VideoTicket: Detecting Identity Fraud Attempts via Audiovisual Certificates and Signatures

D. Nali
School of Computer Science
Carleton University
Ottawa, Canada
deholo@ccsl.carleton.ca

P.C. van Oorschot
School of Computer Science
Carleton University
Ottawa, Canada
paulv@scs.carleton.ca

A. Adler
Systems and Computer Engineering
Carleton University
Ottawa, Canada
adler@sce.carleton.ca

ABSTRACT

Identity fraud (IDF) may be defined informally as exploitation of credential information using some form of impersonation or misrepresentation of identity, in the context of transactions. Thus, IDF may be viewed as a combination of two old problems: user authentication and transaction authorization. We propose an innovative approach to detect IDF attempts, by combining *av-certificates* (digitally-signed audiovisual recordings in which users identify themselves) with *av-signatures* (audiovisual recordings showing users' explicit consent for unique transaction details). Av- certificates may be used in on-site transactions, to confirm user identity. In the case of remote (e.g. web-based) transactions, both av-certificates and av-signatures may be used to authenticate users and verify their consent for transaction details. Conventional impersonation attacks, whereby credentials (e.g. passwords, biometrics, or signing keys) are used without the consent of their legitimate users, fail against `VideoTicket`. The proposed solution assumes that identity thieves have access to such credentials.

1. INTRODUCTION

Identity fraud (IDF) may be defined informally as exploitation of credential information using some form of impersonation or misrepresentation of identity. Javelin Strategy & Research reported that, in 2005, 8.9M American adults became IDF victims [35]. On average, each of these victims was defrauded $6,383, and spent 40 hours to resolve their IDF problem. 47% of IDF cases were detected by victims themselves. We seek to present a method that helps detect IDF attempts. Throughout the paper, we use the term *transaction* to denote any interaction involving two or more parties, and resulting in the issuing of credential tokens (e.g. credit or health cards), access to services (e.g. health care) or goods (e.g. software programs, groceries, etc), and/or financial transfers. We use the expression *remote transaction* to refer to transactions involving at least one remote (e.g. web-based) party, at transaction time. Transactions that are not remote are said to be *on-site*.

Few generic IDF detection systems (i.e. IDF detection systems that can simultaneously be used for remote and on-site transactions, regardless of applications[1]) have been proposed in the academic literature. Application-specific IDF detection methods (such as phishing and key-logging countermeasures [3, 7, 16, 21, 39]) are known, but we seek to design a generic method, which we expect to be more convenient and less expensive for end-users, when considered across applications. We also seek to design an IDF detection method that combines user authentication (since IDF deals with the fraudulent use of identity) and transaction authorization (since IDF consists in exploiting credential information in the context of transactions). Furthermore, we aim at designing a method that does not rely on credentials which can be used fraudulently, i.e. without their legitimate users' explicit consent for specific transaction details. This is a limitation of digitized handwritten signatures, passwords, secret keys, digital signatures, message authentication codes, keys derived from fingerprints, and statements certifying device locations. Contrary to most proposals, we assume identity thieves (already or will) have access to such user secrets, and propose a method to detect IDF despite this assumption [18].

Overview of Proposed Scheme. We propose an IDF detection scheme using audiovisual recordings (*av-recordings*) to simultaneously authenticate users and authorize transactions. At system setup, each user is issued an audiovisual certificate (*av-certificate*), i.e. a data structure composed of: (1) a list of privileges granted to the av-certificate's legitimate holder; (2) an av-recording in which this user shows her face, and identifies herself (e.g. through spoken words); and (3) a digital signature over (1) and (2) computed by a trusted authorized party. At each transaction, the user's av-certificate is combined both with transaction details, and a freshly generated av-signature (i.e. av-recording in which the user conveys consent for these transaction details); we call the combined data structure an audiovisual ticket (*av-ticket*). This av-ticket is sent (e.g. via the web) to a relying party (e.g. a credit card issuing company). The relying party (or a delegate thereof) examines the av-ticket by verifying the following four criteria: (a) the associated av-signature

[1]e.g. credit card payments, border control, health care provision control, etc.

includes all transaction details included in the av-ticket; (b) the associated av-certificate indicates that its legitimate holder has all privileges required for the given transaction; (c) the digital signature included in the av-certificate is that of a trusted and authorized party; and (d) the person shown on the av-certificate's av-recording appears (with a reasonable level of certainty) to be the same as that shown on the av-signature. If these four criteria are met, the transaction associated with the examined av-ticket is authorized. In the case of on-site transactions, only av-certificates are needed (to verify users' identity); in the case of remote transactions, both av-certificates and av-signatures are used (in the form of av-tickets), to authenticate users and confirm their consent for transaction details. In either case, the proposed scheme may be used for a chosen class of transactions, e.g. card issuing and high-value transactions.

The proposed scheme (called `VideoTicket`) combines user authentication and transaction authorization, enables on-site (hence decentralized) verification, can be used for both remote and on-site transactions, and is suitable for multiple classes of applications (see Section 2.4). `VideoTicket` has lighter security requirements than user-based digital signatures (users need no signing keys; hence user-based signing keys need not be protected). To avoid replay attacks or use of av-signatures for unintended purposes, `VideoTicket` also uses unique transaction details (e.g. by including a transaction's date/time or unique transaction identifier generated by a transaction authorizing party). It can therefore be viewed as the combination of (facial, voice, and/or gesture) biometrics and a challenge-response protocol between users and transaction authorizing parties. `VideoTicket` is not resilient to certain classes of av-recording forgery (see Section 3), but makes such both forgery user- and transaction-specific, and thereby less scalable for IDF in comparison to classes of IDF committed with reusable credential (e.g. credit card) information (obtained, e.g., via mass database compromise). `VideoTicket` might require human-based verification of av-recordings. In the case of on-site transactions, this human-based verification consists in verifying users' av-certificates, in the same way clients' handwritten signatures are theoretically verified by cashiers using the back-side of credit cards.

In summary, we propose a generic method to detect IDF attempts by examining and comparing audiovisual recordings of users. `VideoTicket` captures users' biometric identity and consent for transaction details; hence, impersonation attacks, whereby credentials are used (potentially multiple times) without their legitimate users' consent, do not work. `VideoTicket` assumes that identity thieves have access to such credentials. We report on our early-prototype partial implementation of `VideoTicket`. Our work raises interesting questions for biometric research, e.g. the possibility of fully-automated multi-modal biometric authentication schemes combining gesture analysis with face and voice recognition. We wish to stimulate research on the automatability and commercial viability of schemes like `VideoTicket` with present or emerging technologies.

Outline. Section 2 describes the proposed scheme and applications thereof. Section 3 discusses various aspects of `VideoTicket`, including detection effectiveness, financial and time cost, on-site verifiability, scalability, privacy implications, manageability, security requirements, convenience of use, and verification outsourcing capability. Section 4 reports on and presents lessons learned from a partial prototype of `VideoTicket`. Section 5 reviews related work. Section 6 concludes.

2. VIDEOTICKET PROTOCOL

This section describes `VideoTicket`, a generic method to detect IDF attempts by comparing av-recordings. Section 2.1 lists parties involved in the scheme. Sections 2.2 and 2.3 present the two main protocols, namely *Setup* and *Transaction*. Section 2.4 describes practical potential applications using variants of `VideoTicket` (including, notably, a protocol employing av-certificates and audiovisual calls to confirm user identity and consent for transaction details).

2.1 Parties Involved

`VideoTicket` involves four main parties denoted U, R, V, and B (see Fig. 1). U is a legitimate system user who carries a general-purpose storage device d_U (e.g. a flash drive, magnetic-stripe or smart card, or cell phone) used to store credential information allowing credential relying parties to determine whether U has a claimed identity (ID) and set of privileges. U must be able to: (1) obtain and understand transaction details presented to her;[2] (2) show her face and express her will before a camera and microphone (e.g. using words or hand signs); and (3) execute other computer-oriented transaction-related tasks such as card swiping (in the case of on-site transactions), web-browsing, keyboard typing, and mouse pointing and clicking (in the case of remote transactions). R is a credential relying party (e.g. a building entrance control office, credit or student card issuing office, web-based merchant or service provider, or on-site point of sale). V is a party on which R relies[3] to verify users' claimed identities and privileges, and B is a party that issues av-certificates to users so that they prove their claims of ID and possession of privileges. For example, $B = R = V$ could be a government agency (or credit card company) that reviews online applications to issue health cards (or credit cards); each online card application could include an av-ticket. (Section 2.4 outlines more detailed application scenarios.) U, R, V, and B may have various trust relationships. In Section 3.1.2, we identify several such relationships, and discuss their impact on the security guarantees provided by `VideoTicket`.

2.2 Setup Protocol

Public Key Setup. B generates for itself a signature-related public-private key pair (e_B, d_B), and V obtains an authentic copy of e_B. If $R \neq V$, R and V may also obtain authentic copies of each other's signature and encryption-related public keys to realize authenticated, confidentiality-protected, and integrity-protected communication channels between
them.

[2]This may involve using PC keyboard, mouse, and monitor.
[3]In some instantiations, V and R may be collocated or the same entity. A single V may be relied upon by multiple credential relying parties.

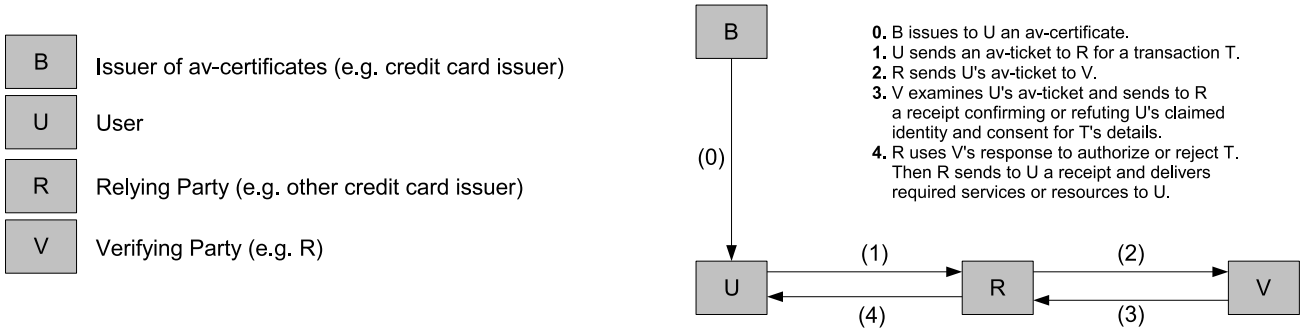

Figure 1: `VideoTicket` Protocol Overview

Replay-Protection Setup. V creates a table E_V (used in Step 4 of Section 2.3) to detect replay attacks. E_V contains identity and transaction authorization information processed by V within the last Δ_V time units (e.g. 2 hours).

Maximal Transaction Duration Setup. R sets an upper bound to the processing time of each user transaction, by initializing Δ_{trans} (e.g. setting $\Delta_{trans} = 5$ minutes).

Credential Information Setup. To obtain an av-certificate from B, U interacts with B as follows.

1. U goes to B in person, and requests from B all pieces of ID and privilege-related credential information she is entitled to receive from B.
2. B ensures that U is who she claims to be (e.g. via a predetermined out-of-band procedure involving presentation of identity-related cards issued by trusted parties, and confirmation of information found on these cards through phone call to their issuers). Then, the following takes place.
   (a) B assigns to U a permanent identifier ID_U, and, if U has privileges π_1 through π_n (where n is a positive integer such as 10), B forms the sequence $(ID_{\pi_1}, \cdots, ID_{\pi_n})$ of privilege identifiers.[4] B also associates with this n-tuple the pair (ID_B, ℓ), where: ID_B is a permanent identifier of B assigned by a trusted naming authority and used to obtain (e.g. through a web query) an authentic copy of e_B; and ℓ is a string encoding π_i's validity time interval (for $i = 1, \cdots, n$), and a description of both h and a public-key signature scheme with associated key size (e.g. 2048-bit RSA-PSS).
   (b) B records a short (e.g. 15-second) audiovisual sequence r_U in which U shows her face (as for passport photos, but under multiple viewing angles), and identifies herself. To identify herself, U may speak a few sentences,[5] use hand signs, and/or demonstrate a physical token showing identification information. In the latter case, the token must be sufficiently large to be examined by V without zoom. U may identify herself using a veronym (i.e. identifier revealing U's identity), or pseudonym assigned by B. The image and sound quality of r_U should be sufficiently high for software-assisted human verifiers employed by V to determine whether r_U is a montage of (shorter) audiovisual clips, or the person in r_U is not the same person as that appearing in another specified audiovisual clip (see Step 4 of Section 2.3). r_U must also be recorded with adequate equipment (e.g. a noise-canceling microphone, and light-and-color-adjusting video camera), in a partially-controlled environment (e.g. suitably-lit private office, or semi-closed booth located in a public area).
   (c) B uses d_B to compute $b_U = ([h(r_U), ID_U, \ell]_{d_B}, [h(r_U), ID_U, \ell, ID_{\pi_1}]_{d_B}, \cdots, [h(r_U), ID_U, \ell, ID_{\pi_n}]_{d_B})$.[6]
   (d) B forms U's av-certificate $c_U = (ID_U, \ell, r_U, ID_{\pi_1}, \cdots, ID_{\pi_n}, b_U, ID_B)$, and has c_U be stored on d_U (e.g. by obtaining d_U from U, and storing c_U thereon; or by sending c_U to d_U via Bluetooth or SMS). c_U is thereby stored on d_U, and U obtains d_U.

2.3 Transaction Protocol

Suppose now that a user U wants to interact, through a transaction T, with a party R, to access a set S of services or resources. Suppose also that, to do so, R requires U to have privileges π_1 through π_n, and assume that R, U, and V refer to S using the string identifier ID_S. For R to determine whether U's claims of ID and possession of privileges are valid, U, R, and V proceed as follows.

1. At time t_{start}, R generates and gives to U a string τ encoding unique and partially unpredictable transaction details associated with T. For instance, let τ consist of a dollar value, R's identifier and geographic location, t_{start}, and a transaction identifier uniformly chosen at random by R from a sufficiently large set of easy-to-pronounce-or-read words (e.g. ZIP or postal codes), or easy-to-reproduce gestures.
2. U gives the av-certificate c_U to R, via a communication channel providing stream-integrity and confidentiality protection, and enabling authentication of R by U.[7]

[4] The concatenation of bit strings x and y is "x, y" or (x, y).
[5] For instance, U could say: "This is Joe Morning, customer at BestBank, at 12:30, on July 13th, 2006, in Boston."
[6] $[x]_{d_B}$ is the digital signature on string x using key d_B.
[7] Such a channel may be instantiated using SSL with server

3. Let s_T be an av-signature of T by U, i.e. an av-recording in which U shows her face and shows consent for the information encoded in τ. Consent for τ's details can be demonstrated through spoken words, hand signs, or showing of information printed on, or electronically displayed by a physical token (e.g. a small movable monitor attached to a kiosk supervised by R). s_T is assumed to have sufficiently high image and sound quality to enable software-assisted human verifiers employed by V to determine whether s_T is a montage of shorter av-recordings, or the person appearing in s_T is not the same as that appearing in r_U. s_T must also be recorded with adequate equipment (e.g. a noise-canceling microphone, and light-and-color-adjusting video camera), in a partially- controlled environment (e.g. suitably-lit private office, or semi-closed booth located in a public area). To improve the verifiability of s_T, U may be asked to position her head in front of a video camera in such a way that her face appears in a box displayed on a screen; the dimension of this box may be chosen so that relevant features of U's face can be discerned by s_T's verifier. Either R records s_T, or s_T is recorded with a microphone and camera-enabled device (e.g. a laptop PC or cell phone) used by U. In the latter case, U's device sends s_T to R through a communication channel providing stream-integrity and confidentiality protection.
4. R forms (p_T, ID_S), where $p_T = (\tau, s_T, c_U)$ is an *av-ticket*. If $V \neq R$, R then sends (p_T, ID_S) to V via a communication channel providing mutual authentication, and stream-integrity and confidentiality protection.
5. If $V \neq R$, V checks whether E_V has an entry containing τ, and if so, V notifies R that p_T has already been processed by V. If $V = R$, V stores p_T in E_V,[8] and uses a software-assisted person, to check whether:
   - C1. s_T does not appear to be a montage of av-recordings, which can be checked by seeking abrupt changes in objects' (e.g. lips or hands) movements, light contrast, image color, sound pitch, or sound volume;
   - C2. the person appearing in s_T is the same as that appearing in r_U;
   - C3. consent for all elements of τ that make T unique and unpredictable (with respect to any other transaction) is shown in s_T; this may involve examination of speech, hand movements, and/or information appearing on a physical token shown by the person appearing on s_T;
   - C4. information encoded in (τ, b_U) and independently verifiable by V (e.g. current time,[9] R's identifier, and inclusion of current time in b_U's validity period) is accurate;
   - C5. the $n+1$ components of b_U are valid signatures by B on $(h(r_U), ID_U, \ell)$, and $(h(r_U), ID_U, \ell, ID_{\pi_1})$ through $(h(r_U), ID_U, \ell, ID_{\pi_n})$ respectively;[10] and
   - C6. the tuple $(ID_{\pi_1}, \cdots, ID_{\pi_n})$ includes the identifiers of all privileges required to access S.

   If conditions $C1$ through $C6$ are met, V sets $a_T = 1$ to indicate that the person M who presented c_U has identity ID_U and possesses all privileges required to access S. Otherwise, V sets $a_T = 0$ to indicate that some of conditions $C1$ through $C6$ are not met, or M is an impersonator of U (i.e. the user of identity ID_U). V also sets e_T to be a short constant string disregarded if $a_T = 1$, and specifying which conditions are not met and what could be done by U, if $a_T = 0$. If $V \neq R$, V then sends $(a_T, e_T, h(p_T))$ to R via a communication channel providing mutual authentication, stream-integrity and confidentiality protection.
6. If $V \neq R$, R uses $h(p_T)$ (given $(a_T, e_T, h(p_T))$) to associate a_T with p_T. Let t_{end} be the time at which R receives a_T. If $t_{end} - t_{start} > \Delta_{trans}$ or $a_T = 0$, R rejects T, and sends to U a transaction receipt z_T including s_T and a short constant string indicating the reason why T was rejected (e.g. a description of e_T concatenated with a note indicating that T's processing duration was too long). If, on the other hand, $a_T = 1$, then R authorizes T and sends to U a transaction receipt z_T including s_T.[11] If T is an on-site transaction, R may also print and give to U a partial transaction receipt (e.g. a portion of z_T excluding s_T).

2.4 Practical Application Scenarios

`VideoTicket` may be used for various practical applications including driver's license, health-care or credit card issuing, financial transaction approval, and customer authentication for remote assistance.

In the case of driver's license and health card issuing, `VideoTicket` could be used with $B = R = V$ being a government agency that issues health cards (HCs) or drivers' licences (DLs). Legitimate HC or DL holders may be required to obtain a new card every t (e.g. $t = 5$) years, either at designated offices, or via the web. In the former case, av-certificates may be used to confirm applicants' identity; in the other case, av-certificates may be combined with av-signatures, to authenticate applicants and confirm their will to be issued a new card.

Another way to use `VideoTicket` is to let $R = V$ be a company that issues credit cards after having verified av-tickets sent by applicants via the web. The av-tickets may need to be issued by select (trusted) banks, credit card companies, or government agencies.

authentication (in the case of online transactions), or the physical insertion of a token (storing c_U) in a trusted input device controlled by R (in the case of on-site transactions).

[8] Recall that E_V stores values temporarily. V removes all entries of E_V that have been stored for Δ_V time units or more.

[9] Some accuracy level (e.g. a 5-minute window) of time synchrony between U, R, and V is hereby assumed.

[10] While verification of $n + 1$ signatures is more computationally intensive than verification of a single signature on $(h(r_U), ID_U, \ell)$ and a combination of $(h(r_U), ID_U, \ell, ID_{\pi_i})$'s, the former saves space in b_U by removing the need to include, in b_U, a large number of signatures by B on $(h(r_U), ID_U, \ell)$ and all possible combinations of $(h(r_U), ID_U, \ell, ID_{\pi_i})$'s.

[11] z_T may be sent to an email or other address specified by U in Step 1 of the *Transaction* protocol. The address may be used only once to obtain z_T, or multiple times for interactions with R or certain classes of credential relying parties.

VideoTicket may also be used as follows: R is a web merchant, and $B = V$ is a credit card company. To detect credit-card-based forms of IDF, B could require that select transactions (e.g. high-value, international, or postdated fund transfers) need the presentation of av-tickets.

Note that av-tickets may be presented to V as explained in Section 2.3, or in a variant protocol whereby av-signatures are requested by V directly from U, under chosen circumstances (e.g. for high-value transactions). For example, U may carry a camera-phone enabling audiovisual calls. Suppose U wants to make an expensive purchase from R over the web. U could fill an associated web form, and send her av-certificate with this form to R. R would delegate the transaction request to V (as explained in Section 2.3). Then, in order to confirm U's identity and consent for the transaction, V could call U (using information extracted from U's av-certificate), and engage in an audiovisual call with the person answering the call. In such a case, the audiovisual call would play the role of the av-signature described in Section 2.3.

Another way to use VideoTicket is to let $B = R = V$ be a company (e.g. bank, Internet service provider, or large corporation) that wants to offer remote assistance (e.g. financial advice or computer support) to its customers/employees. To do so, each customer/employee U of B obtains an av-certificate from B, via a registration procedure. To authenticate its customers, $R = B$ then deploys a secure web-based audiovisual chat-like application that provides a confidential communication channel between U and R. When U seeks assistance from R, U interacts with R via the aforementioned web application, and $V = R$ authenticates U by obtaining (from U or a trusted database) U's av-certificate. V may ask U multiple authenticating questions (as in the case of commonplace telephone assistance), but these questions may be partially replaced by audiovisual evidence which R obtains by downloading U's av-certificate. In the latter case, the time required by R to authenticate U may be shortened, thereby improving both U's experience and R's (time and financial) efficiency at providing remote assistance.

We emphasize that, for each application, VideoTicket may be used for a chosen class of transactions, e.g. card issuing and high-value transactions.

3. DISCUSSION OF VIDEOTICKET

In this section, we discuss the security and several other aspects of VideoTicket, including scalability, privacy implications, convenience of use, and financial and time cost. We also discuss the automatability of the proposed scheme.

3.1 Security Discussion

3.1.1 Threat Model

We consider a number of goals and techniques ID fraudsters may respectively have and employ to attack practical instantiations of VideoTicket. While not exhaustive, these goals and techniques are meant to abstract a number of realistic practical threats. We do not aim to mathematically prove the security of VideoTicket; it partially relies on the reliability of human analysis and comparison of audiovisual recordings. We use the notation of Section 2 to enable more precise discussion.

Adversarial Goals. In practical instantiations of VideoTicket, IDF may take multiple forms depending on system applications (e.g. on-site debit card payment, online issuing of credit cards, on-site access to medical services, or on-site border control). Here, we abstract four goals ID fraudsters may have when they attack our scheme. ID fraudsters may seek to: (G1) *gain money* (or credits corresponding to money); (G2) *gain access to digital or physical services or resources* without paying; (G3) *preserve their anonymity* when accessing physical or digital services or resources; (G4) *frame legitimate users* (e.g. by discrediting or blackmailing them).

Adversarial Techniques. To impersonate a legitimate user U of VideoTicket, an ID fraudster A may use several techniques, including the following (and their combinations).

T1. **Forgery of Identity Claim (e.g. Audiovisual Recording)**. A may attempt to forge an audiovisual sequence $\hat{v}_T$ in which a person appearing to be U shows her face and speaks the information encoded in forged transaction details $\hat{\tau}$ which U does not show explicit consent for. If A also obtains the credential information c_U of U, then A may be able to impersonate U, by sending $\hat{p}_T = (\hat{\tau}, \hat{v}_T, c_U, ID_S)$ to R.

T2. **Forgery of Identity Proof's (e.g. av-Certificate's) Digital Signature**. A may attempt to forge signatures issued by B, e.g. by obtaining a copy of d_B, or using a weakness in the associated signature scheme. This can be used in either of the following scenarios:

- (a) If A is able to forge a signature of B on $(h(r_A), ID_U, \ell)$, and $(h(r_A), ID_U, \ell, ID_{\pi_1})$ through $(h(r_A), ID_U, \ell, ID_{\pi_n})$, where r_A is an av-recording analogous to r_U but showing A, then A can access S without having the required legitimate ID or privileges. This is done by sending $(\hat{p}_T, ID_S)$ to R, where: $\hat{p}_T = (\hat{\tau}, \hat{s_T}, s_A)$; $\hat{\tau}$ are transaction details potentially unknown to U, but agreed upon by A and R; $\hat{s_T}$ is an av-recording in which A shows her face and explicit consent for $\hat{\tau}$; $s_A = (ID_U, \ell, r_A, ID_{\pi_1}, \cdots, ID_{\pi_n}, b_A, ID_B)$; and $b_A = ([h(r_A), ID_U, \ell]_{d_B}, [h(r_A), ID_U, \ell, ID_{\pi_1}]_{d_B}, \cdots, [h(r_A), ID_U, \ell, ID_{\pi_n}]_{d_B})$.
- (b) If $A = U$ and U does not have a privilege π_{n+1} (identified by $ID_{\pi_{n+1}}$) needed to access a service or resource offered by R, then A may proceed as follows, if she is able to forge a signature of B on $(h(r_U), ID_U, \ell, ID_{\pi_{n+1}})$: A uses $[h(r_U), ID_U, \ell, ID_{\pi_{n+1}}]_{d_B}$ as a forged proof that she has privilege π_{n+1} and V (logically but illegitimately) indicates to R to A has π_{n+1}.

T3. **Dishonest Verification**. A may attempt to cause V (or an employee of V) to improperly verify and/or compare av-recordings or digital signatures included in av-signatures and av-certificates, by issuing an illegitimate value of a_T. This would achieve the same result as T5(c), T5(d), or T5(e), but without impersonating V.

T4. **Imitation of Legitimate User**. A may attempt to dress, make up, and speak like U in such a way that V (or an employee of V) cannot distinguish A from U.

T5. **Coercion of Legitimate User**. A may attempt to coerce U into generating a valid av-signature by showing forced consent for transaction details chosen by A; A could then reuse this av-signature and U's av-certificate to generate a valid av-ticket.

T6. **Impersonation of Relying Party (e.g. through Phishing)**. A may attempt to impersonate R using the following techniques:

(a) A impersonates R in such a way that U interacts with R believing that A is R (as in the case of phishing attacks). A can thereby use the credential information provided by U to access resources or services requested by U.

(b) A impersonates R in such a way that V believes that A is R (assuming, of course, that $R \neq V$). In this case, A either: (i) learns the value of a_T; or (ii) is able to make V believe that U interacts with R (while, in fact, it is not the case).

T7. **Impersonation of Verifying Party**. A may attempt to impersonate V in such a way that R interacts with A believing that A is V. A can thereby:

(a) use the credential information provided by R to access resources or services requested by U;

(b) learn the value of p_T, thereby compromising the privacy of U;

(c) illegitimately deny or grant U access to certain resources or services;

(d) issue an illegitimate value of a_T to deny R the privilege of granting certain resources or services to U (e.g. to reduce R's market share, and potentially increase that of other relying parties trusting V).

(e) issue an illegitimate value of a_T to illegitimately influence R to grant a known impersonator of U access to certain resources or services.

T8. **Reusable User Credential Theft/Cloning**. A may attempt to steal (or clone and return) U's reusable credentials (e.g. av-certificate). This does not suffice to generate valid av-signatures (hence av-tickets) in U's name.

T9. **PC-based Keyboard Logging**. A may attempt to surreptitiously record credential information typed by U using a user PC's keyboard. This does not suffice to generate valid av-signatures (hence av-tickets) in U's name.

T10. **PC-based Screen Logging**. A may attempt to surreptitiously record all information seen by U on a user PC's monitor (including, e.g., opened windows, mouse movements, and mouse clicks). This does not suffice to generate valid av-signatures (hence av-tickets) in U's name.

T11. **Replay of Identity Claim (e.g. Audiovisual Recording)**. A may attempt to replay a valid audiovisual sequence s_T in which U shows her face and explicit consent for transaction details τ. Such a replay would, however, be detected by V using E_V.

3.1.2 Threat Discussion

Considering the aforementioned threats and the fact that knowledge of text-based credential information (e.g. passwords) does not suffice to generate av-tickets on behalf of users (and thereby impersonate these users), we conclude that `VideoTicket` is resilient to five main classes of attacks: (R1) theft and cloning of av-certificates and user personal devices; (R2) PC-based keyboard logging; (R3) PC-based screen logging; (R4) replay of av-signatures; (R5) and network-based capture of userids and passwords (e.g. through phishing). In addition, we conclude from T1-T5 that `VideoTicket` detects IDF attempts when the following conditions are met: (D1) av-signatures are not undetectably forged;[12] (D2) B's digital signature is not forged;[13] (D3) V accurately verifies digital signatures or av-recordings; (D4) A is not able to successfully appear to be U, e.g. by dressing, looking, and speaking like U; (D5) U is not coerced into generating a correct av-signature against her will.

A might be discouraged from attempting to forge av-recordings because it may be practically infeasible to generate these automatically, or to reuse them for multiple transactions or users. (This contrasts with other classes of authentication or transaction authorization credential information, such as textual/graphical passwords, credit card numbers, or driver's license numbers.) Moreover, attacks based on the forgery of av-recordings may work in remote transactions (since neither R nor V see A), but may be difficult to carry in the case of on-site transactions (especially if R verifies the av-certificate of A). `VideoTicket` also uses difficult-to-predict transaction identifiers in order to increase the difficult of successful av-recording forgery.

In some applications (e.g. debit or credit card transactions, and driver's license-based ID verification), B may be assumed to: (1) honestly attempt to detect IDF; (2) be able to protect the confidentiality of its signing keys; (3) and use practically unforgeable digital signature algorithms. Given these assumptions, the remaining issues are whether: (a) V accurately verifies digital signatures and av-recordings; (b) U can be imitated by A; and (c) U can be coerced. If $V = B$ (as in the case of health and debit card transactions), D3 may be assumed. Note, however, that the effectiveness of humans at verifying av-recordings may decrease after a number of consecutive work hours. We encourage further research on this topic. An interesting question is whether U can be imitated by A in such a way that V cannot distinguish av-recordings of U and A. For instance: is it possible for A to go, in person, to R, provide a copy of U's av-certificate, and make $V = R$ believe that $A = U$? `VideoTicket` does not counter such attacks. Neither does `VideoTicket` counter attacks whereby U is coerced into generating valid av-signatures against her will.

3.2 Other Practical Considerations

On-Site Verifiability, Scalability, and Privacy. The proposed scheme is designed in such a way that av-tickets

[12]It is assumed that A is not able to forge audiovisual clips in which chosen victims speak chosen transaction details, with chosen voice characteristics, and in such a way that lips and face movements are convincingly synchronized with speech.

[13]`VideoTicket` does not rely on end-users' digital signatures.

can be verified on-site by credential relying parties.[14] This allows not only for a decentralized (hence more scalable) system with no single (verification) point-of-failure, but also for improved user privacy, since, for each user U, the scheme does not require a single (ID and privilege verification) party to track all of U's transactions. U must, however, show her face on av-signatures and av-certificates' av-recordings; this can be perceived as privacy invasive. `VideoTicket` is therefore not suitable for anonymous transaction approval; we focus on applications such as border control, and credit/health/identity card issuing and renewal, which often are non-anonymous.

Manageability and Security Requirements. `VideoTicket` does not employ user-specific signing keys; this implies that users do not need to generate (signature-related) public-private keys pairs, regularly request certification of public keys, and safeguard the privacy of private keys. Moreover, no infrastructure and processes are needed to revoke, announce the revocation, and request the revocation status of user-specific signing keys. Consequently, `VideoTicket` avoids known practical roadblocks of (large-scale) public key infrastructures.

Convenience for End-Users. Let T be a transaction occurring at time t, between a user U and a credential relying party R, located at a location L. If U specifies T's identifier when s_T is recorded, then this identifier need not be unique with respect to R, but to t, L and T's identifier. Consequently, T's identifier may be a positive integer smaller than q ($q \geq 1$), if R is engaged in q simultaneous transactions at time t. However, in order for T's identifier to be unpredictable, it ought to be chosen uniformly at random in a sufficiently large set (e.g. the set of 7-letter lower-case words composed of roman letters, which has a cardinality of about 8 billions). User studies are needed to determine how much time is required, in practice, by users to produce good-quality av-recordings under various conditions.[15]

Verification Outsourcing Capability. V may out-source (i.e. delegate to a trusted party) its verification responsibilities (e.g. for increased cost efficiency). This, however, may introduce privacy concerns, when delegate verifiers have access to users' transaction details.

Financial and Time Costs of Human Verification. The financial cost of av-recording verification by humans is potentially low: 60 online transaction verifications per hour, performed by an employee paid $12 per hour, grossly costs 50 cents per verification;[16] for on-site transactions in which $V = R$, 1-minute verification of a user's ID may be too long, but 30 seconds may be acceptable for some applications (e.g. credit, debit, health or student card issuing). The use of `VideoTicket` could be restricted to transactions whose non-careful examination could lead to costly identity theft. For example, card issuing transactions ought to be carefully examined since identity theft committed with cards whose existence is unknown to their legitimate owners can be difficult to detect. Depending on the time taken in practice to verify av-signature, VideoTicket may not be suitable for classes of transactions such as last minute bid on eBay.

[14] On-site verification here implies credential relying parties need not interact with any remote parties at transaction time.

[15] We envision the use of VideoTicket with a small number of recording devices per user, e.g. one camera-phone and one or two PCs.

[16] Av-ticket verification outsourcing may considerably reduce this estimate (e.g. by an order of magnitude). Several other factors (e.g. cost of equipment, staff training, facilities, and web or phone-based user assistance) should be considered for a complete cost analysis.

3.3 Automated Biometric Verification

Section 2 presents a version of `VideoTicket` designed to use human agents to verify av-recordings. Here, we discuss what may be more interesting for commercial practice (if feasible): the use of automated biometric technologies for verifying av-recordings. Biometric systems allow automatic identification or identity verification of individuals using behavioral or physiological characteristics [42]. `VideoTicket` can be classified as a biometric *verification* scheme, whereby system users assert an identity that needs to be verified. This differs from biometric *identification* in which the identity of an unknown person is sought through examination of a potentially large list of system user records. Biometric verification systems are commonly evaluated using their *false acceptance rate* (FAR) and *false rejection rate* (FRR), as well as their *failure to enroll* (FTE) and *failure to acquire* (FTA) rates [22] (see also [5] Chap. 10). The FTE and FTA rates are typically most affected by user training. For example, in a face recognition system, users who position themselves incorrectly before a camera, or in poorly-lit environment, may not be processed successfully due to the generation of FTE or FTA events at enrollment and processing time respectively. Once the biometric system has successfully captured and segmented the features (in this case, the face) from the image, a *match score* is generated, which is related to the likelihood of a match between the person in the live image and the enrolled image. The match score is compared with a threshold to make a match decision. The choice of the threshold affects the compromise between FAR and FRR. Either error rate may be made arbitrarily low at the expense of the other. Increased false acceptance typically leads to increased financial loss and decreased security,[17] while increased false rejection typically leads to increased user frustration and decreased usability. For this reason commercial biometric systems typically choose a threshold designed to minimize the FRR at a chosen acceptable FAR.

Biometric data captured by `VideoTicket` potentially provides a rich source of biometric information, some of which (e.g. facial and voice information) can be processed using available technologies, and some of which (e.g. gesture information) is the object of current research. `VideoTicket` might be extendable to a biometric processing system based on biometric fusion [38], i.e. the combination of multiple biometric features (e.g. face and voice information) into a single signal. Biometric fusion often provides improved error rates, when compared with systems processing (associated) single biometrics.[18]

[17] More impersonators are susceptible to be falsely accepted.

[18] Fusing *unreliable* biometrics into a single system does make

Biometric data captured by *VideoTicket* could be used for:

1. *Face Recognition from Still Images.* Face recognition from still images is one of the most well understood biometric modalities (coming second after fingerprint recognition, in terms of maturity of the industry [44]). Many large scale tests of face recognition performance have been conducted, such as the FERET [31], FRGC [33] and FRVT [30, 32] series of tests. Face recognition performance has shown continuing improvements over the past 10 years [1], and recent tests indicate that automated face recognition algorithm performance is sometimes equal to or better than that of *untrained* human evaluators [1, 28, 32].[19]

   Face recognition performance depends on the quality and size of input images [30]. The FRVT 2006 analyzed face recognition performance for very high, high, and low resolution images. The low resolution images correspond closely to the requirements of *VideoTicket* – 75 pixels between the centers of the eyes. This results in an (JPEG format) image file size of 10k. The error rates for low-resolution still face recognition with controlled pose and illumination are FRR= 2.39% at FAR= 1%. While these results are promising for `VideoTicket`, they assume *controlled* pose and illumination. These assumptions may not be realistic for `VideoTicket` applications. Hence, lower error rates are expected in practical settings where user pose and illumination are not controlled.

2. *Speaker and Speech Recognition.* Speaker recognition is the automatic verification of a speaker's identity from a voice recording of this speaker. Speaker recognition differs from speech recognition in that the former seeks the identity of the speaker, while the latter seeks to understand what she says [36]. Both types of recognition can be used for `VideoTicket`: the former for user authentication, and the latter to verify users'consent to given transaction details. Speaker recognition is sometimes divided into two categories [2]: 1) text dependent recognition, in which case the user is tested against a specific enrolled passphrase, and 2) text independent recognition, in which case the user is identified using different spoken words than those used during enrollment. The quantity of speech recording used for speaker recognition varies between these categories; text dependent recognition typically uses a passphrase of a few seconds, while text independent recognition can use minutes of voice data. Data captured by `VideoTicket` might be used for a form of speaker recognition lying between text-dependent and independent recognition. We are not aware of other speaker or speech recognition applications designed with requirements similar to those of `VideoTicket`, and wish to stimulate research on these requirements.

   Many large scale tests of speaker recognition performance have been conducted by NIST; the most recent report [23] indicates continual improvement of speaker recognition performance over the test years (1996–2003), with an achieved FRR of about 13% at FAR= 1% (for cellular telephone recordings).

3. *Face Recognition from Video.* Face recognition from video is a research area that has not yet seen the systematic testing observed for face recognition based on still face images. Video-based face recognition typically involves the analysis of video clips to identify high-quality image frames [24]. These frames are typically extracted, and passed to face recognition software which processes them.

   Another approach is to use the video data to build a parameterized face model [20, 45]. A concern with this approach is that the computational work required to build such models from video data may not be sustained by currently-available user PCs (e.g. for preprocessing on the user's side) [20, 45].

4. *Lip Movement, Speech and Gesture Matching.* Other biometric features captured by `VideoTicket` are lip movements, speech and face gestures. The analysis of these features may be used to better protect against attacks whereby a fraudulent voice signal is associated with a legitimate video clip. Previous work has examined the possibility of synthesizing lip movements from speech data and models, e.g. for computer graphics applications [34]. Lip-movement-to-speech matching has been proposed for use in liveness detection [40], and as a technique to enhance speech recognition [12, 43]. Little research appears to have focused on lip-movement-to-speech matching for biometric verification. We encourage further work in this direction.

It is difficult to obtain direct price estimates on the use of biometrics to extend a system like `VideoTicket`, e.g. because most face recognition vendors do not currently make pricing information publicly available. Our informal inquiries indicate that, after a biometric system has been built, the cost of each transaction (over a 3 year period) might be below 5 cents. Thus, an automated extension of `VideoTicket` might provide financial cost benefits over the system described in Section 2 (cf. our discussion above on `VideoTicket`'s financial requirements).

Based on the above review of biometrics research (as it applies to `VideoTicket`), we wish to stimulate research and obtain feedback on the viability of an automated extension of `VideoTicket` for commercial use. To this end, our initial impression is that, while recent advances in face recognition are promising, more research is needed to deal with the analysis of user statements of consent to transaction details.

4. EARLY PROTOTYPE REPORT

In order to partially demonstrate the feasibility of `VideoTicket`, we built a software prototype. Section 4.1 provides an overview of the prototype software. Section 4.2 presents lessons learned from our prototype implementation.

4.1 Implementation Overview

Our prototype consists of three software modules: the *issuing module*, the *transaction request module*, and the *verification module*. The *issuing module* (see Fig. 2 on the

the combined system more reliable than the individual biometrics. We envision the fusion of multiple *reliable* biometrics capturing several classes of user features.

[19]Note that `VideoTicket`, as described in Section 2, assumes the use of *trained* human verifiers.

left) is meant to be used by B to issue av-certificates for users who appear before B in person. This module provides an interface enabling B's operator to view and authenticate (through automated digital signature verification) an av-certificate presented by U as a proof of ID. The *transaction request module* (see Fig. 2 on the right) enables its user to record an av-signature, input textual transaction details, specify the location of an av-certificate, and send the associated av-ticket via email to a specified address. Finally, the *verification module* (see Fig. 4) enables a verifier to review the two av-recordings of an av-ticket (with respect to given textual transaction details). This last module may also be used to approve or reject a transaction request and send the associated transaction status to a specified email address.

The prototype was built in Java, using the Java Media Framework (providing audiovisual recording and playing capabilities), and the Java Mail API and Java Activation Framework (providing email processing capabilities). The user interface was done using the Swing API. 1024-bit DSA with SHA-1 was used for cryptographic operations (using the Java security API). Audiovisual recordings were generated using a 16-bit stereo linear track at 44.1 KHz sampling rate, and a CINEPAK video codec [41] with 320x240 JPEG frames, at a rate of 15 per second. The total code size (for the three modules) is 144Kb. The CPU requirements of cryptographic and av-recording operations were too small to be noticed by humans. However, the sending of email on a 2.8GHz P4 PC running Windows XP with 1Gb of RAM took around 10 seconds when two 15-second av-recordings (taking about 2.8Mb each) were sent as email attachments.

4.2 Lessons Learned

Implementing a prototype of `VideoTicket` helped us identify user interface features that might lead to security-oriented errors if not implemented adequately. These features and associated security errors are presented and briefly discussed below, in the form of lessons learned from our prototype implementation.

L1. **Distinguish av-Certificate Verification from av-Recording Comparison and Comparison of av-Signature and Transaction Details**. Since the correct verification of transaction requests depends on the correct completion of these three tasks, they should be visually distinguishable by av-ticket verifiers. For example, av-ticket verifiers could be required to check a box associated with each task, and then press a transaction approval decision button enabled only if the three previous tasks have been completed.

L2. **Provide Intelligible Transaction Details**. Av-ticket verifiers must compare textual transaction details with transaction details specified in av-signatures; otherwise, attackers may use av-signatures for unintended transactions. Transaction details should therefore be clearly presented (e.g. using adequate fonts) in identified classes of credential information (e.g. Issuer:*MyBank*; Shop:*MyStore*; etc).

L3. **Provide Means to Review av-Recordings Efficiently**. If av-ticket verifiers are not able to efficiently review av-recordings (e.g. using functionalities such as volume up/down, pause, play, fast forward, rewind, and stop), they may not detect forgery attempts of av-signatures.

L4. **Allow Undo of Transaction Request Approvals**. Since people make mistakes (e.g. when unintentionally approving transaction requests), av-ticket verifiers should be able to undo (within a predefined context, e.g. time period) their approval of transaction requests.

Our early-prototype lacks many of these security-oriented user interface recommendations. The prototype was meant to be an early proof-of-concept of the client end.

5. RELATED WORK

Maurer [25, 26] proposes the concept of digital declarations for court resolution of disputes over user liabilities in (high-value) digital contracts. The idea is that each user digitally signs, in addition to a digital contract, a digital recording of a conscious act related to the contract, in order to show user consent. If a user U denies a party C's claim that U has consented to a digital contract d, U goes to court, and requests that C presents: (a) a digital recording showing U's consent for d; (b) a valid digital signature b of d; (c) physical evidence (e.g., paper-based documents signed by a person whom trusted experts say is U) of U's commitment to honor contracts digitally signed with the key associated with that used to verify b. In contrast, `VideoTicket` is concerned with real-time (potentially independent) verification of users' identities and consent for on-site and remote transactions associated with low and high currency values, and authorized by arbitrary credential relying parties (including, but not limited to court judges). Moreover, `VideoTicket` does not primarily rely on cryptographic techniques to verify user commitments. (In particular, it does not use *user-specific* signing keys and the associated management and public-key infrastructures and processes.)

Mobiqa [27] proposes a scheme whereby barcodes displayed by user-held mobile phones are used to verify users' claims of identity and possession of privileges. Suppose that a user U wants to obtain a privilege π from a party B, then U goes to see B in person, B takes a still digital picture of U, and sends to U's mobile phone (via SMS) a barcode that encodes an identifier ID_U. Assume also (without loss of generality), that π grants U access to a concert. Then U goes to the entrance gate of the concert, at the appropriate time, and shows her mobile phone displaying the aforementioned barcode to the gate controller G. G scans the barcode, and uses the corresponding identifier ID to: (1) obtain, from a database controlled by B, a still photo; (2) obtain privileges associated with ID, and verify that they authorize access to the concert; and (3) verify that the still photo obtained from B's database is a photo of the person showing the phone. G then lets U in, if and only if these three conditions are met. In contrast, `VideoTicket` uses av-recordings instead of still pictures (i.e. richer identifying information); utilizes trusted digital signatures to allow third parties to independently verify user ID; and handles both on-site and remote access to services or resources, through the use of transaction details combined with av-recordings of users' consent for specific transaction details.

The SecurePhone project [14] aims to design and implement

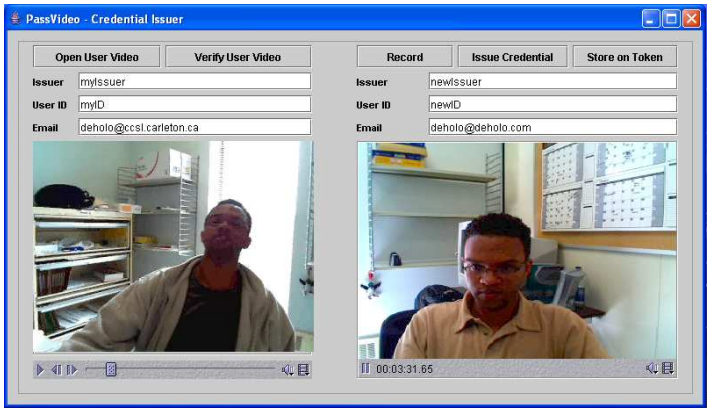

Figure 2: Issuing Module.
With the issuing module, the user reviews and verifies a given av-certificate, records an av-recording, and issues a new av-certificate.

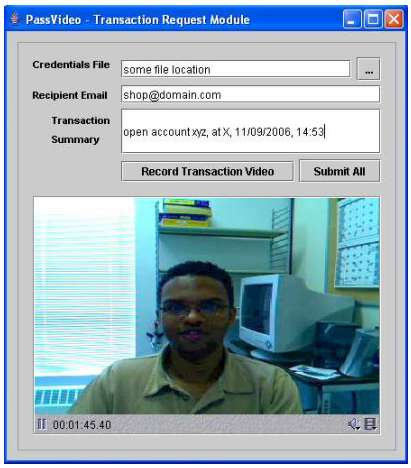

Figure 3: Transaction Request Module.
With the transaction request module, the user records an av-signature, inputs transaction details, selects an av-certificate, and sends an associated av-ticket via email.

Figure 4: Verification module.
User verifies av-certificate and av-signature associated with given av-ticket.

a mobile communication system enabling users to perform legally-binding transactions during cell phone-based conversations. To achieve this, users are first authenticated by their phones, using a (local or remote) multi-modal biometric verifier that examines users' facial, voice, and digitized signature. Each authenticated user U is then given access to cryptographic services provided by her phone's SIM card. Then, these services are employed to issue user-specific digital signatures, and thereby facilitate legally-binding transactions between U and other biometrically-authenticated users talking with U over a phone channel. Unlike `VideoTicket`, the SecurePhone project therefore relies on the following: a large-scale PKI with user-specific signing keys and related (deployment, maintenance, revocation, and revocation notification) infrastructure and processes; user-specific digital signatures (for non-repudiation); non-human-based (user-to-phone) authentication; and phone-based conversations with no visual (face) presentation of communicating parties.

Koreman et al. [17] report on an 84-subject experimental evaluation of a multi-modal biometric technique authenticating users who read prompts into a camera and microphone, and script sign on a touch screen. This technique is reported to have 0.8% equal error rate; the statistical confidence level of this figure is not provided. Karam et al. [13] describe a method allowing an impostor to guide, in real-time, the facial movements and speech of a synthetic face mimicking a chosen person of whom sufficient audiovisual information has been collected in advance. This imposture method is reported to have a 26% equal error rate with 2% of statistical uncertainty. The effectiveness of each impersonation instance is determined by an algorithm whose empirical effectiveness is currently being studied. It therefore remains unclear whether the proposed imposture method would effectively fool a human verifier (e.g. one used in an instantiation of `VideoTicket`).

Gentry et al. [10] propose a general framework for using distributed online human communities to solve problems that are difficult to solve by computers and easier to solve by humans. Av-ticket verification may fall into this class of problems. If so, `VideoTicket` can be viewed as a scheme using an instance of the general concept of distributed human computation.

Cyphermint [6] proposes an authentication scheme whereby: each user U goes to a trusted kiosk R; R takes a still picture of U, and sends this picture to a back-end server V operated by a human verifier; the verifier compares this photo to another picture provided by U in a preliminary registration phase; if the two photos match, the human verifier authenticates U, and V authorizes U to perform financial transactions at R. In contrast, `VideoTicket` uses audiovisual recordings, simultaneously performs user authentication and transaction authorization, and can be used both for remote and on-site transactions (since it does not rely on trusted user terminals).

Choudhury et al. [4] described an algorithm for person verification from audiovisual clips, which achieved 100% verification rate on real-time input from 26 users. In 2000, Matas et al. [8] reported on a person verification contest using still pictures, audio sequences, and video clips from 295 users; the best algorithm in this contest achieved false rejection rates of 2.5% and .8% for false acceptance rates of 2.3% and 46% respectively.

Previous work comparing automated and human person verification includes: models of strategies used by people to recognize and process faces [9, 11, 29, 37]; evaluation of supermarket cashiers'performance at identifying shoppers from photos on credit cards [15]; evaluation of people's ability to match poor-quality video footage against high-quality photographs [19]; and comparison of human vs. automatic face recognition based on still photos [1].

6. CONCLUDING REMARKS AND DISCUSSION POINTS FOR NSPW

This paper focuses on identity fraud detection in both remote and on-site transactions, regardless of applications. The IDF detection proposal can be seen as a novel alternative to digital signatures. Av-signatures may be seen as analogous to digital signatures, and av-certificates as analogous to public-key certificates. As a comparison, digital signature verification typically involves two steps: (1) signature correctness verification with respect to a certified public key (this is similar to av-signature transaction detail verification and comparison with a certified av-recording); and (2) public-key certificate trust verification (which is similar to av-certificate trust verification).

We wish to stimulate research on other non-cryptographic and cryptographic techniques combining user authentication with transaction authorization. An open question is whether it is possible to build a reliable IDF detection scheme that combines these two security goals while hiding the face and voice of legitimate users (e.g. for improved user privacy). Another open question is whether it is possible, given sufficient audiovisual information on a real person, to animate, in real-time, a virtual upper body that speaks, and moves its face and hands like that person, in such a way that the virtual person is indistinguishable from the real one by trained human verifiers. Technology enabling such impersonation could be used to forge av-signatures. Another research direction stemming from our work is the design of fully-automated multi-modal biometric schemes combining hand signs, with other biometric features such as face and voice. We are not aware that such schemes exist; they would help automate transaction request verification in the context of `VideoTicket`. This may reduce the temporal and financial cost of transaction verification in `VideoTicket`, and lower the risks of insider attacks whereby transaction requests that should be rejected are approved.

Automated multi-modal biometric schemes may combine face verification of still pictures (extracted from av-signatures) with speech verification of audio clips (extracted from the same av-signatures). An open question is whether the resolution of still pictures used for face verification could be made sufficiently high in practice to provide low false acceptance and rejection rates. One may also consider using a system whereby human verifiers review av-signatures rejected by an automated biometric subsystem. Such a hybrid system could provide cost benefits (due to the use of automated verification), while maintaining a desired level of

detection effectiveness (gained from human verification).

Another open question is whether is whether trained humans can perform better than automated schemes at recognizing people from speech or multi-modal biometric features. With respect to this point, we wish to stimulate research on the need for user studies involving both human attackers attempting to impersonate legitimate users, and human verifiers adequately compensated for discovering impersonators. We also wish to stimulate research on the long-term value of the "human-authenticating-human" approach to user authentication.

Acknowledgments: The first author thanks Michael Hu for helpful discussions improving a previous version of this paper, and acknowledges partial funding from the Ontario Research Network for E-Commerce. The second author acknowledges NSERC for funding an NSERC Discovery Grant and his Canada Research Chair in Software and Network Security. The third author acknowledges funding from NSERC. All authors thank NSPW'07 referees and attendees for fruitful suggestions and discussions on several aspects of a previous version of this paper.

7. REFERENCES

[1] A. Adler and M. Schuckers. Comparison of Human versus Automatic Face Recognition Performance. *IEEE Transaction on Systems, Man and Cybernetics (to appear)*, Feb 2007.

[2] J. Campbell. Speaker recognition: a tutorial. *Proceedings of the IEEE*, 85:1437–1462, 1997.

[3] N. Chou, R. Ledesma, Y. Teraguchi, D. Boneh, and J. Mitchell. Client-Side Defense Against Web-Based Identity Theft. In *Annual Network and Distributed System Security Symposium (NDSS '04)*, 2004.

[4] T. Choudhury, B. Clarkson, T. Jebara, and A. Pentland. Multimodal Person Recognition using Unconstrained Audio and Video. In *International Conference on Audio and Video-Based Biometric Person Authentication*, pages 176–180, 1999.

[5] L. Cranor and S. Garfinkel. *Security and Usability.* O'Reilly Media, Inc., August 2005.

[6] CypherMint. CypherMint PayCash System. `http://www.cypermint.com`. Site accessed in Nov. 2006.

[7] R. Dhamija and J. D. Tygar. The Battle Against Phishing: Dynamic Security Skins. In *Symposium on Usable Privacy and Security (SOUPS '05)*, pages 77–88. ACM Press, 2005.

[8] J. M. et al. Comparison of Face Verification Results on the XM2VTS Database. In *International Conference on Pattern Recognition*, volume 4, pages 858–863. IEEE Computer Society, 2000.

[9] N. Furl, A. O'Toole, and P. Phillips. Face recognition algorithms as models of the other race effect. *Cognitive Science*, 96:1–19, 2002.

[10] C. Gentry, Z. Ramzan, and S. Stubblebine. Secure distributed human computation. In *ACM Conference on Electronic Commerce (EC '05)*, pages 155–164. ACM Press, 2005.

[11] P. Hancock, B. Bruce, and M. Burton. A Comparison of two Computer-Based Face Identification Systems with Human Perceptions of Faces. *Vision Research*, 38:2277–2288, 1998.

[12] K. Iwano, T. Yoshinaga, S. Tamura, and S. Furui. Audio-Visual Speech Recognition Using Lip Information Extracted from Side-Face Images. *EURASIP Journal on Audio, Speech, and Music Processing*, 2007:Article ID 64506, 9 pages, 2007. doi:10.1155/2007/64506.

[13] W. Karam, C. Mokbel, H. Greige, G. Aversano, C. Pelachaud, and G. Chollet. An Audio-Visual Imposture Scenario by Talking Face Animation. In *Nonlinear Speech Modelling*, volume 3445 of *Lecture Notes in Artificial Intelligence*, pages 365–369. Springer-Verlag, 2005.

[14] W. Karam, C. Mokbel, H. Greige, G. Aversano, C. Pelachaud, and G. Chollet. SecurePhone: A Mobile Phone with Biometric Authentication and e-Signature Support for Dealing Secure Transactions on the Fly. In *Mobile Multimedia/Image Processing for Military and Security Applications*, volume 6250, pages 365–369. SPIE, 2006.

[15] R. Kemp, N. Towell, and G. Pike. When Seeing Should Not be Believing: Photographs, Credit Cards and Fraud. *Applied Cognitive Psychology*, 11:211–222, 1997.

[16] E. Kirda and C. Kruegel. Protecting Users Against Phishing Attacks with AntiPhish. In *Computer Software and Applications Conference (CSAC '05)*, pages 517–524, 2005.

[17] J. Koreman, A. Morris, D. Wu, S. Jassim, H. Sellahewa, J.-H. Ehlers, G. Chollet, G. Aversano, H. Bredin, S. Garcia-Salicetti, L. Allano, B. L. Van, and B. Dorizzi. Multi-modal Biometric Authentication on the SecurePhone PDA. In *Workshop on Multimodal User Authentication (MMUA '06)*, 2006.

[18] K. Kursawe and S. Katzenbeisser. Computing Under Occupation. In *New Security Paradigms Workshop (NSPW '07)*, 2007.

[19] C. Liu, H. Seetzen, A. Burton, and A. Chaudhuri. Face Recognition is Robust with Incongruent Image Resolution: Relationship to Security Video Images. *Applied Experimental Psychology*, 9:33–41, 2003.

[20] X. Liu and T. Chen. Video-Based Face Recognition Using Adaptive Hidden Markov Models. In *Computer Vision and Pattern Recognition*, volume 1, pages 340–345. IEEE Computer Society, 2003.

[21] M. Jakobsson. Modeling and Preventing Phishing Attacks, 2005. Phishing Panel in Financial Cryptography.

[22] A. J. Mansfield and J. L. Wayman. Best Practices in Testing and Reporting Performance of Biometric Devices: Version 2.01, 2002. National Physical Laboratory (UK). Report CMSC 14/02. `http://www.cesg.gov.uk/site/ast/biometrics/media/BestPractice.pdf`. Site accessed in April 2007.

[23] A. Martin and M. Przybocki. NIST Speaker Recognition Evaluation Chronicles, 2004. NIST. `http://www.nist.gov/speech/publications/papersrc/ody2004NIST-v1.pdf`. Site accessed in April 2007.

[24] A. Martin and M. Przybocki. Biometric Sample Quality Standard Draft (Revision 4), 2005.

International Committee for IT Standards. Document number M1/06-0003. http://www.nist.gov/speech/publications/papersrc/ody2004NIST-v1.pdf. Site accessed in April 2007.

[25] U. Maurer. Intrinsic Limitations of Digital Signatures and How to Cope with Them. In *International Conference on Information Security (ISC '04)*, volume 2851 of *Lecture Notes in Computer Science*, pages 180–192. Springer-Verlag, 2003.

[26] U. Maurer. New approaches to digital evidence. *Proceedings of the IEEE*, 92(6):933–947, 2004.

[27] Mobiqa. Mobi-Pass, 2006. http://www.mobiqa.com/prod_pass.html. Site accessed in July 2006.

[28] A. O'Toole, P. Phillips, F. Jiang, J. Ayyad, N. Pénard, and H. Abdi. Face Recognition Algorithms Surpass Humans Matching Faces Across Changes in Illumination. 2007. In Press. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[29] A. O'Toole, D. Roark, and H. Abdi. Recognizing Moving Faces: A Psychological and Neural Synthesis. *Trends in Cognitive Sciences*, 6:261–266, 2002.

[30] P. Phillips, P. Grother, R. Micheals, D. Blackburn, E. Tabassi, and M. Bone. Face Recognition Vendor Test 2002 Overview and Summary, 2003. NIST. http://www.frvt.org/DLs/FRVT_2002_Overview_and_Summary.pdf. Site accessed in April 2007.

[31] P. Phillips, A. Martin, and C. Wilson. An Introduction to Evaluating Biometric Systems. *IEEE Computer*, 33:56–63, 2000.

[32] P. Phillips, W. Scruggs, A. OŠToole, P. Flynn, K. Bowyer, C. Schott, and M. Sharpe. Face Recognition Vendor Test 2006 and Iris Challenge Evaluation 2006 Large-Scale Results, 2007. NIST. http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf. Site accessed in April 2007.

[33] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the Face Recognition Grand Challenge. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 1, pages 947–954. IEEE Computer Society, 2005.

[34] G. Potamianos, C. Neti, and S. Deligne. Joint Audio-Visual Speech Processing for Recognition and Enhancement. In *Conference on Audio-Visual Speech Processing*, pages 95–104. International Speech Communication Association, 2003.

[35] J. S. . Research. 2006 Identity Fraud Survey Report, 2006. http://www.javelinstrategy.com/research. Site accessed in June 2006.

[36] D. Reynolds. An Overview of Automatic Speaker Recognition Technology. In *International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 4072–4075. International Speech Communication Association, 2002.

[37] D. Roark, A. O'Toole, and H. Abdi. Human Recognition of Familiar and Unfamiliar People in Naturalistic Video Analysis and Modeling of Faces and Gestures. In *International Workshop on Analysis Models for Faces and Gestures*, pages 36–41. IEEE Computer Society, 2003.

[38] A. Ross, A. Jain, and J.-Z. Qian. Information Fusion in Biometrics. In *Audio- and Video-Based Biometric Person Authentication*, volume 2091 of *Lecture Notes in Computer Science*, pages 1611–3349. Springer-Verlag, 2001.

[39] B. Ross, C. Jackson, N. Miyake, D. Boneh, and J. Mitchell. Stronger Password Authentication Using Browser Extensions. In *USENIX Security Symposium*, pages 17Ű–32, 2005.

[40] S. Schuckers, L. Hornak, T. Norman, R. Derakhshani, and S. Parthasaradhi. Issues for Liveness Detection in Biometrics. 2002. Biometrics Consortium Conferene. http://www.biometrics.org/html/bc2002_sept_program/2_bc0130_DerakhshabiBrief.pdf. Site accessed in April 2007.

[41] C. Technologies. Cinepak. http://www.cinepak.com/begin.html. Site accessed in Aug. 2006.

[42] J. Wayman. A Definition of "Biometrics", 2001. National Biometrics Test Center Colllected Works. http://www.engr.sjsu.edu/biometrics/nbtccw.pdf. Site accessed in April 2007.

[43] T. Yoshinaga, S. Tamura, K. Iwano, and S. Furui. Audio-Visual Speech Recognition Using Lip Movement Extracted from Side-Face Images. In *Conference on Audio-Visual Speech Processing*, pages 117–120. International Speech Communication Association, 2003.

[44] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 35(4):399–458, 2003.

[45] S. Zhou and R. Chellappa. A Robust Algorithm for Probabilistic Human Recognition from Video. In *Computer Vision and Pattern Recognition*, volume I, pages 226–229. IEEE Computer Society, 2002.