



ELEMENTIVE

Modeling Cryptographic Properties of Voice and Voice-based Entity Authentication

Giovanni Di Crescenzo

Telcordia Technologies

giovanni@research.telcordia.com

Prepared for:

3rd ACM-CCS 07 Workshop on
Identity Management
November 2nd, 2007

Co-authors:

Munir Cochinwala,
Hyong S. Shim
(Telcordia Technologies)

Summary of Presentation

- Problem area and motivation
- Entity Authentication
- Voice: A business case
- An abstraction of voice
 - Modeling (some) non-cryptographic properties of voice
 - Modeling (seemingly) cryptographic properties of voice
- Entity-Authentication Protocols
 - Voice-based protocols
 - Voice-and-password-based protocols
- Conclusions

Problem Area and Motivations

Identity Management: a confluence of (at least) 3 perspectives

Creation,
management
and deletion
of identities
as unique
names

**Pure Identity
Paradigm**

(This paper)

Using biometrics (voice) and/or
secrets (passwords) as **identities**
to gain access to a **service**

Entity Authentication: a technical
cornerstone of identity
management

**Entity Authentication
Paradigm**

Strong and/or Multi-Factor Authentication:
currently preferred approaches to entity authentication

Biometrics (and, in particular, Voice):
one of, or the main factors in multi-factor authentication

Delivering
personalized,
role-based,
online,
on-demand,
multimedia content,
presence-based services
to users and their devices.

**Service
Paradigm**

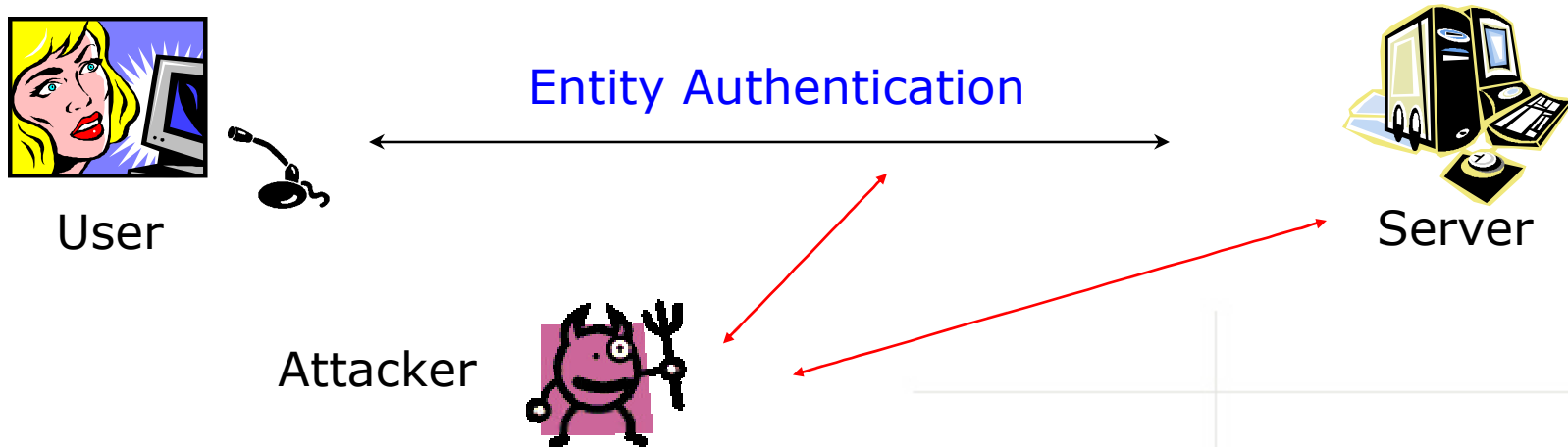
A Business Case for Voice

- IP networks have transformed **service provision**, lowering the barriers to entry for **new service providers** and introducing **new service paradigms**
 - Voice services are now being delivered commercially over the internet (e.g. Voice over IP)
- **Authentication** is crucial for the security (and success) of these paradigms
- **Network operators** have the ability to authenticate end users, increasing their advertising worth and reducing fraud for content providers
- **Telecom operators** can use their services and capabilities to provide authentication (e.g., voice recognition along with SIM card authorization)

Entity Authentication

- Users want to use their identity or any appropriate credentials to gain access to services
- **Traditional methodologies:** Single-factor solutions for authentication include something you (and hopefully only you):
 - “**know**” (e.g., passwords, personal data, personal history, credit report)
 - “**have**” (e.g., token content)
 - “**stored**” (e.g., secrets, cryptographic keys)
- **More recent methodologies:**
 - strong and/or multi-factor authentication
 - something you “**are**” (e.g., biometrics)
 - single sign-on based on strong authentication

Voice-Based Entity Authentication



- A **voice-based entity authentication** protocol should satisfy:
 - **Correctness**: an authorized user is successfully authenticated by the server
 - **Security against entity impersonation**: an attacker is able to impersonate a user and be successfully authenticated by the server only with negligible probability
 - Attacker can overhear the user's voice samples, play as the server, intrude into part of the server's storage, and finally make an impersonation attempt
- Other practical requirements include: **usability** and **reliability**.

Security and Usability within Entity Authentication

- All forms of authentication exhibits tradeoffs between usability and security depending on deployment and attack scenarios
- Take **passwords**, for example:
 - High usability and security for computer users within organizations' networks
 - Low usability and security in many other scenario (e.g., online banking), as they are easily forgot, shared, stolen, etc.
- **Biometrics** are often invoked as a form of authentication that improves security but at the cost of reducing usability in many scenarios
- **This paper's approach:** strengthen security of high-usability forms of biometrics, such as **voice**

A Voice Abstraction

- Modeling (non-cryptographic) properties of a human factor
- For any voice sample, the associated voice signal may depend on a large number of factors which we cluster into 3 groups:
 - the **human** that produces it
 - which **time** it is produced at
 - under which **noise** conditions it is produced
- **Voice abstraction: voice sampling and voice representation algorithms**
 - **Voice sampling**: maps words from a dictionary into voice signals; parameterized by a (unique) identity and a current time, and probabilistic (to model noise)
 - **Voice representation**: maps voice signals into “voice values” in a metric space
- **Two basic properties of voice abstractions:**
 - **Density**: two voice samples produced by the same human, using different words (even at similar times) result in ‘sufficiently distinct’ voice values
 - **Accuracy**: two voice samples produced by the same human at different times (even using the same word) result in ‘sufficiently similar’ voice values

Voice Abstraction – formal definition

- Let VS and VR be a voice sampling and a voice representation algorithm, respectively, and let d, a be positive real values.
- We say that pair (VS, VR) is a **voice abstraction** with parameters (d, a) if the following properties hold:
 - **d-density:**
 - for any identity id , time t and distinct words $w(1), w(2)$ in D , it holds that $\text{dist}(v(1), v(2)) \geq d$, where $v(i) = (v_{\{i1\}}, \dots, v_{\{ir\}}) = VR(VS_{\{id, t\}}(w(i)))$, for $i=1, 2$.
 - **a-accuracy:**
 - for any identity id , word w and distinct times $t(1), t(2)$, it holds that $\text{dist}(v(1), v(2)) \leq a$, where, for $i=1, 2$, it holds that $v(i) = (v_{\{i1\}}, \dots, v_{\{ir\}}) = VR(VS_{\{id, t(i)\}}(w))$.

Voice: a Cryptographic Primitive?

- Modeling Cryptographic Properties of a Human Factor
- Voice Cryptographic Primitive: a Voice Abstraction satisfying the following two seemingly cryptographic properties
- Impersonation hardness:
 - it seems hard to find a human producing voice samples that are, in some formal sense, indistinguishable from another human's voice samples.
- Extrapolation hardness:
 - it seems hard to extrapolate the digital representation of a human's voice sample given some digital representations of distinct voice samples from the same human.

Voice Cryptographic Primitive – formally

- Let pair (VS, VR) be a voice abstraction with parameters (d, a) , and let $\text{eps}(i), \text{eps}(e)$ be positive real values. We say that (VS, VR) is a **voice cryptographic primitive** with parameters $(d, a, \text{eps}(i), \text{eps}(e), q)$ if the following properties hold:
 - **$(\text{eps}(i), q)$ -impersonation hardness**: for any identity id and any adversary algorithm A making at most q queries to its oracle, the experiment $\text{ImpExp}\{VS, VR, A\}$, defined below, returns 1 with probability $\leq \text{eps}(i)$.
 - **$(\text{eps}(e), q)$ -extrapolation hardness**: for any identity id and any adversary algorithm A making at most q queries to its oracle, the experiment $\text{ExtExp}\{VS, VR, A\}$, defined below, returns 1 with probability $\leq \text{eps}(e)$.

- $\text{ImpExp}\{VS, VR, A\}(id, a)$:
 - 1. $(id', t1, t2, w1, w2) \leftarrow A^{\{VS_{\{id, ct\}}\}}(id, a)$
 - 2. let $v(1) = VR(VS_{\{id, t1\}}(w1))$
 - 3. let $v(2) = VR(VS_{\{id, t2\}}(w2))$
 - 4. if id and id' are different and $\text{dist}(v(1), v(2)) \leq a$ then return 1 else return 0

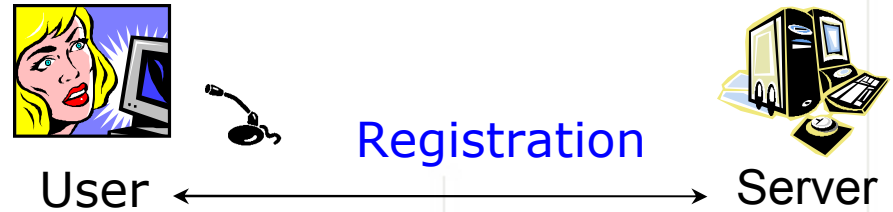
- $\text{ExtExp}\{VS, VR, A\}(id, a)$:
 - 1. $v \leftarrow A^{\{VS_{\{id, ct\}}\}}(id, a)$
 - 2. let $(w(1), \dots, w(q))$ be the words queried by A to VS
 - 3. let $v(i) = VR(VS_{\{id, t\}}(w(i)))$, for $i = 1, \dots, q$
 - 4. if $\text{dist}(v, v(i)) > a$ for $i = 1, \dots, q$ then
 - if there is a w in D s.t. $v = VR(VS_{\{id, t\}}(w))$ then return: 1 else return: 0

A First Voice-Based Entity Authentication Protocol

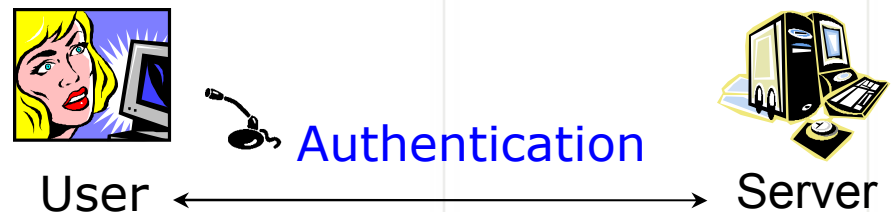
Protocol Design and Properties

- **Single voice sample transmission protocol**
 - Analogue (or closest) of digital password protocol
 - Proof of security validates our model
- **Provably secure assuming a voice cryptographic primitive**
 - Adversary successfully authenticating either manages to do so for a previously seen id (→ violates extrapolation hardness) or for a new id (→ violates impersonation hardness)
 - Quantifiable reduction parameters
- **Correctness**
 - Ok until you exhaust previously stored voice samples

Protocol Structure



Server stores multiple and “sufficiently distant” voice samples from user.



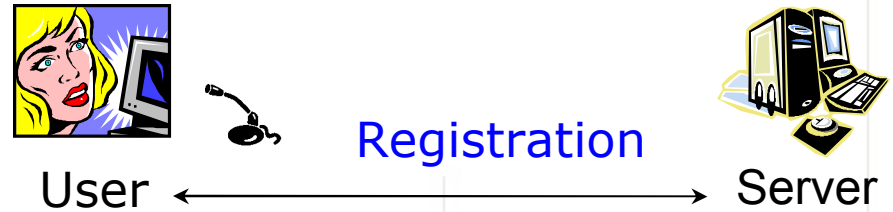
User is asked to produce a previously stored and not yet used voice sample. Server verifies sample authenticity by matching closeness with a previously stored value.

A Second Voice-Based Entity Authentication Protocols

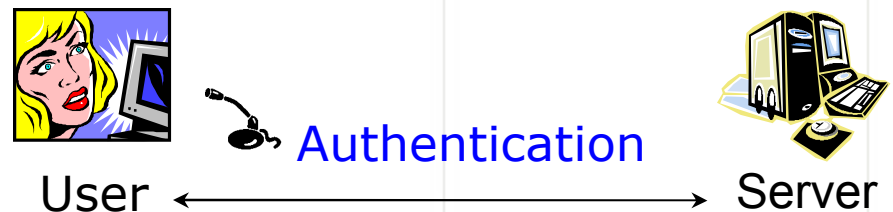
Protocol Design and Properties

- **Multiple voice samples transmission protocol**
 - At registration user stores a larger number of voice samples
 - During authentication, user produces a carefully chosen subset of voice samples
 - New combinatorial objects:
 - Implicitly samplable superpolynomial-size family of cover-free sets
- **Correctness ok for any polynomial number of sessions**
- **Provably secure against adversary witnessing up to an arbitrary polynomial q of sessions, assuming a voice cryptographic primitive**
 - Number of voice samples required per session only $\sim q \log n$

Protocol Structure



Server stores multiple and “sufficiently distant” voice samples from user.



User is asked to produce a random cover-free subset of stored voice samples. Server verifies sample authenticity by matching closeness of all samples with previously stored values.

Voice-and-Password-Based Entity Authentication Protocols

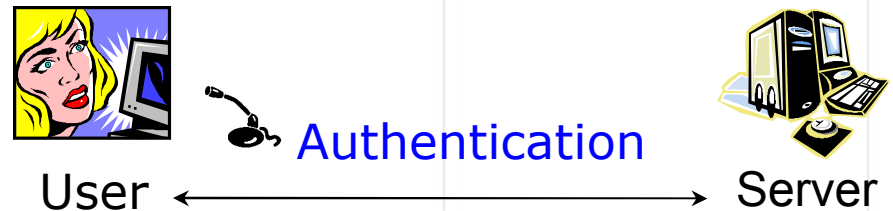
Protocol Design and Properties

- **Goal:** Security against both impersonation and server storage intrusion (bounded retrieval model)
- We combine **two types of protocols**:
 - Our voice-based protocols
 - Previous password protocols secure against partial server storage corruption in the bounded retrieval model
- Using **dispersers** and **pairwise independent hash functions**
- **Usability** remains the **same as before**
 - Only server's program more involved
- **Scheme is provably secure against both impersonation and intrusion assuming the existence of a voice cryptographic primitive**
 - Quantifiable reduction parameters

Protocol Structure



Server stores one or multiple voice samples and a tag based on the user's password.



User sends password and previously stored voice sample (or subset of them).
Server verifies authenticity.

Conclusions

- **Entity Authentication** is a technical cornerstone of identity management
- **Voice-based Entity Authentication** has attractive usability, applicability, and security properties
- We started the area of **quantitative design and analysis** of voice-based entity-authentication and its security
 - High potential for **quantifiable** security in **practical** systems
- Research open problems abound