Smartphone Based User Verification Leveraging Gait Recognition for Mobile Healthcare Systems

Yanzhi Ren[1], Yingying Chen[1], Mooi Choo Chuah[2], Jie Yang[3]
[1]Department of ECE, Stevens Institute of Technology, Hoboken, NJ 07030
{yren2, yingying.chen}@stevens.edu
[2]Department of CSE, Lehigh University, Bethlehem, PA 18015
chuah@cse.lehigh.edu
[3]Department of CSE, Oakland University, Rochester, MI 48309
yang@oakland.edu

Abstract*—The rapid deployment of sensing technology in smartphones and the explosion of their usage in people's daily lives provide users with the ability to collectively sense the world. This leads to a growing trend of mobile healthcare systems utilizing sensing data collected from smartphones with/without additional external sensors to analyze and understand people's physical and mental states. However, such healthcare systems are vulnerable to user spoofing attacks, in which an adversary distributes his registered device to other users such that data collected from these users can be claimed as his own to obtain more healthcare benefits and undermine the successful operation of mobile healthcare systems. Existing mitigation approaches either only rely on a secret PIN number (which can not deal with colluded attacks) or require an explicit user action for verification. In this paper, we propose a user verification scheme leveraging unique gait patterns derived from acceleration readings in mobile healthcare systems to detect possible user spoofing attacks. Our framework exploits the readily available accelerometers embedded within smartphones for user verification. Specifically, our user spoofing attack mitigation scheme (which consists of three components, namely Step Cycle Identification, Step Cycle Interpolation, and Similarity Score Computation) is used to extract gait patterns from run-time accelerometer measurements to perform robust user verification under various walking speeds. Our experiments using 322 smartphone-based traces over a period of 6 months confirm that our scheme is highly effective for detecting user spoofing attacks. This strongly indicates the feasibility of using smartphone based low grade accelerometer to conduct gait recognition and facilitate effective user verification without active user cooperation.

I. Introduction

Smart phone, PDAs, tablets, etc. have become increasingly popular and play significant roles in our daily lives. In particular, with sensors that can be easily attached to smartphones and the plurality of sensors embedded within smartphones, the collected sensing data can be mined for the understanding of people's physical and mental health states. For example, barometer sensor can be attached to smartphones equipped with accelerometer and microphone to collect sensing data, which can be mined to uncover people's daily life activities [1]. Information about users' daily life activities and behaviors can further assist in the development of various emerging applications in the healthcare domain. For instance, walking activities and conversations extracted from collected sensor data can be used to predict users' physical and mental conditions [1].

However, such healthcare systems are vulnerable to user spoofing attacks in which an adversary can distribute his registered device to other users such that data collected from these users can be claimed to be his own. By doing so, the adversary can claim potential health benefits that are allocated to people with certain illnesses even though he may not have any illnesses. For instance, in the social community-based mobile healthcare systems [2] for facilitating epidemiology research and disease propagation control, an adversary can attract additional vaccine allocation by launching user spoofing attacks and thus undermine the regular operations of such mobile healthcare systems.

Nevertheless, mitigating user spoofing attacks is not an easy task. Most smartphones only offer user verification methods which rely on explicit manual entry of a secret PIN number. This is insufficient as many users only go through such a verification process once when a smartphone is switched on [3]. In addition, verification based on PIN numbers are not applicable to the cases when an adversary collude with other users. Recently, new techniques utilizing biometric characteristics such as fingerprints have been proposed for user verifications. However, fingerprint readers are not available on most smartphones, making it less suitable for mobile healthcare systems. Further, this technique also requires an explicit user action for verification, e.g., putting a finger on the fingerprint reader.

In this work, we exploit users' unique physical traits, which are hard to forge, to mitigate user spoofing attacks in mobile healthcare systems. Our design goal is to enable user spoofing attack detection without relying on explicit user cooperation or additional hardware such as a fingerprint reader. The basic idea is to utilize a user's gait pattern because a person's gait is often unique and can serve as a useful discriminator. We design our system to be robust by taking into the fact that several constant spontaneous sub-events embedded within gait patterns can uniquely characterize each user and are hard to imitate. A user may change his/her walking speeds, but the uniqueness embedded in each gait pattern remains unchanged. The presence of user spoofing attacks causes the newly identified gait patterns to be dramatically different from a user's normal gait patterns and hence such attacks can be

978-1-4799-0230-9/13/$31.00 ©2013 IEEE

detected. To the best of knowledge, our work is the first that utilizes gait information to detect user spoofing attacks in mobile healthcare systems.

Specifically, we design a user verification scheme leveraging gait patterns derived from accelerometer readings. Our framework employs readily available accelerometers embedded within smartphones instead of deploying additional hardware for user verification. While gait recognition via accelerometer sensors have been studied using sensors with high sampling rates (e.g., larger than 100 Hz in [4], [5]), we focus on addressing several unique challenges that one faces when using low grade accelerometers in mobile healthcare systems. First, low grade accelerometers (e.g. those in smartphones) has a lower sampling rate (e.g., lower or equal to 50 Hz), posing possible difficulty in capturing each user's unique gait patterns. Second, users' walking speeds may vary during the verification process, making it hard to identify step cycles accurately. Third, the user verification process should be able to complete with small number of measurements. To cope with these challenges, our gait pattern based user verification scheme consists of three components: *Step Cycle Identification*, *Step Cycle Interpolation*, and *Similarity Score Computation*.

During Step Cycle Identification, we utilize the fact that a user's gait patterns should be repeatable, and hence walking traces collected from a user should be highly correlated. We thus construct a template for each user's unique gait pattern by identifying the first distinguishable step cycle, and then utilize the high correlation between a user's step cycles to identify other step cycles within a trace. This approach can derive step cycles more accurately than other methods used in previous studies [4], [5], which identify step cycles by identifying local minimas repeatedly within a trace. Our Step Cycle Identification method has the adaptive learning capability to update a user's step cycle template using real-time feedback. A user's walking speed varies and is determined by many factors such as his/her health conditions, age, gender, environment, and so on. Our goal is to design a scheme that works well irrespective of what speed a user walks at when the accelerometer readings are collected.

Our Step Cycle Interpolation component helps to align identified step cycles of different lengths into normalized cycles of fixed length. This interpolation step allows our scheme to perform gait recognition robust to various walking speeds. Furthermore, a user's walking profile is constructed during a training process. And our scheme only needs the user to upload one accelerometer trace under any speed at its convenience for user profile construction, without requiring extensive uploading of multiple traces to cover different walking speeds. We summarize our main contributions as follows:

- By exploiting the correlation relationship inherent in a user's walking traces, our scheme can achieve more robust step cycle identification compared to previous studies even when a user's walking speeds vary.
- We use several techniques including automatic template update and step cycle interpolation to remove the impact of varying walking speeds, and preserve the uniqueness present in the user's gait pattern for accurate user verification.
- We collect 322 accelerometer traces from multiple users over a period of 6 months. The results show that our scheme is highly effective for detecting user spoofing attacks. Our technique can also be applied to other healthcare systems which utilize human sensing data.

The rest of the paper is organized as follows. We first present some recent researches which are related to our work in Section II. We then present the system model for our mobile phone enabled healthcare monitoring systems and the model for user spoofing attacks in Section III. Next, we present our gait based user verification scheme in Section IV. In Section V, we validate the feasibility of our proposed detection scheme through experiments conducted using real human walking traces. Finally, we conclude our work in Section VI.

II. Related Work

There has been active studies in designing schemes for detecting spoofing attacks in wireless networks [6], in which an adversary device masquerades the identity of a legitimate device. In the mobile healthcare systems we consider in this work, we are more interested in a user spoofing attack where an adversarial user passes his registered device to his friend to collect sensor data on his behalf, which is different from the device-identity spoofing attacks considered in wireless networks.

It may appear that cryptographic authentication schemes [7] are effective for thwarting user spoofing attacks. However, an adversary has access to all the security information stored in mobile devices and hence can pass the security checks easily. Schemes which utilize users' unique physical or physiological characteristics such as fingerprint [8] are attractive. These methods rely on additional hardware or require users to take explicit actions, and may not be suitable for mobile healthcare systems that constantly monitor users' behaviors.

There are also investigations which utilize users' behavioral traits such as gaits for user verification. In [9], a vision based gait recognition scheme has been proposed. The system uses several cameras for recording users' gait information. Some background segmentation techniques are used to extract features from recorded images to verify a user. In floor sensor-based approaches [10], the sensors are placed on the floor and when people walk on the floor, the identity of a user can be authenticated by the exerted force measured by the sensor. However, additional hardware such as cameras and floor sensors is also needed for these schemes to work but such hardware may not be always available.

Furthermore, there are schemes for gait recognition through wearable accelerometer sensors with high sampling rates [4], [11]. The main advantage of using a wearable accelerometer sensor for gait recognition is that it provides unobtrusive verification of a user's identity without requiring his explicit actions. [12] utilizes accelerometer sensors on smart phones to perform physical activity classification. However, they did not use gait patterns for user verification. In [5], the uniqueness of

the gait in case of foot motion with respect to the shoe attribute and axis of the motion is analyzed. It is not clear how their methods can deal with variable walking speeds. Our work is different in that we aim to employ gait information to detect the presence of user spoofing attacks in mobile healthcare systems. Our approach can extract the unique characteristics of a user's gait pattern from sensor data collected from low grade accelerometers embedded within smartphones, and is robust to users' varying walking speeds.

III. Framework Overview

In this section, we first describe our system model and the possible applications in mobile healthcare domain. We then show the adversary model of the user spoofing attack. We next provide an overview of our user verification framework.

A. System Model

We consider a healthcare monitoring system in which each user registering for its service is given a unique user identifier and a monitoring application which runs on a user's smartphone. This monitoring application can collect readings from embedded sensors within smartphones or external sensors attached to smartphones. Such sensor data will be analyzed to assess that user's physical activity levels or physiological conditions. For instance, a user's physical activity level can be assessed by monitoring his conversational activities, while measurements of heartbeats and blood pressure can be used to predict his psychological conditions [13]. Such sensing data collected by the monitoring device e.g. smartphones is sent to a system server. The server can then derive users' physical and mental well beings based on the rich information embedded in the sensing data. The system server then takes relevant followup actions based on such analysis, e.g. rewards those users who have weight problems for increasing their physical activity level. This type of mobile healthcare system is very useful as it utilizes the information derived from users' daily lives, instead of requesting manual reporting from a user which could be inaccurate and error-prone. Emerging applications enabled by such mobile healthcare systems include:

- The medical professionals from healthcare companies can monitor the health conditions of patients with heart diseases by monitoring their heartbeats. Based on patients' health conditions, the healthcare company can determine the frequency at which such patients should visit the doctors [13].
- Users' behavioral patterns and physical activity levels can be tracked by healthcare companies to facilitate early detection of signs of depression [1].
- Companies that sell healthcare related applications e.g."I Do Move" [14] can convince healthy food companies to provide discount coupons for users who use their healthcare applications by sharing some statistics, e.g. total number of walking steps collected by their applications, with these food companies.

B. Adversary Model

Such mobile healthcare system is vulnerable to user spoofing attacks, in which an adversary can collect the sensing data by passing his monitoring device, e.g. his smartphone, to another person for a short period of time, and upload the data collected by the other person in an attempt to gain more health benefits. For example, users, who registered at "I Do Move", upload their total walking steps to earn food discount coupons once the total steps reach a certain milestone. Malicious users can ask others to walk with their devices and hence reach the qualifying milestone faster. Furthermore, the adversary can distribute his monitoring device to others who have physical or psychological problems and share similar interest with him. The data collected from these people will be mistakenly regarded as being obtained from the adversary. Thus, the system may classify the adversary as a person with certain physical or psychological problems. Additional healthcare benefits or treatments will be mistakenly allocated to the adversary. Multiple users may collude to launch user spoofing attacks to fool the mobile healthcare system. This attack will significantly reduce the effectiveness of the healthcare management system and undermine the successful deployment of mobile healthcare applications since healthcare benefits will be given to the users who do not have physical or psychological problems, and are thus not entitled to receive the benefits.

C. User Verification Framework

Instead of using cryptographic-based authentication methods, we explore utilizing users' unique physical traits which are hard to forge for unobtrusive user verification in the mobile healthcare systems. We build a framework that utilizes user gait patterns extracted from accelerometer readings via smartphones. The framework can be implemented in two ways: *server-centric* and *user-centric*.

In the *server-centric* approach, pre-processed accelerometer readings together with sensor data collected from smartphones (and additional external sensors if any) are sent to a secure centralized server for user verification. The user verification is performed based on each user's profile constructed ahead of time. The user's profile contains its unique gait pattern. The details of the user profile construction are described in Section IV. The server will then decide whether to accept the sensor data collected from this user based on the successful outcome of the user verification process.

In the *user-centric approach*, the smartphone will be responsible for performing user verification. A user's profile will be constructed and stored in the smartphone. If the user verification fails, i.e., the user spoofing attack is detected, the sensor data collected from this user's mobile devices will not be reported back to the server.

IV. User Verification Based on Gait Patterns

In this section, we present our user spoofing attack detection scheme, which can be deployed in both server-centric and user-centric framework.

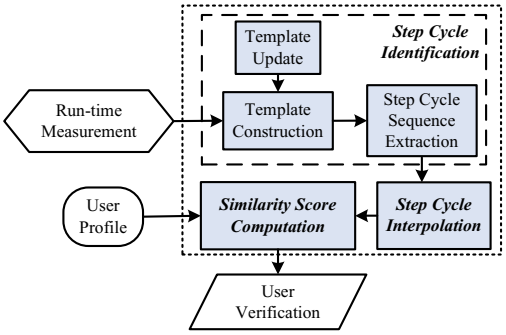


Fig. 1. Flow overview: components of user verification.

A. Challenge and Design Goals

The goal that leverages gait recognition using accelerometer readings on smartphones to mitigate the user spoofing attack is to be able to conduct the user verification without relying on additional infrastructures or explicit user actions. This allows a pure software solution. To fulfill such a goal in mobile healthcare systems, we need to deal with the following challenges.

Robust to Various Walking Speeds. Users' walking speed varies under different scenarios and environments. The gait recognition process should be robust to various walking speeds in order to facilitate an effective user verification.

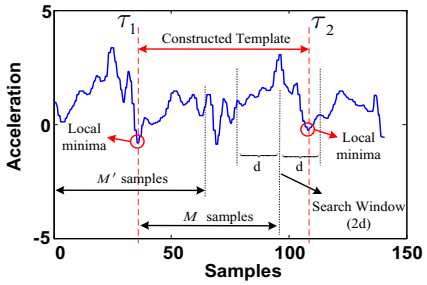
Reasonable Accuracy. Our framework leverages the accelerometers on smartphones with a lower sampling rate (e.g., 50Hz), which is about half the sampling rate of the regular accelerometer sensors. Our technique needs to achieve reasonable attack detection accuracy with readings collected from accelerometers within off-the-shelf smartphones.

Low Detection Latency. Our user verification scheme should be able to detect the presence of user spoofing attack with small number of measurements. In this way, the framework can avoid wasting computational cost spent on processing the sensor data reported from a user's mobile device for the corresponding healthcare needs.

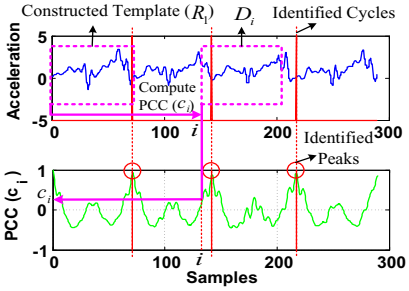
B. Scheme Overview

The basic idea underlying our user verification scheme is based on the observation that the gait pattern is unique for each person and differs between different people. When a user spoofing attack is present, the extracted gait pattern from the run-time accelerometer measurements from smartphones may differ significantly, and hence we make use of vertical acceleration collected from smartphones to perform user verification.

Our scheme, as shown in Figure 1, consists of three main sub-tasks: *Step Cycle Identification*, *Step Cycle Interpolation* and *Similarity Score Computation*. When the verification procedure starts, step cycle sequence needs to be first identified from the run-time accelerometer measurements. A step cycle template based technique is proposed to accurately capture the uniqueness embedded in each person's gait. This template can be dynamically updated when a user's physical/medical situation changes. The identified step cycle sequence is further interpolated to deal with various walking speeds when a user is at different environments. The user verification is then



(a) Template construction



(b) Step cycle sequence extraction by utilizing the template

Fig. 2. Illustration of Step Cycle Identification.

performed by calculating similarity scores between the final interpolated step cycle from the run-time measurements and the preconstructed user profile. A user's profile contains the user's gait pattern and is constructed when a user first submits its accelerometer measurements. The profile is obtained by utilizing the Step Cycle Identification and Step Cycle Interpolation. The Step Cycle Interpolation component allows robust user verification even when the user's walking speed during an run-time measurement is different from that in a user's profile. If a user distributes his device to another person, a lower similarity value will be obtained after the computation because the gait patterns between two people differ dramatically, and consequently the user spoofing attack is detected.

C. Step Cycle Identification

Human gait follows a cyclic pattern. In this work, the event that we use to mark the beginning of the step cycle is the heel strike of the swing leg [9]. At that moment, the person's feet are both on the floor and they are farthest from each other and the vertical acceleration of the impact can be observed as a local minima in the accelerometer readings. Thus, the step cycle can be identified by extracting the timestamps of the heel strikes. However, identifying the step cycle is challenging because the accelerometer readings can be distorted due to the irregular movement of the user's body or the change of walking speeds. The commonly used step cycle identification techniques [11], [12], [15] identify the gait cycles by conducting the typical cycle identification repeatedly in the traces. The problem is that if the cycle identification for one period is not accurate, the detection of the following periods will be affected. The detection errors are propagated and compounded throughout the whole cycle identification. For these reasons, we utilize the fact that the same user's gait patterns are unique and the consecutive step

cycles should present a high correlation in a collected walking trace. We thus extract a person's gait pattern as a template by identifying the first distinguishable step cycle. We then utilize the correlation relationship inherent in the same user's walking trace to search for the maximum correlation between the first distinguishable cycle and the rest of trace to derive the step cycle sequence.

1) Template Construction: Let $\{r(1), ..., r(N)\}$ be a sequence of N accelerometer measurements in the vertical direction from a smartphone and we assume the τ_k-th measurement is the first sample of the k-th step cycle. To construct the step cycle template, we need to find the first two consecutive local minimas $r(\tau_1)$ and $r(\tau_2)$ in the accelerometer readings which represent the beginning and the end of the first distinguishable step cycle $R_1 = \{r(l), \tau_1 \leq l < \tau_2\}$. To identify R_1, we assume the user's maximum and regular step cycles have approximate M' and M samples respectively according to the sampling frequency of the accelerometer. Thus, the beginning of the first step cycle τ_1 can be found by searching the minimum value from the first M' observations:

$$\tau_1 = \arg\min_l \left(r(l)\right), 1 \leq l \leq M' \quad (1)$$

We then search for the end of the first step cycle τ_2 by extending M samples from τ_1. Because the user's walking speed is unknown, the τ_2 can be determined by relaxing the searching range by d samples before and after the M samples:

$$\tau_2 = \arg\min_l \left(r(l)\right), \tau_1 + M - d \leq l \leq \tau_1 + M + d \quad (2)$$

Thus, $r(\tau_1)$ and $r(\tau_2)$ are the first two consecutive local minimas in the sequence of recorded accelerometer readings. The $L = \tau_2 - \tau_1$ consecutive samples of $R_1 = \{r(l), \tau_1 \leq l < \tau_2\}$ in sequence $\{r(1), ..., r(N)\}$ will then be used as a template to identify the rest of step cycles in the collected trace. We illustrate template construction in Figure 2 (a).

Provided with the knowledge about human walking patterns, we can then determine suitable values for M', M and d: The natural cadence of the human walking, irrespective of what speed he/she walks, is usually in the range of $[45, 65]$ step cycles/min [16] and we assume the sampling frequency of accelerometer is 50 samples per second. Thus, the number of samples in one step cycle is in the range $[46, 67]$ samples/step and each regular cycle contains about $(46 + 67)/2 \approx 56$ samples. With the aid of such clues, in this work, we empirically set the M' as the number of samples that a maximum step cycle has with $M' = 67$ samples and M as the number of samples a regular cycle has with $M = 56$ samples, respectively. The search range d is then set as the half of the difference between the maximum and minimum number of samples the step cycle has (i.e., $d = (67 - 46)/2 \approx 11$ samples).

2) Step Cycle Sequence Extraction: The template $R_1 = \{r(l), \tau_1 \leq l < \tau_2\}$ contains the samples of user's first step cycle. We assume there are S step cycles in the collected trace. To identify the subsequent step cycles R_k, $k = 2, 3, ..., S$, in a collected trace, we utilize the step cycle correlation

Algorithm 1 Step Cycle Identification

INPUT:
$Data = \{r(1), ..., r(N)\}$; *A sequence of accelerometer readings*
$R_1 = \{r(l), \tau_1 \leq l < \tau_2\}$; *The constructed template*
$L = \tau_2 - \tau_1$; *Number of samples in extracted template*
$counter = 0$; *Number of peaks in PCC sequence*
PROCEDURES:
for All $i \in [1, N - L]$ **do**
  $D_i = \{r(l), i \leq l < i + L\}$;
  $c_i = corr(R_1, D_i)$;
end for
for All $i \in [1, N - L - 1]$ **do**
  if $c_i > c_{i-1} \& c_i > c_{i+1} \& c_i > threshold$ **then**
    $counter = counter + 1$;
    $\tau_{counter} = i$;
  end if
end for
Return number of step cycles $S = counter - 1$
Return step cycle sequence $R_k = \{r(l), \tau_k \leq l < \tau_{k+1}\}, k = 1, ..., S$

inherent in a user's walking trace. The correlation among step cycles of the same person allows us to extract a user's step cycles accurately due to the fact that the correlation coefficient between two step cycles of the same person is robust to distorted readings caused by irregular movement of the user's body. Further, after examining the correlation coefficient between the template and subsequent step cycles, we can update the template dynamically based on the changes in user's walking speeds. This is because the step cycles should be highly correlated if the speed of the template step and subsequent steps are similar. Thus, a significant decrease in correlation coefficient between two step cycles indicates a large speed change. The template, consequently, should be updated based on the new speed (shown in the next step).

Figure 2 (b) illustrates the step cycle sequence extraction using Pearson correlation method [17]. To identify the subsequent step cycles, the template R_1 is slid across the recorded accelerometer readings and the Pearson correlation coefficients (PCC) between the template R_1 and the consecutive L samples in recorded accelerometer readings are calculated. The Pearson correlation coefficient (PCC) is a statistical method that measures the degree of the linear relationship between two given vectors. The Pearson correlation coefficient value ranges from -1 to 1. Correlation 1 and -1 means that there is a perfect positive/negative linear relationship between the two vectors. Specifically, given the template R_1 with length $L = \tau_2 - \tau_1$ and consecutive L samples $D_i = \{r(l), i \leq l < i + L\}, i = 1, ..., N - L$, from the recorded accelerometer readings $\{r(1), ..., r(N)\}$, the Pearson correlation coefficient is defined as:

$$c_i = corr(R_1, D_i) = \frac{\sum_{l=0}^{L-1} \left(\frac{r(\tau_1+l)-\bar{R}_1}{\sigma(R_1)}\right)\left(\frac{r(i+l)-\bar{D}_i}{\sigma(D_i)}\right)}{L-1} \quad (3)$$

where $\bar{R}_1$ ($\bar{D}_i$, resp.) and $\sigma(R_1)$ ($\sigma(D_i)$, resp) are the mean and standard deviation of R_1 and D_i. The values in Pearson correlation coefficient sequence $C = \{c_i, i = 1, ..., N - L\}$ increase and decrease successively, indicating similarity between the template R_1 and the segment D_i. The peaks arise periodically in PCC sequence C indicating good matches

Algorithm 2 Step Cycle Interpolation

INPUT:
$R_k = \{r(l), \tau_k \leq l < \tau_{k+1}\}, k = 1, ..., S$; *Identified step cycles*
$P = 300$; *Number of samples*
PROCEDURES:
for All $k \in [1, S]$ **do**
$\{r(1,k), ..., r(P,k)\} = Interpolation(\{r(\tau_k), ..., r(\tau_{k+1})\})$;
end for
for All $j \in [1, P]$ **do**
$\bar{r}(j) = \sum_{k=1}^{S} \frac{r(j,k)}{S}$
end for
Return interpolated step cycle $I = \{\bar{r}(1), ..., \bar{r}(P)\}$

between the template and the subsequent D_is. Thus, these periodical peaks can be used to identify the subsequent step cycles. The local maximas in C are detected and marked as beginning points of each walking step, which occur at the heel strikes of a swing leg. The algorithm of step cycle sequence extraction is provided in Algorithm 1.

In Figure 2(b), the blue line in the upper plot represents the accelerometer readings on smartphones. The green line in the lower plot represents the correlation coefficient sequence C computed between the step cycle template and each data segment D_i. The step cycles are identified by searching periodical local peaks in sequence C. The identified step cycle sequence R_k is:

$$R_k = \{r(l), \tau_k \leq l < \tau_{k+1}\}, k = 1, ..., S \quad (4)$$

3) Template Update: The length of the step cycle changes as the user's walking speed varies. With the help of the correlation coefficient between the template and the subsequent step cycles, we are able to tell when the user's speed changes, and update the template timely once the speed change is detected. The update decision may be system triggered. Particularly, the system automatically searches the peaks in the Pearson correlation coefficient sequence: if most of the peaks (e.g.,80%) in a past time period (e.g., a few minutes) are lower than a threshold (e.g., 0.8), the template update is triggered. A new template R_1 will be generated using the Template Construction scheme on newly collected accelerometer readings.

D. Step Cycle Interpolation

A user usually walks at different speeds in different scenarios such as taking a leisure walk after dinner or walking rapidly to catch a commuter train after work. Furthermore, the walking speed of a user during the run-time data collection process is most likely different from the speed when the user profile is constructed. The number of samples in step cycles varies as the user's walking speed changes. To deal with variable walking speeds, our framework preforms step cycle interpolation. This interpolation step allows us to perform robust user verification by directly measuring the similarity between the step cycle sequence in the user profile and the interpolated sequences obtained from run-time measurements under different walking speeds. More importantly, by using Step Cycle Interpolation, a user only needs to upload one accelerometer trace under any speed at its convenience for user profile construction without requiring extensive uploading of multiple traces to cover different speeds.

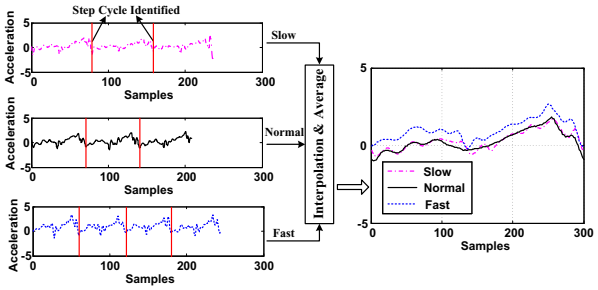


Fig. 3. Illustration of Step Cycle Interpolation for a user under three typical walking speeds: slow, normal and fast.

To perform step cycle interpolation, we align the extracted step cycle sequence to a reference step cycle with length P by using cubic spline interpolation [18], a fast, efficient and stable method of function interpolation. Further, we choose a large P (e.g., $P = 300$ samples) so that it is larger than any user's longest one step cycle irrespective of what speed the user walks. The step cycle sequence after interpolation are represented as:

$$N_r = \{r(1,k), ..., r(P,k)\}, k = 1, ..., S. \quad (5)$$

To capture the pattern of all the step cycles, we average over the interpolated step cycles. Thus, the final interpolated step cycle can be represented as $I = \{\bar{r}(1), ..., \bar{r}(P)\}$ with:

$$\bar{r}(j) = \sum_{k=1}^{S} \frac{r(j,k)}{S}, j \in [1, P] \quad (6)$$

The algorithm Step Cycle Interpolation is provided in Algorithm 2. Figure 3 shows an example on how the interpolated step cycle is extracted under different walking speeds for a specific user. In Figure 3, the collected acceleration readings under three representative speeds (i.e., *slow, normal, and fast*) are depicted in the left side of the figure. The detailed description of these three speeds are presented in Section V-A. The final interpolated step cycles corresponding to these three different speeds are shown in the right side figure. Before interpolation, it is hard to directly compare the step cycles under different speeds due to different lengths of the step cycles. After Step Cycle Interpolation, we find that the interpolated step cycles under three different speeds are highly correlated regardless of the walking speeds. This results is encouraging as it indicates a particular user's gait pattern is unique and not sensitive to a user's walking speeds.

E. Similarity Score Computation

The interpolated step cycle represents a user' gait pattern. Based on the foot motion, a step cycle can be further decomposed into several sub-events such as initial contact, loading response, and midstance [19]. The user's gait pattern in certain sub-events may remain constant while others change. The sub-events within a gait pattern remain constant should be treated more significantly since they can better represent the uniqueness of the user's gait pattern. Thus, to capture this observation in a quantitative way, we propose to use *weighted* Pearson correlation coefficients when computing the similarity between the extracted gait patterns and the user profile.

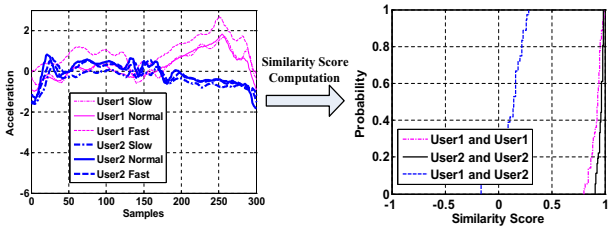


Fig. 4. An illustration of interpolated step cycles of user 1 and 2 under different walking speeds and CDF of their similarity scores.

We next calculate the weights from sub-events in a user's step cycle. Based on the interpolated step cycle sequence N_r, we first equally divide P samples in the interpolated cycle into B (e.g., $B = 6$) blocks: $\{P_n, ..., P_{n+1}\}, n = 0, ..., B-1$ with $P_0 = 1$ and $P_B = P$. Thus, the average sample distance over these blocks can be represented as: $Dist = \{\bar{d}_n, n = 0, ..., B-1\}$, where each $\bar{d}_n$ is defined as:

$$\bar{d}_n = \frac{\sum_{c=P_n}^{P_{n+1}} \sum_{\substack{k,l \in [1,S] \\ k \neq l}} |r(c,k) - r(c,l)|}{(S-1) \times S \times (P_{n+1} - P_n + 1)} \quad (7)$$

Each $\bar{d}_n$ in $Dist$ measures the average sample distance in the n-th block between each pair of S interpolated step cycles. Based on the sample distance, we define *weights* over these blocks as $\{w_n, n = 0, ..., B-1\}$, where each w_n is defined as: $w_n = 1/\bar{d}_n$.

We then define the similarity score between the interpolated step cycle obtained from run-time measurement $I_g = \{\bar{r}^g(1), ..., \bar{r}^g(P)\}$ and the user profile $I_h = \{\bar{r}^h(1), ..., \bar{r}^h(P)\}$ by computing weighted Pearson correlation coefficient with the weight as $\{w_n, n = 0, ..., B-1\}$:

$$\mathbf{C}(I_h, I_g) = \frac{\sum_{n=0}^{B-1} corr\big(\{\bar{r}^h(P_n), ..., \bar{r}^h(P_{n+1})\}, \{\bar{r}^g(P_n), ..., \bar{r}^g(P_{n+1})\}\big) w_n}{\sum_{n=0}^{B-1} w_n} \quad (8)$$

If the similarity scores are lower than a pre-defined threshold, the framework will declare the presence of the user spoofing attack for this particular user ID.

Feasibility Study. We study how the similarity scores change when acceleration readings are collected from different users under three typical walking speeds (i.e., *slow*, *normal*, and *fast*). We collect 6 traces per user with 2 traces per walking speed. Figure 4 plots the interpolated step cycles generated for these two users and the cumulative distributed function (CDF) of the similarity score. From the left subfigure in Figure 4, we observe the interpolated step cycles under three walking speeds within a particular user are very similar, while the interpolated step cycles between two users differ significantly. Furthermore, from the right subfigure, the similarity scores are high (larger than 0.8) for the same user regardless of walking speeds. Whereas the similarity scores reduce to $[-0.2, 0.3]$ between two users. These observations strongly confirm the feasibility of using our gait recognition based method to detect user spoofing attacks.

V. Performance Evaluation

In this section, we conduct experiments using accelerometer traces collected from volunteers using smartphones to evaluate the effectiveness of our approach in detecting user spoofing attacks. The following subsections detail our experimental methodology and results.

A. Experimental Methodology

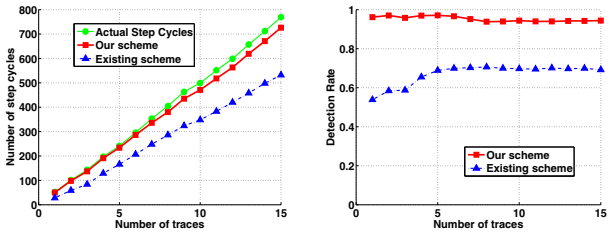
We use a HTC EVO smartphone equipped with accelerometer that supports 50 Hz sampling rate for data collection from volunteered users. Each HTC EVO smartphone runs Android operating system with 192 MB RAM and a 528MHz MSM7200A processor. The accelerometer readings are collected when the users are walking and then written into a log file on a smartphone. During the experiments, we let users put the phone in the hip pouch position. We defer the study of using traces collected from other body positions as our future work. Such phone position is natural since many people carry their smartphones in a similar position [11]. Besides, users of our system can be asked to put phones in such positions for verification purposes. Of the 3 dimensional accelerotmeter signals retrieved from the smartphone, only the acceleration in vertical direction is used. We experiment with three representative user walking speeds, namely *slow* (slower than 0.7m/s), *normal* (about 0.7m/s-1.1m/s), and *fast* (faster than 1.1m/s).

We conduct experiments using 322 accelerometer traces collected from 23 volunteers over 6 months to evaluate the effectiveness of our approach. A size of 23 users is also typical for user monitoring and verification studies [5], [20]. Unless otherwise stated, each collected trace represents accelerometer readings of a user walking for a time period of about 30 minutes. For each trace, the user is either walking at a constant speed (i.e., slow, normal, or fast) or walking at varying speeds by changing his/her speed every 10 seconds in their natural walking style. In total, we have 14 traces for each user collected over a period of six months: four traces per constant walking speed and two traces for varying speed. Unless otherwise stated, we choose a user's trace under normal walking speed to construct the user profile. The remaining traces under constant speeds and all the traces under varying speeds are used for testing.

We use a trace-driven approach to evaluate our gait pattern based user verification scheme in a personal computer, which corresponds to the situation in the server-centric framework. We defer the implementation of the user-centric framework as our future work. We emulate the user spoofing attack scenario by comparing a user's testing traces with a different user's profile. To study the statistical characteristics of our approach, in total, we create 12,144 (i.e., $(3 \times 22) \times ((2 \times 3) + 2) \times 23$) attacking instances and 552 (i.e., $3 \times ((2 \times 3) + 2) \times 23$) non-attack instances based on the number of the accelerometer traces we collected from all the volunteers.

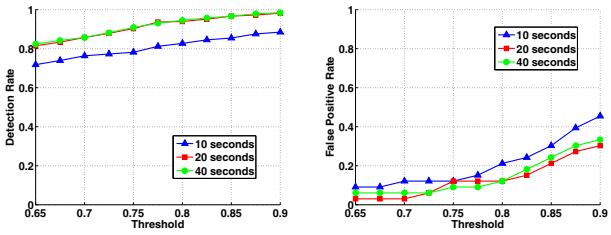
We use the detection rate and false positive rate to evaluate the effectiveness of our scheme. They are defined as:

- **Detection rate**: the percentage of attack instances that are correctly identified by our scheme;
- **False positive rate**: the percentage of non-attack instances that are mistakenly detected as attack instances.



(a) Number of cycles identified (b) Detection rate

Fig. 5. Comparison of step cycles identification by using different schemes with 15 traces from 5 users under different walking speeds.



(a) Detection rate (b) False positive rate

Fig. 6. Constant speed study: Performance of detecting user spoofing attack by using the same duration of run-time measurement trace and user profile trace.

B. Comparison of Step Cycle Identification

In the first set of experiments, we evaluate the effectiveness of our proposed step cycle identification scheme via comparing it with an existing method that identifies cycles by conducting the typical cycle identification repeatedly in a trace [11], [12], [15] (i.e., only based on local minimums searching). For comparison, we use 15 walking traces from 5 users with one minute length for each trace. Thus, there are 3 traces from each user under 3 different walking speeds. We compare the *step cycle detection rate*, which is the percentage of step cycles that are accurately identified, of our proposed method to the existing method.

Figure 5 (a) and (b) depict the cumulative number of identified step cycles and the corresponding detection rate with increasing number of walking traces for both our proposed method and the existing method. First, we observe that the number of the step cycles identified by our method stays very close to that of the actual number of step cycles present in each trace (reported by each user), whereas the gap between the curve of using the existing method and that from the actual step cycles is significantly larger than that of our proposed method. Further, we found that the detection rate of our scheme is significantly higher than that of using the existing method: our method can achieve a detection rate over 90% with different number of walking traces, while the detection rate ranges from 50% to 70% for the existing scheme. These observations indicate that our proposed step cycle identification scheme can derive step cycles much more accurately than the existing schemes [11], [12], [15]. This is because the existing schemes only rely on local minima searching, which is easily affected by the noise caused by irregular movement of the user's body and the detection errors propagate and affect the accuracy of subsequent cycle detections. Whereas our method exploits the high correlation inherent within a user's step cycles and is more robust to such noise.

C. Constant Walking Speed Study

We next evaluate the effectiveness of our gait based user spoofing attack detection method by using the run-time measurements with constant walking speeds.

1) Detection Latency Analysis: The detection latency analysis evaluates the detection performance when run-time measurements of different durations are used for attack detection. Specifically, we evaluate the detection performance when the run-time measurement trace length equals to "10 seconds", "20 seconds" and "40 seconds" respectively, under normal walking speed. The same corresponding length of the user profile trace is used. The time length of "10 seconds", "20 seconds", and "40 seconds" corresponds to about 9, 18 and 36 step cycles respectively under the normal speed.

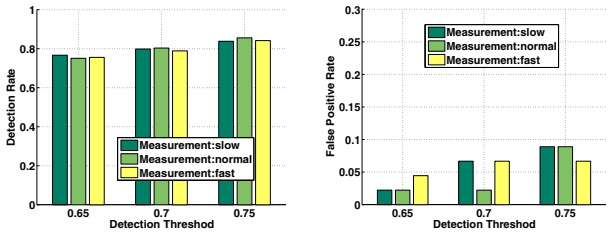
Figure 6 (a) and (b) present the detection rate and false positive rate under different detection thresholds. We first observe that the longer traces result in better detection performance. In particular, our scheme can achieve over 80% detection rate with less than 10% false positive rate when the trace length is longer than 20 seconds. This is because more step cycles can be identified in a longer trace which result in a more accurate capture of a user's unique gain pattern. The encouraging observation is that a trace length of 20 seconds is sufficient for our scheme to achieve a reasonable detection rate and a low false positive rate.

2) Robustness against Different Walking Speeds: We next study the robustness of our method under the scenarios where the run-time measurement traces are of different walking speeds from that used for a user profile. Figure 7 (a) and (b) present the detection rate and false positive rate under different detection thresholds when the run-time measurement traces are in slow, normal, and fast walking speeds, respectively. The duration of both user profile traces and run-time measurement traces is set as 20 seconds and a user profile is constructed from traces collected with a normal walking speed.

We observe that the detection rate increases as the detection threshold increases. This is because with higher detection threshold, it is easier for our scheme to detect traces which are from different users. Further, we find that the overall detection rate remains around 80% and the false positive rate is lower than 10%. Moreover, similar detection rate and false positive rate are achieved even if the traces of run-time measurements are collected using different walking speeds from that used to construct a user profile. This demonstrates that our scheme is robust against any potential attacks using different walking speeds.

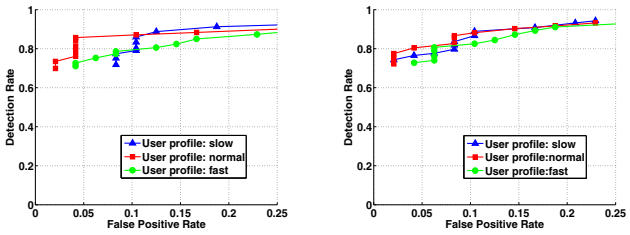
D. Varying Walking Speed Study

Finally, we evaluate the effectiveness of our method using the run-time measurement traces with varying speeds. We keep the user profile trace as 20 seconds while varying the duration of the run-time measurement traces. Figure 8 plots the Receiver Operating Curve (ROC) of our scheme when the



(a) Detection rate (b) False positive rate

Fig. 7. Constant speed study: Performance of detecting user spoofing attack by keeping the duration of both run-time measurement trace and user profile trace to 20 seconds with run-time measurement traces of different walking speeds.



(a) Run-time measurements: 20 seconds (b) Run-time measurements: 40 seconds

Fig. 8. Varying speed study: ROC curve of user spoofing attack detection with varying run-time measurement duration and user profile traces of different walking speeds and fixed 20 seconds duration.

detection threshold is changed from 0.65 to 0.9. The legend "*User profile: slow*", "*User profile: normal*" and "*User profile: fast*" denote the traces for constructing a user's profiles are chosen from constant speed traces with slow, normal and fast walking speeds, respectively. Figure 8 (a) and (b) show the run-time measurement traces are 20 and 40-second long, respectively.

Similarly, the overall performance of our method can achieve over 80% detection rate with less than 10% false positive rate. This shows that our method is robust to the dynamic changes of the users' walking speeds. Further, we observe that the performance using user profiles constructed from traces of different speeds become comparable when the false positive rate is around 10%, indicating our method is not sensitive to the walking speeds of training traces which are used to construct a user profile. In summary, our extensive experimental results demonstrate the effectiveness of leveraging gait patterns to perform user verification. Our correlation based step cycle identification results in better performance compared with the existing method. Our step cycle interpolation approach makes our scheme robust to varying user walking speeds, which are not investigated in other works. We anticipate utilizing some data mining techniques to further improve detection accuracy of our system, and we leave this to future work.

VI. Conclusion

In this paper, we address the problem of user spoofing attacks in emerging mobile healthcare systems. We propose a user verification scheme leveraging gait patterns derived from acceleration readings to mitigate against user spoofing attacks. Our framework employs readily available accelerometers embeded within smartphones instead of deploying additional hardware or requiring explicit user action for user verification. Our framework exploits the correlation relationship inherited from a user's walking traces and develops a step cycle template based technique that can identify the user's gait pattern more accurately than existing studies. Furthermore, our step cycle interpolation method can perform robust detection of the presence of user spoofing attacks under various walking speeds. Through experiments using 322 traces collected over a period of 6 months, we show that using smartphone based low grade accelerometers can achieve reasonable detection accuracy with only small number of run-time measurements for effective user verification in mobile healthcare systems. We will further analyze the robustness of our scheme when the adversary can study other users' walking styles in our future work.

Acknowledgement: This work is supported in part by National Science Foundation Grants CNS1016303, CNS1217387, CCF1018270, CNS1016296 and CNS1217379.

References

[1] M. Rabbi and et al, "Passive and in-situ assessment of mental and physical well-being using mobile sensors," in *Proceedings of UbiComp*, 2011.
[2] Y. Ren, J. Yang, M. C. Chuah, and Y. Chen, "Mobile phone enabled social community extraction for controlling of disease propagation in healthcare," in *Proceedings of IEEE MASS*, 2011.
[3] N. L. Clarke and S. Furnell, "Authentication of users on mobile telephones - a survey of attitudes and practices," *Computers and Security*, pp. 519–527, 2005.
[4] J. Mantyjarvi and et al, "Identifying users of portable devices from gait pattern with accelerometers," in *Proceedings of ICASSP*, 2005.
[5] D. Gafurov and E. Snekkenes, "Gait recognition using wearable motion recording sensors," *EURASIP J. Adv. Signal Process*, 2009.
[6] J. Yang, Y. Chen, W. Trappe, and J. Cheng, "Determining the number of attackers and localizing multiple adversaries in wireless spoofing attacks," in *Proceedings of IEEE INFOCOM*, 2009.
[7] A. Wool, "Lightweight key management for ieee 802.11 wireless lans with key refresh and host revocation," *Wireless Networks.*, vol. 11, no. 6, Nov. 2005.
[8] X. Chen and et al, "A secured mobile phone based on embedded fingerprint recognition systems," in *Proceedings of IEEE ISI*, 2005.
[9] T. Teixeira and et al, "PEM-ID: Identifying people by gait-matching using cameras and wearable accelerometers," in *Proceedings of ICDSC*, 2009.
[10] J. Jenkins and C. Ellis, "Using ground reaction forces from gait analysis: body mass as a weak biometric," in *Proceedings of Pervasive*, 2007.
[11] D. Gafurov, E. Snekkenes, and P. Bours, "Spoof attacks on gait authentication system," *IEEE Transactions on Information Forensics and Security*, 2007.
[12] M. Alzantot and M. Youssef, "UPTIME: Ubiquitous pedestrian tracking using mobile phones," in *Proceedings of WCNC*, 2012.
[13] J. J. Oresko and et al, "A wearable smartphone-based platform for real-time cardiovascular disease detection via electrocardiogram processing," *IEEE Transactions on Information Technology in Biomedicine*, 2010.
[14] "I Do Move Work out and Win," Sep. 2012. [Online]. Available: http://itunes.apple.com/us/app/idomove-work-out-and-win/id510602229?mt=8
[15] M. O. Derawi and et al, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *Proceedings of IIH-MSP*, 2010.
[16] C. BenAbdelkader and et al, "Stride and cadence as a biometric in automatic person identification and verification," in *Proceedings of FG*, 2002.
[17] G. Casella and R.L.Berger, *Statistical Inference*. Duxbury Press, 1990.
[18] R. V. Dukkipati, *Numerical Methods*. New Age International Pvt Ltd Publishers, 2010.
[19] C. Vaughan, B.Davis, and J.O'Cononor, *Dynamics of Human Gait*. Kiboho Publishers, 1999.
[20] M. Lin and et al, "A scalable approach for multidimensional wellbeing monitoring: Community and energy based adaptation of mobile sensing and feedback," in *Proceedings of Wireless Health*, 2012.