Continuous Verification Using Multimodal Biometrics*

Sheng Zhang, Rajkumar Janakiraman, Terence Sim, and Sandeep Kumar

School of Computing, National University of Singapore,
3 Science Drive 2, Singapore 117543
{zhangshe, janakira, tsim, skumar}@comp.nus.edu.sg

Abstract. In this paper we describe a system that continually verifies the presence/participation of a logged-in user. This is done by integrating multimodal passive biometrics in a Bayesian framework that combines both temporal and modality information holistically, rather than sequentially. This allows our system to output the probability that the user is still present even when there is no observation.

Our implementation of the continuous verification system is distributed and extensible, so it is easy to plug in additional asynchronous modalities, even when they are remotely generated. Based on real data resulting from our implementation, we find the results to be promising.

1 Introduction

For most computer systems, once the identity of the user has been verified at login, the system resources are typically made available to the user until the user exits the system. This may be appropriate for low-security environments but can lead to session "hijacking" (akin to hijacking [1]) in which an attacker targets a post-authenticated session. In high risk environments or where the cost of unauthorized use of a computer is high, continuous verification, if it can be realized efficiently is important to reduce this window of vulnerability. By this we mean that biometric verification is not merely used to authenticate a session on startup, but that it is used in a loop throughout the session to continuously authenticate the presence/particapation of the user. Examples where continuous verification is desirable include the usage of computers for airline cockpit controls, in defense establishments, and in other processing that affects the security and safety of human lives. In such situations, the desirable default action might be to render the computer system ineffective when the authorized user is not the one controlling it.

One way to realize (an approximation of) continuous verification is to use passive but accurate biometric verification. However, a single biometric may be inadequate for passive verification either because of noise in data samples or because of unavailability of a sample at a given time. For example, face verification cannot work when frontal face detection fails because the user presents

* This work was funded by the National University of Singapore, project no. R-252-146-112.

D. Zhang and A.K. Jain (Eds.): ICB 2006, LNCS 3832, pp. 562–570, 2005.

a non-frontal pose. To overcome this limitation, researchers have proposed the use of multiple biometrics, and have demonstrated increased accuracy of verification with a concomitant decrease in vulnerability to impersonation [4]. Use of multiple biometrics has led to the investigation of integrating different types of inputs (modalities) with different characteristics. Kittler et al. [2] experiment with six fusion methods for face and voice biometrics, using the sum, product, minimum, median, and maximum rules. In our work, we follow a similar approach: we combine face and fingerprint to do continuous verification.

For a continuous verification system, three criteria are important with regard to biometrics fusion:

1. The different reliability of the various modalities must be accounted for. That is, any fusion method must factor in the reliability of each modality.
2. Older observations must be discounted, to reflect the increasing uncertainty of the continued presence of the legitimate user.
3. Any fusion method should be able to handle lack of observations in one or more modalities, which arises from a normal usage pattern, i.e., when the user looks away from the camera.

Thus the usual fusion methods of sum, product etc. cannot be directly used because they do not satisfy the above criteria.

The key to continuous verification is the integration of biometric observations across both modality and time. Up to now, the task of integrating data across both modality and time has not been addressed satisfactorily. In this paper, we propose a Holistic Fusion method that combines face and fingerprint across modalities and time simultaneously and in a way that satisfies the above three criteria. This is realized by using the Hidden Markov Model (HMM). We experimentally compare our fusion method with a few alternatives – Time-first, Modality-first, and Naive Integration – and show that our method is superior.

2 Theory

The goal of verification is to determine whether the person with the claimed identity is who he claims to be. Two situations can occur: either the verifier *accepts* the claim as genuine, or the verifier *rejects* it (and decides that the user is an imposter).

In our case, the verification uses two types (modalities) of observations: fingerprint and face images. The challenge is to integrate these observations across modality and over time. To do this, we devised the integration scheme shown in Figure 1. Currently we implement a face verifier and a fingerprint verifier, other modalities are possible in the future. Each verifier computes a score from its input biometric data (fingerprint or face), which is then integrated

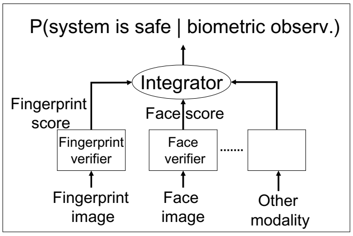

Fig. 1. Integration scheme

(fused) by the Integrator. The output from the Integrator is then used by the operating system kernel to delay or freeze user processes. For implementation details, please refer to [3].

2.1 Fingerprint Verifier

We acquire fingerprint images using the SecureGen$^{\text{TM}}$ mouse, which incorporates a fingerprint scanner ergonomically where the thumb would normally be placed. This makes the mouse a passive (non-intrusive) biometric sensor, ideally suited for continuous verification. The mouse comes with a SDK that matches fingerprints, i.e., given two images, it computes a similarity score between 0 (very dissimilar) and 199 (identical). Unfortunately, the matching algorithm is proprietary and is not disclosed by the vendor. Nevertheless, it is enough to get good results using the score generated by the proprietary algorithm.

First, we collect 1000 training fingerprint images from each of four users. For each user, we compute two probability density functions (pdf) - the intra-class and inter-class pdfs (represented by histograms). If we denote the similarity score by s, the intra-class set by Ω_U, and the inter-class set by Ω_I, then these pdfs are $P(s \mid \Omega_U)$ and $P(s \mid \Omega_I)$. The pdfs are similar to those in Figure 2 (which are for faces), but have smaller overlap, indicating that fingerprint verification is reliable (high verification accuracy).

Given a new fingerprint image and a claimed identity, the image is matched against the claimed identity's template (captured at registration time) to produce a score s. From this we compute $P(s \mid \Omega_U)$ and $P(s \mid \Omega_I)$. These values are then used by the Integrator to arrive at the overall decision. See Section 2.3 for more details.

2.2 Face Verifier

Our Face Verifier is also based on intra- and inter-class pdfs, except that the score s is now an image distance, rather than a measure of similarity. To train the Face Verifier, we first capture 500 images of each user under varying head poses, using a Canon VCC4 video camera and the Viola-Jones face detector [6]. The images are resized to 28×35 pixels. For each user, the training images are divided into the intra-class and inter-class sets. For each set, we calculate the pairwise image distance using the L_p norm (described below). This is similar to the ARENA method [5]. These distances are now treated as scores s, and the pdfs $P(s \mid \Omega_U)$ and $P(s \mid \Omega_I)$ estimated as before.

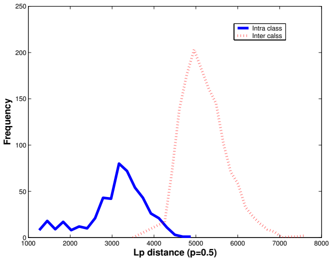

Fig. 2. Face intra-class and inter-class pdfs for a typical user

The L_p norm is defined as $L_p(\mathbf{a}) \equiv (\sum |a_i|^p)^{\frac{1}{p}}$, where the sum is taken over all pixels of image $\mathbf{a}$. Thus the distance between images $\mathbf{u}$ and $\mathbf{v}$ is $L_p(\mathbf{u} - \mathbf{v})$. As in ARENA, we found that $p = 0.5$ works better than $p = 2$ (Euclidean). Given a new face image and a claimed identity, we compute the smallest L_p distance between the image and the intra-class set of the claimed identity. This distance is then used as a score s to compute $P(s \mid \Omega_U)$ and $P(s \mid \Omega_I)$, which in turn are used by the Integrator.

2.3 Holistic Fusion

The heart of our technique is in the integration of biometric observations across modalities and over time. This is done using HMM, which is a sequence of states x_t that "emit" observations z_t, for time $t = 1, 2, \ldots$ Each state can assume one of two values: $x_t \in \{Safe, Attacked\}$. *Safe* means that the logged-in user is still present at the computer console, while *Attacked* means that an imposter has taken over control. It is also possible for the user to be absent from the console, but for a high security environment, this is considered to be the same as *Attacked*. Each observation z_t is either a face or fingerprint image, or equivalently, its corresponding score (See Sections 2.1, 2.2). Note that the states are hidden (unobservable), and the goal is to infer the state from the observations.

The result of the fusion is the calculation of P_{safe}, the probability that the system is still in the *Safe* state. This value can then be compared to a pre-defined threshold T_{safe} set by the security administrator, below which appropriate action may be taken. A key feature of our method is that we can compute P_{safe} at any point in time, whether or not there are biometric observations. In the absence of observations, we decay P_{safe}, reflecting the increasing uncertainty that the system is still *Safe*.

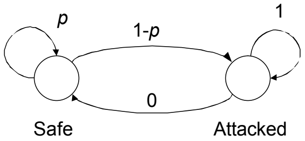

Fig. 3. State transition model

Let $\mathcal{Z}_t = \{z_1, \ldots, z_t\}$ denote the history of observations up to time t. From a Bayesian perspective, we want to determine the state x_t that maximizes the posterior probability $P(x_t \mid \mathcal{Z}_t)$. Our decision is the greater of $P(x_t = Safe \mid \mathcal{Z}_t)$ and $P(x_t = Attacked \mid \mathcal{Z}_t)$. Equivalently, we seek to determine if $P(x_t = Safe \mid \mathcal{Z}_t) > 0.5$, since the probabilities must sum to 1. We may rewrite:

$$P(x_t \mid \mathcal{Z}_t) \propto P(z_t \mid x_t, \mathcal{Z}_{t-1}) \cdot P(x_t \mid \mathcal{Z}_{t-1}) \tag{1}$$

$$P(x_t \mid \mathcal{Z}_{t-1}) = \sum_{x_{t-1}} P(x_t \mid x_{t-1}, \mathcal{Z}_{t-1}) \cdot P(x_{t-1} \mid \mathcal{Z}_{t-1}) \tag{2}$$

This is a recursive formulation that leads to efficient computation[1]. The base case is of course $P(x_0 = Safe) = 1$, because we know that the system is *Safe* immediately upon successful login. Observe that the state variable x_t has the effect of summarizing all previous observations. Because of our Markov assumptions, we note that $P(z_t \mid x_t, \mathcal{Z}_{t-1}) = P(z_t \mid x_t)$, and $P(x_t \mid x_{t-1}, \mathcal{Z}_{t-1}) = P(x_t \mid x_{t-1})$.

However, $P(z_t \mid x_t)$ is simply the intra-class pdf (when $x_t = Safe$) or the inter-class pdf (when $x_t = Attacked$). As for $P(x_t \mid x_{t-1})$, this is described by the state transition model shown in Figure 3. In the *Safe* state, the probability

[1] At time t, if there exists a biometric observation, we use Equation 1 to compute P_{safe}, otherwise Equation 2.

of staying put is p, while the probability of transitioning to *Attacked* is $(1-p)$. Once in the *Attacked* state, however, the system remains in that state and never transitions back to *Safe*.

The value of p is governed by domain knowledge - if there is no observation for a long period of time, we would like p to be small, indicating that we are less certain that the user is still safe (and thus more likely to have been attacked). To achieve this effect, we define $p = e^{k\Delta t}$, where Δt is the time interval between the current time and the last observation, and k is a free parameter that controls the *rate of decay*, which the security administrator can define. For instance, if the security administrator decides that p should drop to 0.5 in 30 seconds, then $k = -(\log 2)/30$.

In general, any decay function may be used to specify p, with a suitable rate of decay. We chose an exponential function for its simplicity: a value of $k = 0$ means that the user is never attacked ($p = 1$), while a very large value of k indicates that attacks are very likely.

3 Discussion

We compare our method with other alternatives: Temporal-first, Modality-first and Naive Integration.

3.1 Temporal-First and Modality-First Integration

Figure 4 shows how observations from different modalities present themselves over time. Observations from a single modality are shown horizontally, while observations across time are shown vertically. Note that at time t_3, only fingerprint is observed and also for ease of understanding, we show observations a and d as aligned vertically. In practice we allow a and d to occur within a small window of time apart.

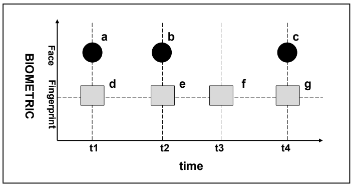

Fig. 4. Combining multiple biometric modalities

One common method of fusion is the following: let $P(x_t \mid \mathcal{Z}_t^{m_j})$ denote the posterior probability of being safe at time t for modality m_j. To combine across time, we compute the weighted sum:

$$P(x_t \mid \mathcal{Z}_t^{m_j}) = \frac{1}{N} \sum p(x_{t_i} \mid z_{t_i}^{m_j}) \cdot e^{k\Delta t_i} \tag{3}$$

Where Δt is the time difference between the current time and observation time, N is the number of observations. This decays older observations by the weight $e^{k\Delta t}$ such that it satisfies Criterion 2 for continuous verification.

To combine over modalities, we may again use a weighted sum:

$$P(x_{t_i} \mid z_{t_i}) = w^{m_1} \cdot P(x_t \mid z_{t_i}^{m_1}) + w^{m_2} \cdot P(x_t \mid z_{t_i}^{m_2}) \tag{4}$$

Note that here the two weights are w^{m_1} and w^{m_2}. They should be chosen to reflect the reliability of each modality, in order to satisfy Criterion 1. We will use the area under the ROC curve to represent the reliability.

Thus, Temporal-first implies the application of Equation 3 followed by Equation 4. Similarly, Modality-first changes this construction by applying Equation 4 first, then Equation 3. Note that if there is only a single modality (i.e., time t_3 in Figure 4), we just use the modality (no weight applied) as the combined result. Likewise if there is only one observation across time, then we just decay the observation by $e^{k\Delta t}$. In practice, for computational efficiency, we combine observations that occur within a recent history H from the current time, since observations that are too old have negligible weights.

3.2 Naive Integration

Since fingerprint is more reliable than face and also more reliable than the two combined (See Section 4.1), the idea of naive integration is to use the most reliable modality available at any time instant. More precisely,

1. At any time t, if a fingerprint observation exists, then $P(x_t \mid \mathcal{Z}_t) = P(x_t \mid z_t^{m_2})$ $(m_2 = fingerprint)$ whether or not face observation exists.
2. Otherwise if there exists only face observation, then $P(x_t \mid \mathcal{Z}_t) = P(x_t \mid z_t^{m_1})$ $(m_1 = face)$, since now face is the most reliable biometric that is available.
3. Else if no biometric observation is available, then we just decay the probability $P(x_t \mid \mathcal{Z}_t) = P(x_{t-1} \mid z_{t-1}) \cdot e^{k\Delta t}$. Where $P(x_{t-1} \mid z_{t-1})$ is calculated from Step (1) or (2), depending on the last biometric observation (fingerprint or face). Here Δt is the time interval between the current time and the latest observation time.

It is clear that Naive Integration satisfies the three criteria in Section 1.

4 Experiments

All the experiments were conducted on real users using an Intel Pentium 2.4 Ghz Workstation with 512MB RAM. The captured images are 384 × 288, 24-bit deep taken using a Euresys Picolo capture card with a Canon VCC4 camera.

Ideally all the biometric data are acquired at fixed times. But in reality the observations greatly depends on how the user presents himself to the Biometric system. Following are the possible cases where there could be no observation. (1) User is not using the mouse or not placing his thumb on the fingerprint scanner. (2) User is not presenting a frontal face to the camera.

4.1 ROC Curve Analysis

For assessing the Receiver Operator Characteristic (ROC) our system, we run 6 sets of experiments for each user under the different combinations of legitimate user versus imposter for face and fingerprint modalities.

The area under ROC curve is the reliability measure. From the fused probabilities of the above experiments, we compute ROC's for face verifier, fingerprint verifier and both combined. The ROC areas for fingerprint-only, combined-modality, and face-only verifiers are 0.9995, 0.989, and 0.970, respectively. Thus verification using fingerprint alone is the best, followed by combining the two modalities. Face verification alone is the least reliable.

However, for continuous verification, combining multimodal biometrics is preferred over using just a single modality. The lack of observations from a single modality can be compensated by using a second modality. Also it is more difficult for an imposter to impersonate multiple biometrics.

4.2 Comparing the Fusion Methods

We run four experiments to evaluate how the system behaves when one or both of the biometrics are impersonated. In these we take turns to impersonate each modality one at a time. Because each user presents his biometric in a different way, we cannot average the curves from different users. Figure 5(a) 5(b) 5(c) show five plots each in the following order: individual probabilities, Holistic Fusion, Naive Integration, Modality-first, Temporal-first Integration. In these experiments, $\Delta t = 1.5s$ is used for modality integration, $H = 30s$ for temporal integration and $k = -log(2)/30$ for the decay function. There can be no observation at some time periods. In these situations in order to maintain the system integrity we choose to lock the system. The user has to re-login to regain access. These four setups can be classified into three cases.

Legitimate user using the system. Figure 5(a) shows the biometric observation for 15 minutes. The individual probabilities P_{safe} (5(a)-1) are not consistently high, it occurs in a sporadic manner. This means that any value for the threshold T_{safe} will result in significant False Accept (FAR) and False Reject (FRR) rates. In continuous verification, a False Accept is a security breach, while a False Reject inconveniences the legitimate user, because he must re-authenticate himself. Ideally P_{safe} should not fluctuate, but be equal to 1 as long as observations are available. Of the four fusion methods, Holistic Fusion comes closest to this ideal (5(a)-2). It computes a P_{safe} value close to 1, except for the periods when there are no observations from both modalities (around 300s and 600s). At such times P_{safe} decreases gradually according to the decay function. By comparison, the P_{safe} computed by Naive Integration (5(a)-3) fluctuates wildly, because only a single modality is used any at time. Again, this means no T_{safe} value will make both FRR and FAR small. As for Modality-first (5(a)-4) and Temporal-first (5(a)-5) Integration, the plots are similar. The P_{safe} values are not close to 1. Moreover in the absence of observations P_{safe} drops abruptly to zero resulting in sudden lock outs. From these plots, it is clear that Holistic Fusion is superior to the other fusion methods.

Imposter taking over the system. Figure 5(b) shows the observations when an imposter takes over the system at some time instant (around 38s). The probabilities of individual biometrics (5(b)-1) as well as P_{safe} for all integration methods drop to near zero after the attack. The goal here is to detect the attack as soon as possible so that damage to the system is minimized. Both Holistic Fusion (5(b)-2) and Naive Integration (5(b)-3) detect this situation sooner than the other two methods. However, P_{safe} for Naive Integration does not remain consistently low; it fluctuates widely. This implies that FAR >0 for most values of T_{safe}. For Modality-first (5(b)-4) and Temporal-first (5(b)-5) Integration, the system takes longer to detect the imposter (when $T_{safe} = 0.5$). Choosing a larger value for T_{safe} can reduce the time to detection, but at the expense of a higher FRR. The best method is Holistic Fusion, which detects the imposter quickly (within 5s in our experiments), and whose P_{safe} remains low after the attack.

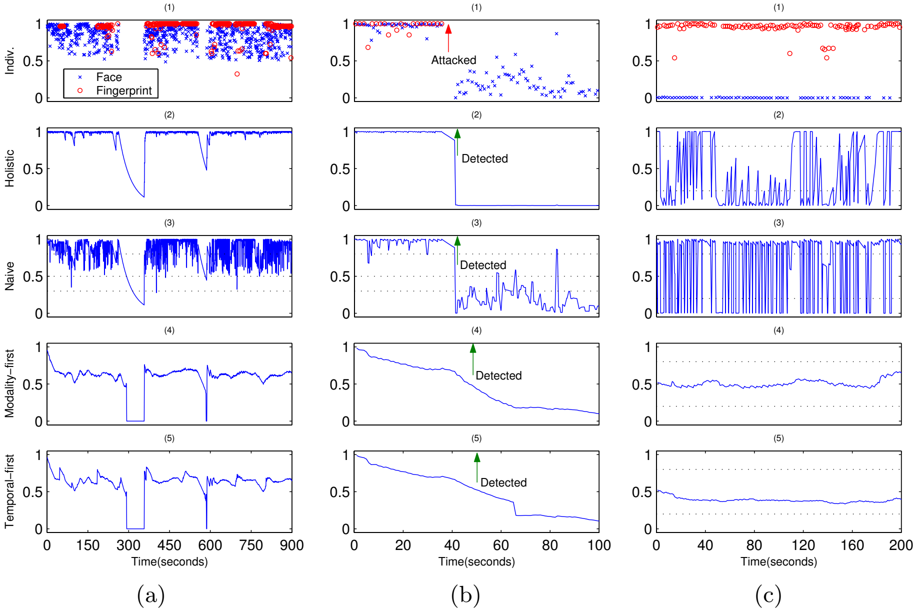

Fig. 5. (a) Legitimate user using the system for 15 minutes. (b) Imposter taking over the system. (c) Partial impersonation: Genuine fingerprint + Fake face. Experiments conducted with Fake fingerprint + Genuine face produced similar results as (c).

Imposter successful in faking one of the biometric (Partial impersonation). Figure 5(c)-1 depicts a situation where the imposter has successfully faked the fingerprint but not face. The individual probabilities contradict each other, and results in wildly fluctuating plots in both Holistic Fusion (5(c)-2) and Naive Integration (5(c)-3). This gives us a way to detect partial impersonation: We may just take two thresholds, one high and one low (say: 0.8 and 0.2) and simply count the number of times within a fixed time interval that P_{safe} jumps between these thresholds. However, comparing Figures 5(c)-3 and 5(a)-3, we see that Naive Integration cannot distinguish between partial impersonation and the legitimate user. Fluctuating P_{safe} values seem to be an inherent property of

Naive Integration. The plots for Modality-first (5(c)-4) and Temporal-first (5(c)-5) Integration are relatively flat, and are in fact similar to those in Figure 5(a) (except when there are completely no biometric observations). Again, this means these two methods cannot distinguish between partial impersonation from legitimate usage. Only Holistic Fusion provides a way to detect partial impersonation that is different from detecting the real user.

What happens if an imposter is careful not to present any observation (neither face nor fingerprint)? In this case, P_{safe} decreases to zero due to the decay function. This is also the situation if the legitimate user has left the console without logging off. In either case, system integrity is ensured.

5 Conclusion

In summary, our work has the following key features:

1. We propose a Holistic Fusion approach that satisfies all the three criteria for continuous verification.
2. We experimentally show that our Holistic Fusion is superior to other alternative methods: Temporal-first, Modality-first and Naive Integration. It is the only method that (a) achieves a low FAR and FRR, (b) detects an attack quickly after it occurs, and (c) is able to detect partial impersonation.
3. In our system, there is only one free parameter k that governs the decay rate. This is intuitively specified by the security administrator based on security requirements.

In the near future, we plan to incorporate keyboard dynamics as another biometric modality. We also plan to make face verification more robust by using incremental training.

References

1. Laurent Joncheray. A Simple Active Attack Against TCP. *Proceedings of the 5th USENIX Security Symposium*, pages 7–19, 1995.
2. J. Kittler, M. Hatef, R. P. W. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, Mar. 1998.
3. Sandeep Kumar, T. Sim, Rajkumar Janakiraman, and S. Zhang. Using continous biometrics verification to protect interactive login sessions. *To appear in the 21st Annual Computer Security Applications Conference*, 2005.
4. A. Ross and A. K. Jain. Information fusion in biometrics. *Pattern Recognition Letters*, 24(13):2115–2125, 2003.
5. T. Sim, R. Sukthankar, M. Mullin, and S. Baluja. Memory-based Face Recognition for Visitor Identification. In *IEEE International Conference on Automatic Face and Gesture Recognition*, 2000.
6. Paul Viola and Michael Jones. Robust real-time object detection. *International Journal of Computer Vision*, 2002.