AA system is shown in Figure 3(b). At the initiation of the device, the owner of the device is prompted to enroll herself and features of enrolled data are evaluated in advance to form a set of signatures. In the authentication phase, when a probe is presented, his/her likeliness is sensed through the mobile sensor (a face picture in this case) and relevant features are extracted. Extracted features are compared against the enrolled signatures independently and the distance score generated by the best match is considered. Probe is authenticated by thresholding the distance score.

At the first glance, it is evident that the AA process is very similar to an identity verification scenario. Since verification is a very well studied problem in the literature, it may appear that solution to AA problem is trivial. However, AA itself is a unique problem with a number of inherent constrains. The unique nature of the application poses a number of challenges.

- 1) Sensor data acquired from mobile devices are often partial or noisy. For example, images of the user captured from the front-facing camera could be clipped depending on the device orientation. Therefore, the similarity between gallery and a true probe could be considerably low at times.
- 2) Only a set of samples belonging to the true user is available during enrollemnt. However, complete information is absent as to how all non-users look like. Therefore, mobile AA problem can be viewed as an



(a)



(b)

Fig. 3. (a) Match and non-match distributions corresponding to a user in the UMDAA-01 dataset. (b) An overview of a typical mobile AA system.

open set verification problem [27], [25]. This constraint disrupts the possibility of producing a classifier based on discrimination (for example a classifier based on SVM cannot be trained since complete information on non-matches are unknown).

- 3) Limited memory and processing power in mobile devices require the use of easily extracted features and simple matching algorithms. Hence, mobile AA systems cannot use more complex and sophisticated classification schemes.

Due to these constraints, practical classification methods employable for AA problems are inherently poor in nature. Therefore, there exists a considerable overlap between score distributions generated by true and false probes. This phenomenon is clearly observed in Figure 3(a). In this paper, we refer to scores that falls in this region as *ambiguous scores*. In the context of AA, ambiguous scores predominantly appear at the tail of the match distribution. This is the key contributor in performance degradation in mobile AA systems.

From a practical perspective, when the obtained distance score is considerably large, it is safe to decide that the probe image does not belong to the enrolled user. On contrary, when a very low distance score (closer to 0) is obtained, user's authenticity is apparent. However, when distance score is in the ambiguous score region, the decision is not so straight forward. Therefore, the ambiguous score region plays a significant role in the performance of any AA system. One of the biggest challenges in the AA systems is to reject intruder faces when they produce an ambiguous score.

Due to the poor nature of the AA classifiers, performance enhancement of the classifier even by the slightest amount is greatly valued. The most straight forward approach to enhance recognition performance is to reject scores that fall in to the ambiguous region and take an alternative sensor reading instead [20]. However, in an event where an intruder's likeliness is relatively similar to that of the enrolled user, the similarity score generated will continue to fall in the ambiguous score region. Therefore, a more scientific approach is desired to differentiate ambiguity scores generated

by intruders from that of the legitimate user to enhance the continuous authentication performance.

### III. EXTREME VALUE-BASED ENHANCEMENT

#### A. Empirical Tail Estimation

Motivated by the significance that the tail region of a match distribution holds, we propose a performance enhancement scheme based on extreme values of the match distribution. For an observation  $x$ , since matched scores are bound to take lower score values, probability for a probe  $x$  been matched is given by  $1 - F_X(x)$ , where  $F_X$  is the cumulative distribution function of the match score distribution. If the distribution of the extreme values  $F_Z$  is known in advance, a two-fold similarity assessment of the probe can be done in order to incorporate information about the tail (ambiguous region). If all values above  $u$  are regarded as ambiguous, let  $\alpha = F_X(u)$ . Then, probability of been matched  $P_M(x)$  can be re-written as,

$$P_M(x) = \begin{cases} 1 - (\alpha + (1 - \alpha) \times F_Z(x - u)), & F_X(x) > \alpha \\ 1 - F_X(x), & \text{O.W.} \end{cases} \quad (2)$$

However, most of the time distribution of the extreme values are not known in advance. In addition, what values fall under extreme values themselves are distribution specific. Therefore, it is necessary to infer these information about the match distribution prior to such a modification. An intuitive way of estimating the match distribution is to use the training data itself for estimation. Let us assume that the training data (enrollment data) of the legitimate user is divided in to two portions. If one portion serves as signatures of the user, then the other portion can be used to construct the matched distribution of the user. A high percentile of the distribution (95 % is commonly used to estimate the tail [4]) can be used to determine a probability boundary that defines the tail of the distribution. For example, given a set of data  $x_1, x_2, \dots, x_n$ , if  $1 - \alpha$  is the fraction of data that belongs to the tail, the boundary  $u$  that defines the tail would be

$$u = x_{(\text{round}(n\alpha))}, \quad (3)$$

where  $x_{(i)}$  is the  $i^{\text{th}}$  order statistic of  $x_1, x_2, \dots, x_n$ . The tail distribution can be estimated by the distribution of all  $x > u$ . With this formulation in hand, one could expect to obtain better results through (2). However, such empirical tail estimation methods do not generalize well due to the low number of extreme events present in a finite training set (for a typical AA dataset, this number ranges from 30-100). As a result, an empirical formulation over fits the tail distribution thereby producing an inapt estimate.

Due to this limitation, we introduce the statical EVT-based modeling scheme to estimate the tail distribution of the match samples instead of using the raw empirical data.

## B. Extreme Value Theory

1) *The Fisher-Tippett Theorem*: Consider  $n$  number of independent identically distributed (i.i.d.) samples  $X = \{x_1, x_2, \dots, x_n\}$  drawn from the same unknown continuous distribution with a CDF of  $F_X(x)$ . If the maximum of the samples are  $M_n$ , where  $M_n = \max(X)$ , there exists a sequence of real numbers  $(a_n, b_n)$  such that  $a_n > 0$  for all  $n$  and a distribution  $H(x)$  such that

$$\lim_{n \rightarrow \infty} P\left(\frac{M_n - b_n}{a_n} \leq x\right) = H(x). \quad (4)$$

The limit distribution  $H(x)$  takes the form of a Generalized Extreme Value (GEV) distribution given by

$$H(x; \zeta, \mu, \sigma) = \begin{cases} \exp(-[1 + \zeta(x - \mu)/\sigma]^{-1/\zeta}), & \text{if } \zeta \neq 0 \\ \exp(-\exp^{-(x-\mu)/\sigma}), & \text{if } \zeta = 0, \end{cases}$$

where,  $1 + \zeta(M_n - \mu)/\sigma > 0$  [2], [4], [12]. The parameters  $\mu$ ,  $\sigma$  and  $\zeta$  correspond to scalar, tendency and tail index of the distribution, respectively. Larger tail index results a CDF with a fat tail and when it is zero, the CDF reduces to a Gumbel distribution. When the tail index is negative and positive the distribution reduces to Weibull and Frchet distributions, respectively. The asymptotic maximal distribution can be estimated without any assumptions on the underline CDF  $F_X(x)$ . In its original form Fisher-Tippett Theorem is formulated focusing on discrete batches of data. In order to extend its applicability to continuous data, Picklands, Balkema and de Haan formulation was later introduced.

2) *Picklands, Balkema and de Haan Formulation*: Consider a set of samples from  $X$  that exceed a sufficiently high threshold  $u$ . If  $F_U$  is the cumulative distribution of the excess of  $X$  over  $u$  such that

$$F_U(x) = P(X - u \leq x | X > u), \quad (5)$$

where  $x > 0$ , then the CDF of  $F_U$  can be approximated using a Generalized Pareto Distribution (GPD)

$$G(x; \xi, \beta) = \begin{cases} 1 - (1 + \frac{\xi x}{\beta})^{-1/\xi}, & \text{if } \xi \neq 0 \\ 1 - e^{-x/\beta}, & \text{if } \xi = 0, \end{cases}$$

such that  $-\infty < \xi < \infty, 0 < \beta < \infty, x > 0$  and  $\xi x > -\beta$  [2], [4], [12]. When  $\xi = 0$ , GPD reduces to an exponential distribution with mean  $\beta$ . When  $\xi > 0$  and  $\xi < 0$ , the GPD takes the form of Pareto distribution and Pareto II distribution, respectively.

3) *Parameter Selection for GPD*: The main challenge that arises in using GPD for tail distribution is the ambiguity in selecting the parameter  $u$ . According to Picklands, Balkema and de Haan result, it is stated that selecting a sufficiently large  $u$  would ensure that the tail distribution can be approximated using a GPD. However, a method to select a specific value for parameter  $u$  is not specified. In practice, heuristic values such as 80%, 90% and 95% have been used in the literature [16]. In this section, we present an automated method to estimate the threshold  $u$  based on the mean excess function (MEF) [11].

For the random variable  $X$ , the mean excess function  $M(u)$  is defined as

$$M(u) = E[X - u | X > u], \quad (6)$$

provided that  $E[X]_+ < \infty$ . For a given  $m$ , the number of i.i.d. samples  $X$ , the empirical mean excess function  $\hat{M}(u)$  for parameter  $u$  can be empirically calculated as follows [11]

$$\hat{M}(u) = \frac{\sum_{i=1}^n (X_i - u) I(x_i > u)}{\sum_{i=1}^n I(x_i > u)}, \quad (7)$$

where  $I$  is the indicator function. It should be noted that the mean excess function stays finite for all  $X$  when  $\xi < 0$ . If the random variable  $X$  is from a GPD, the mean excess function is linear with a positive slope [2], [11]. It can be shown that the mean excess function takes the following form [2]

$$M(u) = \frac{\beta}{1 - \xi} + \frac{\xi}{1 - \xi} u. \quad (8)$$

This theory essentially provides us a way to estimate the parameter  $u$  while ensuring that the defined tail could be approximated using a GPD. The proposed algorithm for parameter estimation is depicted in Algorithm 1. Here, all entries of the training data sample are sorted in the ascending order and the MEF at each point is calculated. Maximum number of extremal data points that would lie linearly in the mean excess plot is determined by assessing how good a straight line could fit the MEF. Parameter  $u$  is estimated based on this information.

## IV. EVT-BASED AA PERFORMANCE ENHANCEMENT

In this section, we present the performance enhancement framework based on EVT for the unimodal AA systems (AEVT). It is possible to invoke EVT to the AA problem given that data beyond a threshold  $u$  is considered. The threshold  $u$  is decided based on the score distribution of the training data according to Algorithm 1.

1) *Training*: As shown in Figure 2, sensor data captured during the enrollment phase are used to derive features for training. The entire training process is presented in Algorithm 2. A portion of enrollment features are used to construct user specific signatures. The total obtained feature set is then compared with the obtained signatures one at a time and the lowest obtained distance value is recorded.

Once the distribution of the lowest scores is obtained, its tail is modeled using EVT. First, the tail bound of the distribution (parameter  $u$ ) is found by using Algorithm 1

**input** : Set of training data  $x$  of size  $n$   
**output**: Estimate for parameter  $u$ , tail proportion  $\alpha$

For each data point calculate the Meand Excess  
Function (MEF);

```

Sort x ;
for i ← 1 to n - 1 do
  M = [ ]; // Initialize M;
  for j ← 1 to n do
    if  $x_j > x_i$  then
      //Store values greater than the
      considered entry ;
       $M = M \cup x_j$  ;
    else
      end
  end
  MEF[i] = mean(M)/length(M) ;
end
For different number of extremal points fit a line;
for k ← 1 to n/10 do
  Sort MEF retaining k largest values of MEF and
  x ;
  [m, c] = FitLine({MEF}, {x});
  //Find the MSE at data points for the fitted line ;
  E[k] ← MSE(line(m, c, x), x);
end
//Determine bound of the tail based on best possible
line fit;
u = argmink{E} ;
//Find value of  $\alpha$  ;
N = [ ];
for j ← 1 to n do
  if  $x_j > u$  then
    //Store values greater than u ;
    N = N ∪ xj ;
  else
  end
end
 $\alpha = \frac{n - \text{length}(N)}{n}$ ;

```

**Algorithm 1:** Parameter estimation for GPD.

on the distribution. Then, based on the estimated parameter, tail of the distribution is extracted and parameters of the corresponding GPD is found through Maximum Likelihood estimation. In the case when the Maximum Likelihood process fails for a GPD, the tail of the distribution is modeled using an exponential distribution.

2) *Testing*: When probe data is obtained, as illustrated in Figure 2, same features as in the enrollment phase are extracted from the probe. Extracted probe features are compared against the obtained signatures in the enrollment phase and the resulting minimum distance score is recorded. This score is then enhanced by the means of update equation (2). The entire testing procedure is presented in Algorithm 3.

**input** : Set of training data  $x$  of size  $n$   
**output**: GPD distribution parameters  $\xi$  and  $\beta$  for the tail distribution

```

Divide data into two portions randomly;
Signatures ← Rand(x,m);
foreach element x[i] do
  for j ← 1 to m do
    |  $d[j] \leftarrow \text{distance}(x[i], \text{signature}[j])$  ;
  end
;
//Get the minimum matching distance;
D[i] = Min(d);
//Parameters for GPD is found using Algorithm 1;
[alpha, u] = Algorithm1(D);
//Select data points over the threshold u ;
M = []; for j ← 1 to n do
  if  $x_j > u$  then
    //Store values greater than the considered
    entry ;
     $M = M \cup x_j$  ;
  else
  end
end
//Fit a GPD for the tail ;
[ $\xi, \beta$ ] = gpdfit(M, u);

```

**Algorithm 2:** Training procedure for the AA system.

## V. EXPERIMENTAL RESULTS

We evaluated the performance of the proposed method using four publicly available unconstrained AA datasets - MOBIO face dataset [19], UMDAA-01 face dataset, UMDAA-01 touch gesture dataset [9] and BTAS-2013 touch gesture dataset [28]. The following features and matching methods were used to generate scores from these datasets:

- 1) UMDAA-01 face dataset: LBP features and cosine distance
- 2) UMDAA-01 face dataset: HOG features of facial components [7] and cosine distance
- 3) MOBIO face dataset: LBP features and cosine distance
- 4) MOBIO face dataset: facial attribute-based features [24] and cosine distance
- 5) UMDAA-01 touch gesture dataset: 27-dimensional features [33] and SVM with RBF kernel
- 6) BTAS 2013 touch gesture dataset: 28-dimensional features [28] and SVM with RBF kernel

We use the scores obtained using the specified feature-matching pairs as the baseline for comparisons. The Receiver Operating Characteristic (ROC) curves are used to measure the performance of EVT-based enhancement method (denoted by EVT), automated EVT-based enhancement method (denoted by AEVT) and raw scores (denoted by original). Experiments carried out on the UMDAA-01 face dataset showed that the best possible performance for EVT is obtained when  $\alpha = 0.95$  for the dataset. Based on this result, we use  $\alpha = 0.95$  to construct the EVT curves in

**input** : Probe data  $y$ ,  $m$  number of Signatures from training,  $\xi, \beta, \alpha$ , CDF of Signatures  $F_Y$   
**output**: Probability score  $s$

```

for  $j \leftarrow 1$  to  $m$  do
  |  $d[j] \leftarrow \text{distance}(y, \text{signature}[j])$  ;
end
//Get the minimum matching distance;
 $D = \text{Min}(d)$ ;
if  $D > u$  then
  | //If  $y$  lies in the tail, use EVT ;
  |  $s = 1 - (\alpha + (1 - \alpha) \times \text{gpcdf}(y, \xi, \beta))$ ;
else
  | //Else use the original score ;
  |  $s = 1 - F_Y(y)$  ;
end

```

**Algorithm 3:** Testing procedure for a probe

all experiments. Experiments described below show that the automatic threshold selection-based AEVT closely resembles the performance of EVT.

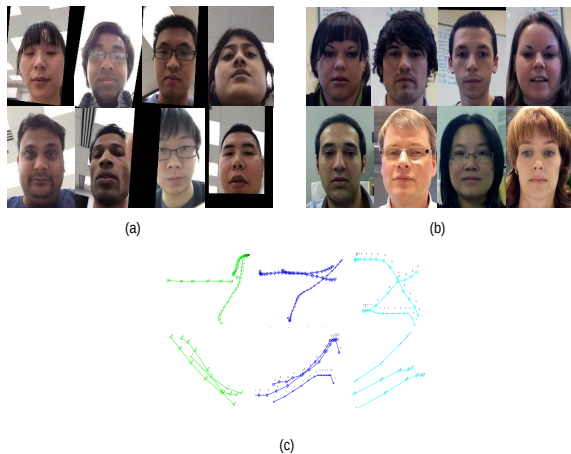


Fig. 4. Sample data from (a) UMDAA-01 face dataset, (b) MOBIO face dataset, and (c) UMDAA-01 touch gesture dataset.

### A. MOBIO Dataset

The MOBIO dataset [19] contains videos of 152 subjects taken across two phases where each phase consists of six sessions each (See Figure 4(b)). Videos in this dataset are acquired using a standard 2008 Macbook laptop computer and a NOKIA N93i mobile phone. Following the protocol defined in [24], video frames of the 12th session were considered as the enrollment data and video frames of all other sessions were used as probes. We conducted our experiments on the laptop image data based on the LBP features and the facial attributes. Obtained ROC curves for each cases are illustrated in Figure 5(a) and (b). As can be seen from these ROC curves, the proposed EVT-based method induces a significant performance improvement on the benchmark AA systems considered. Moreover, the automated parameter selection method has performed better than when a fixed parameter (where  $\alpha = 0.95$ ) was used. The True Accept

Rate (TAR) values for a series of False Accept Rate (FAR) values are tabulated in Table I. It is evident from Table I, that in general, the performance of the automated parameter selection based on the MEF is generally on par with the performance when the parameter is handpicked manually.

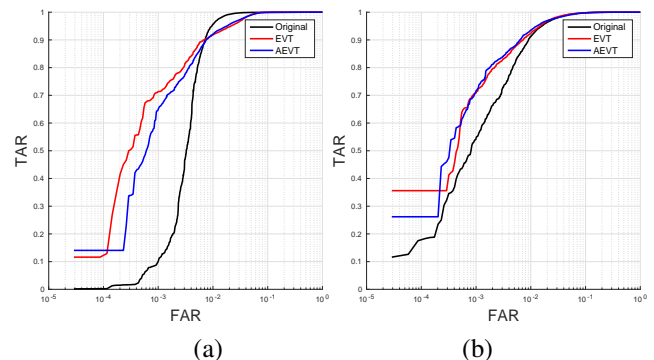


Fig. 5. Performance evaluation on the MOBIO dataset. (a) Results corresponding to the LBP features. (b) Results corresponding to the facial attributes.

### B. UMDAA-01 Dataset

The UMDAA-01 dataset [9] consists of images and touch gestures of 50 individuals taken from an iPhone 5 device across three sessions performing five tasks including an enrollment task. Sample face and touch data from this dataset are shown in Figure 4 (a) and (c), respectively.

1) *Face Dataset*: Following the protocol defined in [9], video frames of the enrollment task were used as enrollment data and frames of all other tasks were used as probes for testing. The ROC curves obtained using the attribute features and the LBP features along with the improvement induced by the EVT are shown in Figure 6. TAR values for a series of FAR values are also tabulated in Table I.

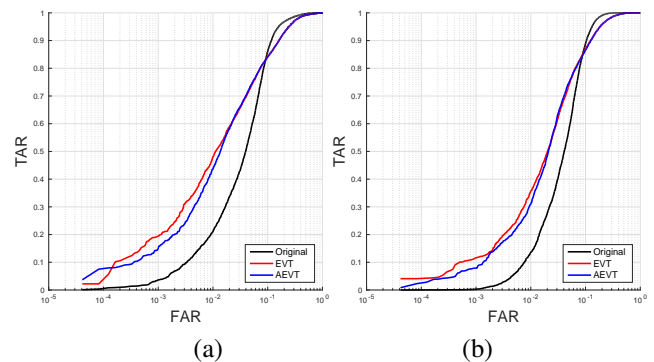


Fig. 6. Performance evaluation on the UMDAA-01 face dataset. (a) Results corresponding to the LBP features. (b) Results corresponding to the HOG features.

It can be seen from these experiments that the proposed method has increased the TAR values for very low FAR values. Practical AA systems operate in very low FAR regions facilitate the security of the system [3]. Therefore, induced improvement in this region would enhance the usability of the user while maintaining the system security. Moreover, the automated parameter selection method has performed

FAR%	TAR %																	
	MOBIO						UMDAA-01						BTAS2013					
	Face Attributes			Face LBP			Face HOG			Face LBP			Touch Gestures			Touch Gestures		
	[24]	EVT	AEVT	[31]	EVT	AEVT	[7]	EVT	AEVT	[31]	EVT	AEVT	[33]	EVT	AEVT	[28]	EVT	AEVT
0.05	5.45	<b>61.9</b>	46.2	43.4	<b>60.5</b>	59.0	0.05	<b>10.2</b>	6.85	1.67	<b>15.6</b>	12.2	13.8	59.2	<b>60.7</b>	7.79	<b>37.7</b>	35.4
0.1	9.73	<b>71.3</b>	64.5	54.9	71.3	<b>71.8</b>	0.29	<b>11.7</b>	7.9	3.67	<b>19.2</b>	15.8	14.1	<b>63.0</b>	62.5	9.81	41.8	<b>42.8</b>
0.25	34.1	<b>78.1</b>	75.3	68.7	81.5	<b>83.0</b>	2.01	<b>17.3</b>	15.8	8.01	<b>28.1</b>	22.4	14.6	<b>67.3</b>	66.3	14.6	48.2	<b>48.7</b>
0.5	77.2	<b>87.3</b>	84.7	81.3	87.3	<b>87.8</b>	5.87	<b>23.9</b>	21.1	13.9	<b>36.3</b>	31.3	19.2	<b>69.4</b>	68.5	23.0	<b>55.5</b>	54.5
1	<b>95.7</b>	91.9	92.0	91.3	92.3	<b>93.5</b>	13.7	<b>35.7</b>	31.0	21.1	<b>47.7</b>	44.6	25.6	<b>72.5</b>	71.7	33.6	61.5	<b>61.8</b>

TABLE I  
TABULATION OF THE TAR-FAR VALUES FOR THE EXPERIMENTS CONDUCTED IN THIS PAPER.

on par with the EVT method based on manual selection of parameters. Therefore, it can be concluded that the automated parameter estimation method provides a reasonable estimate for practical applications.

2) *Touch gesture dataset*: Each touch gesture is represented by a 27-dimensional feature vector as described in [33]. Half of the touch gestures corresponding to each user were used to construct signatures. The remaining data were used for testing. A one class SVM with RBF kernel ( $\gamma = 0.1$ ) was used to train a classifier for each user. In addition a window of size 12 was used when a probe was matched to obtain a distance figure as described in [33]. Shown in Figure 7 are the ROC curves obtained for this experiment. In contrast to earlier experiments, there is a significant improvement obtained by the proposed method compared to the raw score-based benchmark. This is mainly due to the fact that there is a significant overlap between the match and non-match distributions as compared to face datasets.

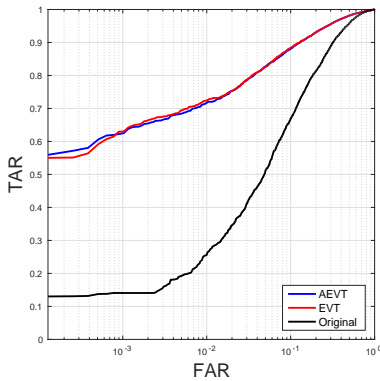


Fig. 7. Performance evaluation on the UMDAA-01 touch gesture dataset.

### C. BTAS2013 Touch Gesture Dataset

The touch gesture dataset introduced in [28] contains swipes from 190 users collected across two sessions. Data collection has been done on a Google Nexus S device running on Android 4.0. Similar to the UMDAA-01 experiment, a 28 dimensional feature was extracted [28] from each swipe and half of the swipe features were used to construct a

user specific signature <sup>2</sup>. A one class SVM based on the RBF kernel ( $\gamma = 10$ ) was used to produce the matching scores for each probe. The ROC curves corresponding to different methods are shown in Figure 8. The corresponding TAR values for a selected set of FAR values are tabulated in Table I. As can be seen from these results, both EVT and AEVT has outperformed the benchmark by a significant margin for low FAR values.

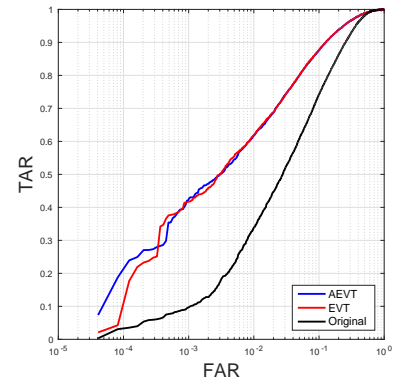


Fig. 8. Performance evaluation on the BTAS2013 touch gesture dataset.

## VI. CONCLUSION

We presented a performance enhancement mechanism based on the statistical EVT for the existing unimodal mobile AA systems. Statistical EVT was used to model the tail of the score distribution along with the introduced parameter selection mechanism. It was shown that the proposed method improves the performance of the existing face and touch gesture-based AA systems. This improvement is more significant when there is a considerable overlap between the original match and non-match distributions. Moreover, it was shown that the proposed method does not depend on the underlying features or the classifiers used - it only depends on the generated score values. As a result, the performance of our method can be even further improved by using better features and classifiers to generate the matching scores. In the future, we will examine how EVT can be used for multi-modal fusion applications in AA.

<sup>2</sup>There is an additional feature corresponding to the pressure sensor in this dataset. Hence, the feature dimension is 28 compared to 27 in the UMDAA-01 dataset.

## REFERENCES

- [1] A. J. Aviv, K. Gibson, E. Mossop, M. Blaze, and J. M. Smith. Smudge attacks on smartphone touch screens. In *Proceedings of the 4th USENIX Conference on Offensive Technologies*, pages 1–7, 2010.
- [2] Y. Bensalah. *Steps in Applying Extreme Value Theory to Finance: A Review*. Bank of Canada, 2010.
- [3] C. Braz and J.-M. Robert. Security and usability: The case of the user authentication methods. In *Proceedings of the 18th Conference on L'Interaction Homme-Machine*, pages 199–203, 2006.
- [4] J. B. Broadwater and R. Chellappa. Adaptive threshold estimation via extreme value theory. *IEEE Transactions on Signal Processing*, 58(2):490–500, 2010.
- [5] G. D. Clark and J. Lindqvist. Engineering gesture based authentication systems. *IEEE pervasive computing mobile and ubiquitous systems*, 14(1):18–25, 2015.
- [6] D. A. Clifton, S. Hugueny, and L. Tarassenko. Novelty detection with multivariate extreme value statistics. *Journal of signal processing systems*, 65(3):371–389, 2011.
- [7] D. Crouse, H. Han, D. Chandra, B. Barbello, and A. K. Jain. Continuous authentication of mobile user: Fusion of face image and inertial measurement unit data. In *International Conference on Biometrics*, 2015.
- [8] C. Deitrick. Smartphone thefts drop as kill switch usage grows, 2015.
- [9] M. E. Fathy, V. M. Patel, and R. Chellappa. Face-based active authentication on mobile devices. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2015.
- [10] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song. Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148, Jan 2013.
- [11] S. Ghosh and S. Resnick. A discussion on mean excess plots. *Stochastic Processes and their Applications*, 120.
- [12] X. Gibert-Serra, V. M. Patel, and R. Chellappa. Sequential score adaptation with extreme value theory for robust railway track inspection. In *IEEE International Conference on Computer Vision (ICCV) workshop on Computer Vision for Road Scene Understanding and Autonomous Driving (CVRSUAD)*, 2015.
- [13] A. Hadid, J. Heikkilä, O. Silven, and M. Pietikainen. Face and eye detection for person authentication in mobile phones. In *ACM/IEEE International Conference on Distributed Smart Cameras*, pages 101–108, Sept 2007.
- [14] F. Juefei-Xu, C. Bhagavatula, A. Jaech, U. Prasad, and M. Savvides. Gait-id on the move: Pace independent human identification using cell phone accelerometer dynamics. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 8–15, Sept 2012.
- [15] S. Kotz and S. Nadarajah. *Extreme value distributions*, volume 31. World Scientific, 2000.
- [16] W. J. S. Lalit P. Jain and T. E. Boulton. Multi-class open set recognition using probability of inclusion. In *The European Conference on Computer Vision (ECCV)*, September 2014.
- [17] D. L.B. Milstein and J.K. Wolf. Robust detection of binary signal in additive noise, using extreme value theory. *IEEE Transactions on Information Theory*, IT-15:370–375, 1969.
- [18] lookout.com. Phone theft in america, 2015.
- [19] C. McCool, S. Marcel, A. Hadid, M. Pietikainen, P. Matejka, J. Cernocky, N. Poh, J. Kittler, A. Larcher, C. Levy, D. Matrouf, J.-F. Bonastre, P. Tresadern, and T. Cootes. Bi-modal person recognition on a mobile phone: Using mobile phone data. In *IEEE International Conference on Multimedia and Expo Workshops*, pages 635–640, July 2012.
- [20] K. M.S. Hossain and V.V. Phoha. New impostor score based rejection methods for continuous keystroke verification with weak templates. In *IEEE Fifth International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, September 2012.
- [21] V. M. Patel, R. Chellappa, D. Chandra, and B. Barbello. Continuous user authentication on mobile devices: Recent progress and remaining challenges. *IEEE Signal Processing Magazine*, 33(4):49–61, July 2016.
- [22] P. Perera and V. M. Patel. Quickest intrusion detection in mobile active user authentication. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2016.
- [23] S. J. Roberts. Novelty detection using extreme value statistics. *IEE Proceedings-Vision, Image and Signal Processing*, 146(3):124–129, 1999.
- [24] P. Samangouei, V. M. Patel, and R. Chellappa. Attribute-based continuous user authentication on mobile devices. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, 2015.
- [25] W. J. Scheirer, L. P. Jain, and T. E. Boulton. Probability models for open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 36, November 2014.
- [26] W. J. Scheirer, N. Kumar, P. N. Belhumeur, and T. E. Boulton. Multi-attribute spaces: Calibration for attribute fusion and similarity search. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2012.
- [27] W. J. Scheirer, A. Rocha, A. Sapkota, and T. E. Boulton. Towards open set recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence (T-PAMI)*, 36, July 2013.
- [28] A. Serwadda, V. Phoha, and Z. Wang. Which verifiers work?: A benchmark evaluation of touch-based authentication algorithms. In *IEEE International Conference on Biometrics: Theory, Applications and Systems*, pages 1–8, Sept 2013.
- [29] Z. Sitova, J. Sedenka, Q. Yang, G. Peng, G. Zhou, P. Gasti, and K. S. Balagani. HMOG: A new biometric modality for continuous authentication of smartphone users. *CoRR*, abs/1501.01199, 2015.
- [30] R. L. Smith. Extreme value analysis of environmental time series: an application to trend detection in ground-level ozone. *Statistical Science*, pages 367–377, 1989.
- [31] A. H. T. Ahonen and M. Pietikainen. Face description with local binary patterns: Application to face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28:2037–2041, 2006.
- [32] R. M. Walter J. Scheirer, Anderson Rocha and T. E. Boulton. Robust fusion: Extreme value theory for recognition score normalization. In *The 11th European Conference on Computer Vision (ECCV)*, September 2010.
- [33] H. Zhang, V. M. Patel, M. Fathy, and R. Chellappa. Touch gesture-based active user authentication using dictionaries. In *2015 IEEE Winter Conference on Applications of Computer Vision*, pages 207–214, 2015.
- [34] Y. Zhong and Y. Deng. Sensor orientation invariant mobile gait biometrics. In *IEEE International Joint Conference on Biometrics*, pages 1–8, Sept 2014.