Two-Factor Web Authentication Via Voice

Jörg Tacke & Andreas Wolf

VOICE.TRUST AG[1]
Landshuter Allee 12-14, 80637 Munich, Germany
{jt|aw}@voicetrust.de

Abstract: The increasing pressure for network accessible information and business applications demands secure, interoperable, cost efficient and user self-service authentication procedures. This paper describes the authors' work on 2-factor authentication for Web applications encompassing speech biometrics, and its relevance to security and cost effectiveness.

1 Introduction

The development of Internet communication shows a strongly increasing number of available e-business applications and network enabled transactions. Sensitive data as well as personal data involved in these procedures demand privacy and strong authentication. Passwords are the generally applied method of network authentication. The lack of security [Sb00] has led to two factor systems using tokens as a second factor of security additionally to passwords. The security impact and cost related to issuing and handling a physical device motivated the development of 2-factor-authentication deploying a password and speech biometrics of a user.

2 Authentication and Related Work

Authentication services have the primary task to prove and to guarantee the claimed identity of an accessing user according to the identities known at the accessed IT infrastructure. Authentication services should support the operator of an IT application to solve the following tasks:

- Quality determination of the executed authentication schema. Schemes differ with respect to theoretical security, user trust, and acceptance. Schema types are *What I know*, *What I have*, and *What I am*. The later are the biometric schemes.
- Provide data which can be used for user authorization or for session management, like a user ID.
- Support the integration into existing environments via standards like LDAP or JAAS.

Security. If access to resources shall be restricted or controlled, the determination of identities is essential. Therefore the quality of verification technologies has an increasing

[1] VOICE.TRUST AG, founded in July 2000, is the leading producer of high-security authentication solutions via digital voice verification. VOICE.TRUST solutions lead to a dramatic reduction in operating costs over conventional authentication solutions of up to 80%. Areas of use include secure authentication for PIN and password reset, Single Sign-On, remote access, or 2-factor authentication in the areas of network security, call center, mobile banking and e-commerce.

importance. Speech biometrics based authentication provides adequate security as well as user acceptance. The second factor demanding authentication is accountability for billing purposes. Rapidly growing non free internet services enter the mass market, so simple but secure techniques for Web based authentication are required.

Acceptance. Speech technologies requiring only telephone access may be used without any additional hardware. So they are excellent candidates for cost efficient authentication techniques. Furthermore, they do not force any changes in the usual behavior of the users (i. e., to make phone calls).

1- and 2-Factor Authentication. The major techniques for Web authentication are cookies or HTML rewriting using session IDs, mostly based on basic authentication. So the major Web authentication technique is based on passwords. This procedure is simple, but of limited security. The lack of security and the "secure password paradox" have led to the development of token-based 2-factor authentications. These systems combine passwords (What I know) with tokens (What I have) and increase the overall security due to almost independent errors and reduced misuse possibilities. But tokens have to be rolled out and are expensive. Speech capabilities are rolled out by birth, and the appropriate scanner device (phone) is already available almost everywhere. Only an enrollment of the users, i. e., the registration of their voices, is necessary and can be done easily via phone.

3 Biometrics

Utilizing knowledge (PIN, Password) as a first factor for authentication, often possession based factors are utilized as the second one (key, seal, smart card, token). Biometrics are the third factor in authentication. Active techniques like speech, signature, or behavior, and passive ones like face geometry, retina, iris, or fingerprint are available. In order to achieve an acceptable strength of authentication without the disadvantages associated with token-based systems, we have applied the authentication factor biometrics instead of possession.

Speech Biometrics. The authors suggest to choose speech as preferred biometric method in network based 2-factor authentication. Speech input devices are globally available, thus eliminating investment and management of additional equipment like scanners etc. Signal quality of telephone devices and transmission infrastructures are standardized and known. User acceptance is high; the user does not have to perform other tasks than performing a familiar one: speaking on the phone and answer questions. Overall error rates of speech authentication show good false acceptance rates in combination with acceptable false rejection rates, leading to high security and a high customer service rate based on total system performance [Aj01]. User support is available via the same media. In case of rejections or errors, the user is informed via voice, thus eliminating the need for extra devices like, e.g., displays. Rollout time and cost is reduced due to the lack of having to install physical devices. Training time is minimized due to online help.

Quality Impact of Telephony Devises. Typically, the human speech communication is limited to a bandwidth of 7-8 kHz. Bandwidth of ITU-T (former CCIT) standard telephony channel are limited to 3100Hz, from 300 to 3400 Hz [MAN01]. ITU-T standardized two PCM transmission modes. According to the scanning theorem, a minimal scanning frequency f_A is required to digitize an analogue signal. f_A must be at least twice the highest frequency f_S to be scanned: $f_A=2*f_S$. The required number of quantification intervals was empirically determined by the recognition quality of spoken syllables. Hence, by deploying binary coding, 256 intervals with a scanning rate of 8 khz and 8 bit coding were standardized, resulting in a bit rate of 64 kbit/s. Starting from the analogue CCITT telephone channel, the G.711 standard with a scanning frequency of 8 kHz and a scanning period of 125 ms was defined. Using various coding techniques, (CELP), bandwidth for IP- telephony reaches 8 kbit/s. Standards like G.729a (CISCO IP telephony) enable a bandwidth of 8 kbit/s, G.723 (Microsoft Netmeeting) equals 24 kbit/s. The quality of transmitted speech depends on bandwidth and the transmission mode inherent quality of service. Comparing indicator for quality of speech is the Mean Opinion Score (MOS) [RZN00]. MOS was derived from scientific audio testing, ranging from 1 (good) to 5 (poor).

MOS	Bandwidth	Codec
4,1	64 kbit/s	G.711
3,85	16 kbit/s	G.726
3,92	8 kbit/s	G.729a

Table 1: Mean Opinion Score

When selecting transmission devices, it is important to consider, that the Internet does not provide consistent bandwidth nor a guaranteed QoS. Hence, even if accepting the lower maximum transmitted frequency range, translating into poor authentication level of quality, consistent authentication quality at that level cannot be expected.

Speaker Verification. There are two main approaches for obtaining frequency information from a time domain signal such as a speech waveform. These are Fast Fourier Transformation (FFT) which converts time domain data to frequency domain data, and autoregressive modeling which processes time domain data to obtain parameters that are representative of the frequency domain data [DHP02]. In order to verify a speaker, he must enroll general speech (text independent) or specific words (text dependent). Text dependent verification tends to achieve better performance rates than text independent verification: Total mean error of 0.6% versus 0.065% [Lj99].

Speaker Authentication. Achieving user identity by a claimed ID such as user name via Web interface, speech verification will be executed only if user name and associated PIN or password match; then a call to the user is triggered, prompting the user for the required voice sample.

4 2-Factor Speech Authentication

The applied architecture strategies are to use available standards, thus avoiding proprietary interfaces or protocols, usage of available devices, procedures and technologies in order to grant maximum interoperability. Standards applied include TCP/IP for network communication, ISDN for telephony, ODBC and LDAP for database access.

Functionality. A user requesting access will enter his user name in the appropriate field and instead of entering his password, he will enter his PIN and his phone number in the password field. The Web or Web application server will initiate authentication against the 2-factor speech authentication server, which acts like a LDAP server. The 2-factor speech authentication server will match user name info and PIN. If data entered by the user matches data stored, the server will call the user at the entered telephone number. The user is prompted to speak short phrases required for multilevel verification depending on the desired security level. In case of positive verification the server returns positive authentication signal to the requesting server.

Interfaces. Depending on back end procedures deployed, we suggest to generally interact with APIs of applications requesting user authentication.

Authentication Protocols. To avoid the need of permanent interface integration, we realized communication via LDAP. LDAP allows database communication as well as communication with authentication protocols like TACACS+, Radius or RadiusX.

VoiceXML. Upcoming voice applications probably will apply VoiceXML 2.0 for application flow descriptions. If large parts of the application code are implemented using XML, porting to other hardware and IVR platforms will get much simpler than in the conventional case using proprietary dialogue description languages.

Performance. Depending on the security level required and thus the number of verifications, time per authentication takes between 10 and 30 seconds. First time users will listen to the full voice prompt. Experienced users will cut the voice prompt taking advantage of the barge in function, thus reducing call duration and authentication time. Multi-level speaker authentication performance test have shown operating points, where no errors (FAR; FRR) were observed for all volunteers [GT02]. Additional security provided by the first factor PIN was not part of the test.

Load Balancing and Scalability. Load balancing is performed by the telephony equipment as it handles the calls, or by the Web servers with their own load balancing techniques. Scalability primarily depends on the simultaneous authentication calls performed by the server. Tests show that a single 1 GHz processor can handle up to 120 simultaneous verification calls.

5 Conclusion

As a solution to missing secure, interoperable, cost efficient and user self servable Web Authentication procedures, we suggest two factor speaker authentication using telephony as scanning device, implemented on available standards to achieve high security, performance and user convenience. Presented test results show unachieved performance.

References

[Aj00] Ashbourn, J.: Biometrics: Advanced Identity Verification. pp. 70-78. Springer-Verlag

[DHP99] Deller, J. R., Hansen, J. H. L., and Proakis, J. G.: Discrete Time Processing of Speech Signals, pp. 225-408. IEEE Press Classic Reissue, Wiley Interscience

[GT02] Grans, K., Tekampe, N.: Dokumentation der Testumgebung und Testergebnisse Voice.Trust Server

[Lj99] Luettin,j.: Speaker Verification Experiments XM2VTS Database, IDIA Research Report

[MAN01]Vorlesung Rechnernetze Uni Mannheim: www.informatik.uni-mannheim.de/informatik/pi4/stud/veranstaltungen/ss2001/rechnernetze/rn2a.pdf, page 44

[RZN00] RZNet Solution Competence Center: www.rznet.de/SCCnews8.html?opendocument&expandview&count=99999

[Sb00] Schneier, B.: Secrets and Lies: Digital Security in a Networked World. Wiley Publishing