Sistema para Autenticação de Usuários por Voz em Redes de Computadores

Adriano Petry[1,2], Sidcley da Silva Soares[1], Dante Augusto Couto Barone[3]

[1]Área Tecnológica e Computação – Universidade Luterana do Brasil (ULBRA)
Rua Miguel Tostes 101 – Canoas – RS – Brasil

[2]Unidade Guaíba – Universidade Estadual do Rio Grande do Sul (UERGS)
Estrada Santa Maria 2300 – Guaíba – RS – Brasil

[3] Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Av. Bento Gonçalves 9500 Bloco IV – Porto Alegre – RS – Brasil

adpetry@ulbra.tche.br, sidcley@inf.ulbra.tche.br, barone@inf.ufrgs.br

Abstract. *This work describes a system for user authentication in computer networks using speech. Initially, it is given a general overview of the system, focusing in the client/server architecture used. After this, the main characteristics of the system building blocks are shown, concerning the techniques for speaker recognition, data base, client/server communication and management and user interface application. At last, the difficulties encountered are discussed and future works are mentioned.*

Resumo. *Este trabalho apresenta um sistema para autenticação de usuários por voz em redes de computadores. Inicialmente é dada uma visão geral do sistema, focando na arquitetura cliente/servidor utilizada. A seguir, são mostradas as principais características dos blocos constituintes, abrangendo as técnicas para reconhecimento de locutor, características da base de dados, comunicação cliente/servidor e aplicação responsável pelo gerenciamento e interface com o usuário. Por fim são discutidas as dificuldades encontradas e são apontados trabalhos futuros.*

1. Introdução

Nos últimos anos, empresas de diversos setores têm buscado novas formas de oferecer produtos e serviços a seus clientes. O uso da Internet para transações comerciais vem assegurando maior visibilidade de empresas por seus clientes, além de facilitar a comercialização de produtos em qualquer parte do mundo. A utilização bem sucedida da Internet para a disponibilização de produtos e serviços requer meios de garantir a segurança e privacidade das informações trocadas, além de uma inequívoca identificação e autenticação dos usuários. Diversas técnicas vem sendo desenvolvidas para a realização dessas tarefas. A criptografia desempenha um papel importante nesse contexto, possibilitando a troca segura de dados. A utilização de senhas robustas de usuários também contribui para diminuir as chances de sucesso dos ataques. Contudo, ainda é possível aumentar o grau de segurança oferecido, através do desenvolvimento

contínuo de novas técnicas e ferramentas que possibilitem realizar transações através de redes de computadores de forma mais confiável.

A autenticação de usuários pode ser realizada através de diferentes métodos, que apresentam características distintas. Uma autenticação através de uma informação secreta (ou senha) de conhecimento exclusivo do usuário é bastante utilizada atualmente. Esse tipo de metodologia apresenta vulnerabilidades relacionadas principalmente à escolha de senhas que podem ser facilmente descobertas, ou ao roubo da informação secreta utilizando, por exemplo, algum tipo de vírus de computador. Outra forma de dificultar a ação de fraudadores é a utilização de elementos físicos, como cartões magnéticos, para a realização da autenticação. Esse tipo de proteção atualmente é adotado em muitos sistemas que exigem alto grau de segurança, como em transações bancárias. Ainda assim, o extravio do dispositivo de acesso e a descoberta da senha secreta possibilitam a realização de transações remotas por pessoas estranhas, que efetuam o acesso como um usuário válido. A utilização de medidas biométricas do usuário para sua autenticação aparece como uma alternativa para diminuir as chances de falha em uma autenticação. A análise biométrica sem intervenção humana (automática) pode ser utilizada junto com os métodos já empregados ou até mesmo substituindo-os.

Alguns exemplos importantes de técnicas biométricas automáticas são o reconhecimento da íris, reconhecimento de face, análise da impressão digital, reconhecimento de voz, verificação das características geométricas da mão, análise da assinatura, e outros [Chirillo e Blaul 2003] [Woodward Jr., Orlans e Higgins 2003]. Cada técnica apresenta suas particularidades e indicações de aplicação. A autenticação biométrica através da voz, por vezes chamada de reconhecimento automático de locutor (RAL), possui algumas vantagens intrínsecas, quando comparada a outras técnicas biométricas. Pode-se salientar a facilidade de captura da medida biométrica (voz) utilizando hardware de baixo custo, a possibilidade de aquisição do sinal sem desconforto e sem a necessidade de contato físico do usuário (método pouco intrusivo), e a ausência de treinamento prévio para utilização do sistema. Da mesma forma que ocorre com outras técnicas, também apresenta desafios importantes a serem considerados, como questões relacionadas a patologias no aparelho vocal, à mudança no estado emocional do usuário, e ao controle sobre o ambiente de gravação.

Este trabalho apresenta uma aplicação de técnicas de RAL como medida biométrica para autenticação remota de usuários em redes de computadores. São descritas as principais características de um protótipo capaz de autenticar usuários por voz pela Internet, desenvolvido através de um projeto de pesquisa que está sendo realizado junto à Universidade Luterana do Brasil (ULBRA), com apoio do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Inicialmente é mostrada uma visão geral do projeto de pesquisa. A seguir, é detalhado o funcionamento do protótipo desenvolvido e, por fim, são apontados trabalhos futuros ainda a serem desenvolvidos no âmbito do projeto.

2. O Projeto RALNET

Com o objetivo de propor alternativas ao processo de autenticação de usuários, o projeto "RALNET: Desenvolvimento e Aplicação de Tecnologias de Reconhecimento Automático de Locutor para Autenticação de Usuários em Redes de Computadores"

propõe o uso de técnicas de RAL em sistemas computacionais conectados em rede, para autenticação remota de usuários. Iniciado em janeiro de 2003, esse projeto tem duração prevista de dois anos, e está sendo desenvolvido junto ao laboratório de Redes e Hardware da ULBRA em parceria com o Instituto de Informática da Universidade Federal do Rio Grande do Sul (UFRGS), contando com o apoio do CNPq.

Dentre os objetivos propostos nesse projeto, salientam-se o estudo de tecnologias de RAL aplicáveis em sistemas de segurança de redes de computadores, o planejamento de um sistema capaz de integrar as tecnologias de RAL e sistemas para transações pela Internet, e a construção e validação de um protótipo capaz de efetivamente utilizar tais tecnologias e efetuar a autenticação por voz de usuários localizados remotamente.

3. Sistema Proposto

O sistema construído funciona através de uma arquitetura cliente-servidor. O lado servidor conta com o acesso a um banco de dados para armazenamento e consulta de informações de usuários e padrões vocais. Além disso, a geração dos padrões de voz e a autenticação das informações extraídas a partir do sinal de voz do usuário é realizada também no lado servidor. A comunicação se dá através da Internet, utilizando técnicas de criptografia para troca de dados. No lado cliente, a interação com o usuário e captura do sinal de voz se dá pela manipulação de uma placa de som comum, que deverá estar presente no computador cliente. Parte do processamento digital do sinal de voz é realizado já no lado cliente, que envia apenas as informações extraídas a partir do sinal de voz, reduzindo assim o tráfego da rede e a carga de processamento no computador servidor. A figura 1 ilustra os principais blocos constituintes do protótipo desenvolvido, que são detalhados a seguir.

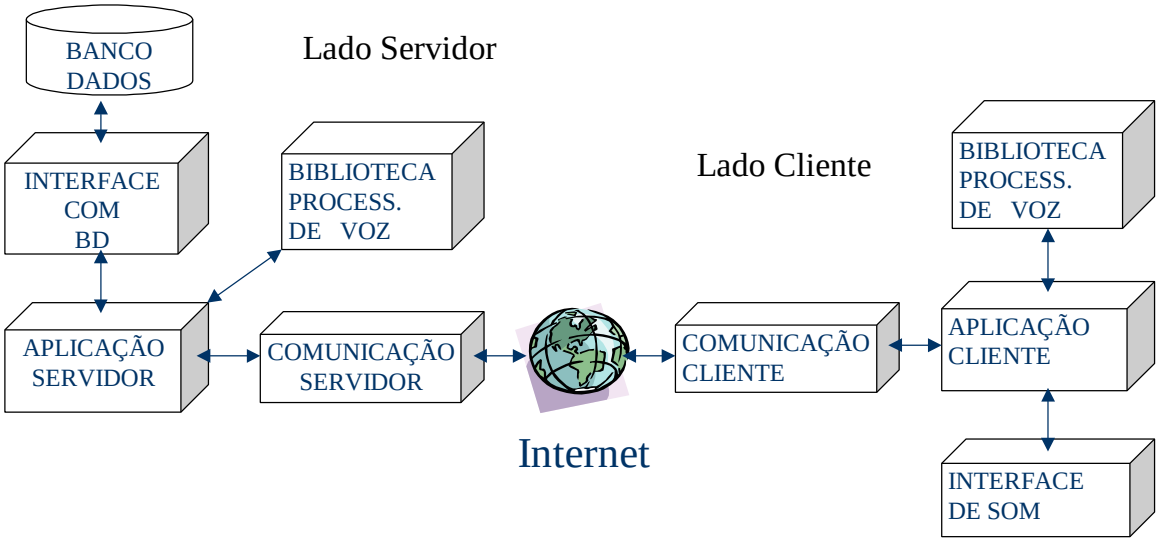

Figura 1. Ilustração da arquitetura utilizada na construção do protótipo de autenticação remota por voz

A linguagem de programação Java foi adotada para a construção do protótipo por oferecer recursos poderosos de programação para comunicação entre computadores, além de suportar diversas arquiteturas de hardware e sistemas operacionais. O ambiente de desenvolvimento utilizado foi o NetBeans IDE 3.5.1, com o kit de desenvolvimento

Java 2 SDK *standard edition* versão 1.4.1. Apenas a biblioteca de processamento de voz foi implementada utilizando uma linguagem de programação diferente, a linguagem C++, através da ferramenta Visual Studio 6.0. Essa escolha foi feita principalmente pela maior velocidade de processamento oferecida e possibilidade de reutilização de código já construído ao longo de trabalhos anteriores [Petry 2002]. O acesso à base de dados foi feito através de *Java Database Connectivity* (JDBC), utilizando um servidor de banco de dados MySQL [Horstmann e Cornell 2001].

3.1. Técnicas para RAL em uma Arquitetura Cliente-Servidor

As técnicas empregadas para o RAL são objeto de estudos e aperfeiçoamentos de pesquisadores há vários anos. Atualmente os algoritmos disponíveis oferecem taxas de reconhecimento elevadas, permitindo a construção de sistemas bastante confiáveis. Um sistema para RAL é composto basicamente pelas etapas de treinamento e reconhecimento. A etapa de treinamento normalmente é realizada antes de se tentar reconhecer qualquer amostra de voz. Essa etapa abrange a aquisição do sinal vocal dos locutores que serão cadastrados no sistema, extração de informações úteis dessas amostras, geração dos padrões de voz que serão utilizados como referência na etapa de reconhecimento e identificação dos limiares de similaridade associados aos padrões gerados, para o caso de verificação de locutor. A etapa de reconhecimento ou autenticação abrange a aquisição do sinal de voz que será avaliado, extração de informações úteis e comparação dessas informações com os padrões gerados na etapa anterior. Nesse momento, o limiar de similaridade indicará se a identidade foi ou não aceita. A figura 2 ilustra as etapas de treinamento e reconhecimento para um sistema genérico de RAL.

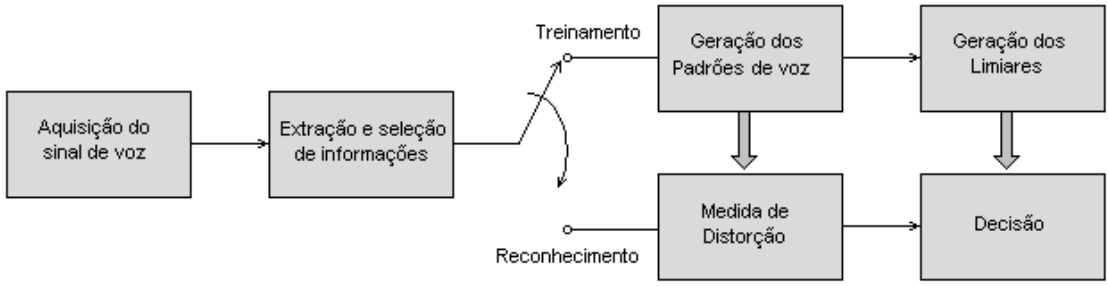

Figura 2: Ilustração das etapas de treinamento e reconhecimento para um sistema genérico de RAL

Neste trabalho, o processamento dos sinais de voz foi feito de acordo com resultados obtidos a partir de diversos experimentos práticos, que também são similares a vários outros trabalhos dos autores na área [Petry 2002] [Petry e Barone 2003] [Petry e Barone 2002] [Petry, Zanuz e Barone 1999]. Os algoritmos foram implementados utilizando a linguagem de programação C++, objetivando principalmente a redução no tempo de processamento requerido. Eles foram disponibilizados ao programa principal (escrito em linguagem Java) através de uma biblioteca de vínculo dinâmico sendo acessada via *Java Native Interface* (JNI) .

No sistema proposto, as amostras de voz utilizadas são gravadas a uma freqüência de amostragem de 8000 Hz, com resolução de 16 bits por amostra em apenas um canal (mono). A seguir, um filtro digital de pré-ênfase com resposta em freqüência

de −6 dB/oitava é aplicado ao sinal de forma a atenuar as componentes em baixa freqüência, nivelando o espectro do sinal. Posteriormente o sinal é dividido em janelas de curta duração, dentro das quais o sinal de voz pode ser considerado estacionário. As janelas do tipo hamming foram utilizadas com duração de 45 ms. Essas janelas são sobrepostas, sendo aplicadas a cada 10 ms do sinal de voz, de forma a suavizar a extração das informações. As janelas de voz que contiverem apenas silêncio são descartadas, baseado na análise da energia do sinal contido na janela. A partir de cada janela de voz são extraídos 16 coeficientes mel-cepstrais e 16 coeficientes delta mel-cepstrais, que serão utilizados para a caracterização do locutor. Os procedimentos de aquisição da voz, pré-ênfase e extração de informações úteis são detalhados em diversos trabalhos, como em [Rabiner e Juang 1993] [Deller Jr., Proakis e Hansen 1987]. O classificador utilizado é baseado em modelos de mistura de gaussianas (GMM), tendo sido empregadas 80 gaussianas adaptadas a partir de um modelo universal (UBM) para representação de cada padrão de voz. Maiores detalhes relativamente à implementação das técnicas empregadas na classificação podem ser obtidos em [Petry e Barone 2003] [Reynolds, Quatieri e Dunn 2000].

Ao se trabalhar com uma arquitetura cliente-servidor, é importante definir-se onde será realizado o processamento da voz adquirida no computador cliente. É possível transmitir-se todo o sinal de voz adquirido para o computador servidor. Entretanto, essa alternativa imporia um alto tráfego na rede, uma vez que os dados brutos de voz ocupam muitos bytes. Por exemplo, um sistema de cadastramento de um usuário que requerer 12 repetições de uma senha vocal composta de 5 dígitos, como sugerido em [Chirillo e Blaul 2003] em um produto comercialmente disponível, pode facilmente obter um minuto de fala. Se os dados brutos forem gravados a uma taxa de amostragem de 8000 Hz, 16 bits por amostra em apenas um canal, o espaço ocupado pelo sinal completo chegaria próximo a 1 MB. Além disso, o processamento de todo esse sinal (pré-ênfase, janelamento, extração de informações e geração do padrão de voz) seria realizado no computador servidor, elevando a carga de processamento requerida. Por outro lado, se a geração completa do padrão de voz fosse feita no computador cliente, que então enviaria ao servidor apenas o padrão gerado, poderiam haver sérios problemas relativos à segurança, como o envio de um padrão não confiável. Isso porque o computador cliente disporia dos algoritmos utilizados e dos padrões gerados, que poderiam ser utilizados para testes de padrões de autenticação, visando a quebra da segurança imposta pelo sistema. A situação seria pior se a fase de autenticação também fosse feita no computador cliente. Uma autenticação negativa poderia ser facilmente desprezada através de poucas mudanças no código do sistema. Assim, é importante que o computador servidor não sirva apenas como um repositório de padrões de voz e receptor de uma decisão sobre autenticação tomada em outro lugar, mas que o padrão de voz em si seja gerado assim como a comparação de padrões na autenticação sejam feitos lá.

No sistema proposto, as etapas de aquisição da voz, pré-ênfase, janelamento e extração de informações são realizadas no computador cliente, tanto no treinamento quanto no reconhecimento. Apenas as informações úteis extraídas são enviadas ao computador servidor, que gera o padrão de voz na fase de treinamento ou autentica o usuário, liberando o acesso a algum serviço restrito, na fase de reconhecimento.

3.2 Base de Dados

Visando uma ágil e robusta forma para o armazenamento e consulta das informações utilizadas no sistema proposto, foi construído um banco de dados para armazenar os padrões de voz e informações do usuário, assim como um conjunto de dados relativos à utilização do sistema, tais como: endereço IP do computador cliente, data de conexão, ações tomadas pelo cliente durante o decorrer da conexão. Foram utilizadas as tecnologias de acesso à base de dados *Java Database Connectivity* (JDBC) e servidor de banco de dados MySQL [Horstmann e Cornell 2001]. Deve-se salientar que o sistema proposto não está atrelado a nenhum banco de dados específico, visto que é disponibilizada na aplicação servidora a opção de conexão com os servidores de banco de dados mais utilizados atualmente no mercado.

Informações como nome de usuário, padrões de voz, informações sobre aquisição da voz e extração de informações úteis, e a data em que o último padrão de voz foi atualizado são guardados na tabela *Users*. A Tabela *Historic* armazena informações a respeito de conexões realizadas por cada um dos usuários da aplicação. São armazenadas as informações que identificam o usuário, o IP do computador cliente, a data de conexão e as informações úteis extraídas a partir da voz do usuário. Esses dados podem ser usados para uma possível avaliação futura de desempenho do sistema. O diagrama entidade-relacionamento (E/R) do banco de dados está ilustrado na figura 3.

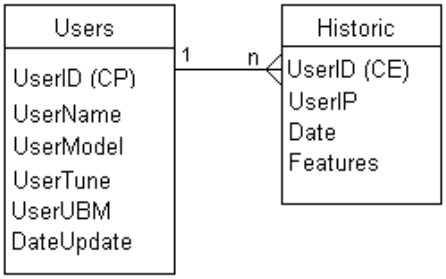

Figura 3: Diagrama E/R do banco de dados do sistema proposto

3.3 Comunicação

A possibilidade de dispor de uma comunicação segura é de importância fundamental para qualquer sistema de autenticação. A partir do estabelecimento da comunicação segura, todo o tráfego de informações entre o computador cliente e servidor deve ser realizado utilizando técnicas criptográficas para garantir o sigilo e a integridade dos dados.

Neste projeto, optou-se por utilizar um algoritmo de chave única para cifrar e decifrar os dados, o *Data Encryption Standard* (DES) [Tanenbaum 1997] [Schneier 1996] [Stallings 1999]. Um dos fatores que contribuiu para a escolha desse algoritmo refere-se à disponibilidade do cifrador junto ao kit de desenvolvimento utilizado (Java 2 SDK *standard edition* versão 1.4.1). Outro fator importante diz respeito à rapidez de execução do DES na cifragem e decifragem, quando comparado com algoritmos de chave assimétrica. O desafio para a perfeita utilização de algoritmos de chave única foca na troca segura da chave criptográfica entre os computadores cliente e servidor. Uma vez que inicialmente a comunicação segura ainda não existe enquanto a chave não é

conhecida pelo lado cliente e pelo lado servidor, deve-se buscar meios para que ambos lados conheçam a chave que será utilizada. O envio “em aberto” de uma chave pode ser interceptado, invalidando assim toda a segurança da comunicação. Para resolver o problema de definição de uma mesma chave criptográfica no computador cliente e servidor, sem ter que envia-la de forma não segura, foi implementado o protocolo mostrado na figura 4. Nesse protocolo, foram utilizadas classes que tratam especificamente da parte de criptografia - classes CriptoClient e CriptoServer - e classes utilizadas no gerenciamento da aplicação e comunicação via *Transmission Control Protocol* (TCP) - classes Cliente e Servidor.

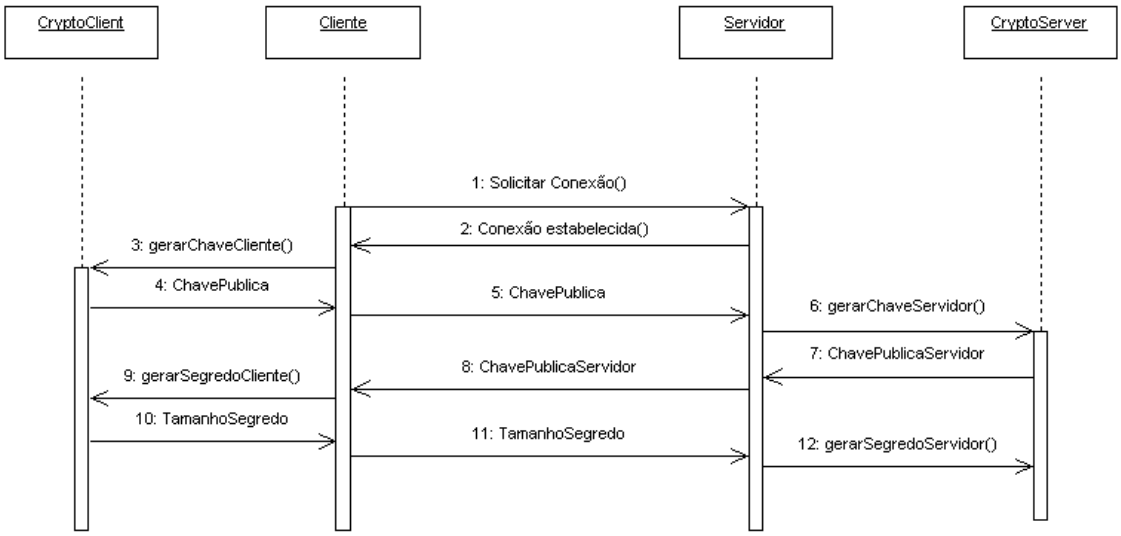

Figura 4: Protocolo para estabelecimento de chave criptográfica

Inicialmente, o computador cliente solicita uma conexão ao computador servidor, que estabelece a conexão. A partir de então, o computador cliente gera um par de chaves assimétricas, utilizando o algoritmo Diffie-Hellman com extensão de 1024 bits [Schneier 1996], enviando para o servidor apenas a chave pública. Da mesma forma, o computador servidor gera um par de chaves assimétricas e envia ao cliente sua chave pública. Nesse momento, tanto o lado cliente quanto o lado servidor possuem seu par de chaves assimétricas, e a chave pública do outro lado. O lado cliente gera então uma informação (segredo) utilizando sua própria chave privada (de conhecimento exclusivo e nunca transmitida) e a chave pública do computador servidor. A seguir o cliente envia ao servidor o tamanho do segredo gerado. O mesmo segredo gerado no cliente pode ser agora gerado no lado servidor. Isso é feito utilizando a chave pública do computador cliente, a chave privada do servidor (de conhecimento exclusivo e nunca transmitida) e sabendo-se o tamanho do segredo que deve ser gerado. Dessa forma, ambos lados cliente e servidor compartilham o mesmo segredo (utilizado como chave de extensão de 56 bits para o algoritmo DES), que foi gerado utilizando dados nunca transmitidos (chave privada).

3.4 Aplicação Cliente e Servidor

As técnicas apresentadas devem ser utilizadas de acordo com a evolução da interação entre o computador cliente e servidor. Assim, foram desenvolvidas as chamadas aplicação cliente e aplicação servidora. Elas são responsáveis basicamente por

implementar um protocolo de comunicação previamente definido para que se possa efetivamente realizar um cadastramento ou autenticação por voz. Além disso, devem ser capazes de utilizar os outros módulos disponíveis para acessar a base de dados, processar sinais de voz, comunicar-se com o outro lado de forma segura e interagir com o usuário a fim de adquirir amostras de sua voz.

A aplicação cliente e a aplicação servidora também são responsáveis pela interação com o usuário através de interface gráfica. Na janela construída para o lado cliente, mostrada na figura 5, é possível: configurar parâmetros como nome do usuário que utiliza o sistema, definir se será realizado um cadastramento ou autenticação, inserir dados sobre o computador servidor e ajustar controles de áudio (em desenvolvimento).

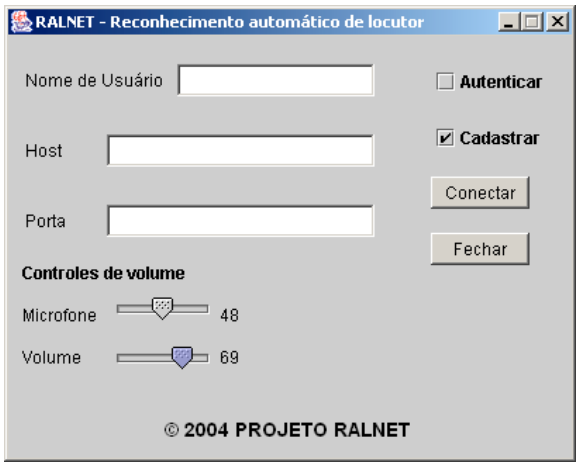

Figura 5: Interface gráfica para o lado Cliente

4. Resultados Experimentais Preliminares

Após a construção do protótipo foram conduzidos alguns testes iniciais para averiguação do perfeito funcionamento e possível correção de problemas. Os experimentos realizados foram executados inicialmente apenas entre os pesquisadores envolvidos na construção do sistema. Planeja-se para os próximos meses a execução de experimentos mas abrangentes, mas até o momento não foi desenvolvida uma metodologia de testes mais representativa que envolva maior número de usuários, a fim de identificar os melhoramentos que poderão ser executados no sistema.

Um dos fatores críticos que deverá ser revisto refere-se a parte de aquisição do sinal de voz no computador cliente. Ficou clara a dificuldade que poderá surgir se não houver uma prévia calibração dos recursos de áudio no computador cliente. Coisas como ajuste automático do volume de gravação do microfone, desabilitação de qualquer dispositivo de reprodução durante a gravação da voz, e uma clara orientação ao usuário para o momento exato que deverá iniciar sua fala, poderão facilmente resultar na má utilização do sistema.

5. Conclusões e Trabalhos Futuros

Este trabalho apresentou um protótipo desenvolvido para autenticação de usuários por voz em redes de computadores. O funcionamento geral do sistema construído foi descrito, assim como foram detalhados seus principais blocos constituintes. Diversos

desafios foram encontrados ao longo do desenvolvimento deste protótipo, relacionados a problemas específicos de construção de alguma parte ou integração e teste do sistema completo.

A partir dos testes de funcionamento serão identificados problemas e vulnerabilidades do sistema, no que diz respeito a perfeita autenticação biométrica dos usuários. Além disso, o protótipo construído deve servir apenas como intermediador para acesso a algum serviço remoto, como por exemplo comércio eletrônico, *internet banking* ou acesso a e-mails. Assim, deve haver uma preocupação também sobre formas para disponibilizar tal serviço com segurança.

Outra questão que deve ser levada em consideração refere-se às brechas de segurança que esse tipo de sistema pode apresentar. Pode-se imaginar que a pré-gravação da voz de uma pessoa e sua reprodução para o sistema de RAL possa acarretar a aceitação incorreta de impostores. Isso é especialmente verdade com a utilização de recursos de gravação e reprodução sofisticados. Os sistemas para RAL pouco podem fazer contra esse tipo de erro. Os métodos mais eficientes que se pode conceber seriam uma vigilância permanente do dispositivo de aquisição de voz, ou o sistema requerer uma palavra pseudo-aleatória específica. Uma vigilância permanente, a fim de garantir que aparelhos de reprodução sonora não sejam utilizados, normalmente não é algo desejado em um sistema. Para o aspecto de o sistema requerer uma palavra pseudo-aleatória, palavras diferentes da requerida apresentariam um valor superior de distorção, impedindo uma falsa aceitação. O fato de a palavra ser pseudo-aleatória dificulta o sucesso de uma pré-gravação [Petry 2002].

Referências

Chirillo, J. e Blaul, S. “Implementing Biometric Security”, Wiley Publishing Inc., 2003.

Woodward Jr., J. D., Orlans, N. M. e Higgins, P. T. “Biometrics: Identity Assurance in the Information Age”, McGraw-Hill, 2003.

Petry, A. “Reconhecimento Automático de Locutor Utilizando Medidas de Invariantes Dinâmicas Não-Lineares”, Tese de Doutorado, Instituto de Informática, Universidade Federal do Rio Grande do Sul, 2002.

Horstmann, C. S. e Cornell, G. “Core Java 2 – Recursos Avançados”, Makron Books, 2001.

Petry, A. e Barone, D. A. C. (2003) “Preliminary Experiments in Speaker Verification using Time-dependent Largest Lyapunov Exponents”, Computer Speech and Language, v. 17, p. 403-413.

Petry, A. e Barone, D. A. C. (2002) “Speaker Identification Using Nonlinear Dynamical Features”, Chaos Solitons and Fractals, v. 13, p. 221-231.

Petry, A., Zanuz, A. e Barone, D. A. C. (1999) “Utilização de Técnicas de Processamento Digital de Sinais para a Identificação Automática de Pessoas pela Voz”, In: Simpósio sobre Segurança em Informática, São José dos Campos, SP.

Rabiner, L. e Juang B. “Fundamentals of Speech Recognition”, Prentice-Hall Inc., 1993.

Deller Jr., J. R., Proakis J. G. e Hansen, J. H. L. “Discrete-time Processing of Speech Signals”, Prentice-Hall Inc., 1987.

Reynolds, D. A., Quatieri, T. F. e Dunn, R. B. (2000) “Speaker Verification Using Adapted Gaussian Mixture Models”, Digital Signal Processing, v. 10, p. 19-41.

Tanenbaum, A. S. “Redes de Computadores”, Editora Campus Ltda, 1997.

Schneier, B. “Applied Cryptography: Protocols, Algorithms, and Source Code in C”, John Wiley & Sons Inc., 1996.

Stallings, W. “Cryptography and Network Security: Principles and Pratice”, Prentice-Hall Inc., 1999.