Tiago de Freitas Pereira

Estudo comparativo de técnicas de detecção de ataques diretos à sistemas de autenticação de face

Campinas
2012

Universidade Estadual de Campinas
Faculdade de Engenharia Elétrica e de Computação

Tiago de Freitas Pereira

Estudo comparativo de técnicas de detecção de ataques diretos à sistemas de autenticação de face

Qualificação de Mestrado apresentada na Faculdade de Engenharia Elétrica e de Computação como parte dos requisitos exigidos para a obtenção do título de Mestre em Engenharia Elétrica. Área de concentração: Computação

Orientador: Professor Doutor José Mario De Martino

Este exemplar corresponde a versão final do exame de qualificação apresentado pelo aluno, e orientado pelo Prof. Dr. José Mario De Martino

Campinas
2012

Resumo

Resumo

Palavras-chave:

Abstract

Abstract

Key-words:

Sumário

Capítulo 1

Introdução

1.1 Contextualização e motivação

Em uma sociedade moderna, o processo de autenticação é uma tarefa importante para proteger dados e recursos sejam ele físicos ou digitais. Consistindo da confirmação de uma identidade requerida, o processo de autenticação é o primeiro e o mais crítico na cadeia de segurança restringindo acesso a usuários não autorizados.

Para a tarefa de confirmação de uma identidade, utilizam-se elementos que devem corresponder unívocamente ao identificador associado a um determinado usuário. Estes elementos são chamados de fatores de autenticação. Centralizados no usuário que está requerendo a identidade, estes fatores podem ser utilizados isoladamente ou combinados a fim de reforçar a a segurança. Os fatores de autenticação são classificados em aquilo que o usuário:

- **Sabe**: Por exemplo, uma senha ou uma frase de segurança;
- **Possui**: Por exemplo, um *token* de segurança, uma chave de cadeado ou um cartão;
- **É**: Por exemplo, uma característica física ou comportamental.

Cada um destes fatores apresentados possui um conjunto de vantagens e desvantagens. O uso mais comum de senhas, é acesso lógico a sistemas computacionais (computadores, e-mail, banco, cartão de crédito e muitos outros). A senha possui a vantagem de ser naturalmente imutável ao longo do tempo, ou seja, caso a mesma não seja mudada, ela continuará tendo o mesmo valor ao longo do tempo. Senhas contudo podem ser tão complexas quanto se queira, ficando a critério de seu detentor criar uma senha que ao mesmo tempo seja segura (de difícil adivinhação para um eventual atacante) e de fácil memorização. Estes critérios, claramente antagônicos, é o principal ponto de ataque a sistemas computacionais baseados em autenticação com senhas. Como um exemplo de vulnerabilidade no uso de senhas, em fevereiro de 2012 78 contas de e-mail de membros do governo da Síria foram invadidas divulgando informações confidenciais[1]. Destas 78 contas de e-mail, 33 a senha era '12345' ou '123456' incluindo a senha do próprio presidente.

[1] http://www.dailymail.co.uk/news/article-2100111/New-York-spin-doctor-coached-Syrian-dictator-Assad-swing-sympathies-US-public.html

Tokens geralmente são assoaciados a um segundo fator de autenticação. Como exemplo, um cartão de crédito é um *token* que acompanhado com a senha do cartão reforça a segurança do mesmo. Há bancos que disponibilizam para seus clientes tokens que geram um número aleatório a cada 30 segundos com a finalidade de reforçar o acesso à serviços de *internet banking*. Na Figura 1.1) pode-se o observar um token em formato de chaveiro. Estes *tokens* são objetos físicos que podem ser facilmente perdidos, e uma vez perdidos a autenticação fica comprometida por parte do usuário.

Figura 1.1: Token em formato de chaveiro

Biometria é a ciência de reconhecer a identidade de uma pessoa baseada em seus atributos físicos e/ou comportamentais, tais como a face, as impressões digitais, veias da mão, voz e a íris (**?**). O uso da biometria como fator de autenticação possuí algumas vantagens. Naturalmente, não é possível esquecer uma característica biométrica e dificilmente esta característica desaparece repentinamente (talvez em casos de acidentes graves). Características biométricas são intrínsecas a pessoa que as possui e portanto é intransferível. Como desvantagem, a biometria pode variar drasticamente ao longo do tempo. Como exemplo, a voz humana pode variar repentinamente quando estamos doentes; nossos traços faciais infelizmente envelhecem ao longo do tempo. Vale ressaltar que métodos de autenticação baseados em biometría são probabilísticos, ou seja, pode ser que o sistema de autenticação rejeite uma entrada autêntica devido à uma série de fatores externos.

As características humanas para serem utilizadas em uma método de autenticação biométrica devem satisfazer alguns requisitos, dentre eles destacam-se:

- Universalidade (toda pessoa deve possuí-la);
- Unicidade (deve permitir distinguir as pessoas);
- Estabilidade (não deve se alterar demasiadamente ao longo do tempo);
- Coletabilidade (deve poder ser medida quantitativamente);
- Desempenho (deve possibilitar um reconhecimento preciso, em tempo hábil);

- Aceitabilidade (deve ser aceitos facilmente por seus usuários);
- Circunvenção (deve dificultar a possibilidade de fraudes).

A tabela 1.1 apresenta um comparativo realizado por (**?**) entre as características biométricas mais utilizadas. É possível observar que nenhuma das biometrías apresentadas consegue atender todos estes requisitos com excelência e a escolha de qual utilizar deve levar em conta a natureza e as exigências de cada aplicação (**?**).

Tabela 1.1: Comparação das características biométricas mais utilizadas

Característica	**Universalidade**	**Unicidade**	**Estabiliade**	**Coletabilidade**	**Desempenho**	**Aceitabilidade**	**Circunvenção**
Face	Alta	Baixa	Média	Alta	Baixa	Alta	Baixa
Impressão Digital	Média	Alta	Alta	Média	Alta	Média	Média
Geometria das mãos	Média	Média	Média	Alta	Média	Média	Média
Veias da mão/dedo	Média	Média	Média	Média	Média	Média	Alta
Íris	Alta	Alta	Alta	Média	Alta	Baixa	Alta
Assinatura	Baixa	Baixa	Baixa	Alta	Baixa	Alta	Baixa
Voz	Média	Baixa	Baixa	Média	Baixa	Alta	Baixa

Sistemas de autenticação biométrica podem ser grosseiramente representados segundo o fluxograma da Figura **??**.

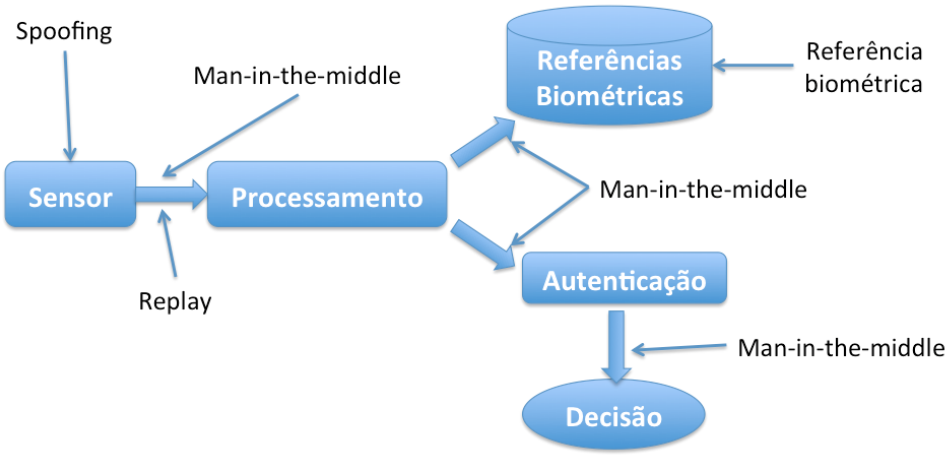

Figura 1.2: Fluxograma de um sistema de autenticação biométrica

Primeiramente o trato biométrico é capturado via algum tipo de **sensor**. Após esta captura, o trato biométrico capturado é **processado** a fim de extrair as características biométricas e geração da referência biométrica. Quando se está efetuando o cadastro de uma referência biométrica, estas características são **armazenadas** em uma base de dados para acessos futuros. Quando se está efetuando a **autenticação**, estas características biométricas serão utilizadas no processo de comparação com alguma identidade requerida no banco de dados. Conforme pode ser observado na figura ataques podem ser efetuados em qualquer ponto da arquitetura (**?**). As próximas subseções irão discorrer sobre cada um dos possíveis ataques e soluções para mitigá-los.

1.1.1 Ataque de *replay*

O ataque de replay consiste da utilização de dados previamente submetidos da identidade alvo para o sistema de autenticação a fim de obter o acesso não autorizado. Estes dados podem ser obtidos interceptando (*sniffing*) o canal de dados entre o sensor e a unidade de processamento de dados biométricos durante uma autenticação bem sucedida da identidade alvo. Para deter ataques dessa natureza, o sistema de autenticação biométrica deve assegurar que o dado fornecido não foi injetado artificialmente. Uma forma de se defender deste tipo de ataque é fazer uso da característica probabilística da própria biometria. É praticamente impossível dois processos de captura independentes e em intervalos de tempo distintos gerar exatamente o mesmo dado biométrico. Se isto ocorrer é provável que este dado foi interceptado e está sendo injetado no sistema de autenticação (**?**).

1.1.2 Ataque na referência biométrica

O ataque de referências biométricas consiste em atacar o local de armazenamento das referências biométricas. Com este tipo de ataque pode-se: adicionar uma biometria falsa no sistema de armazenamento, copiar as referências biométricas armazenadas, remover alguma referência biométrica ou modificar as referências biométricas existentes (**?**). Dentre estas possibilidades a mais perigosa é a cópia de referências biométricas, pois as mesmas podem ser usadas, através de engenharia reversa, para gerar biometrias falsas. (**?**) demonstrou que é possível gerar impressões digitais falsas através do processo de engenharia reversa com referências biométricas baseadas em minúcias. Com estas impressões digitais fabricadas, foi possível violar um sistema de autenticação baseado em impressões digitais. Um outro ponto de destaque neste tipo de ataque é que uma vez violada uma referência biométrica, a mesma perde a característica da unicidade.

1.1.3 Ataque *man-in-the-midle*

O ataque *man-in-the-middle* ou homem do meio é uma forma de ataque em que os dados trocados entre os componentes do sistema de biometria são, de alguma forma, interceptados e alterados pelo atacante. Neste tipo de ataque o atacante pode: interceptar dados sensor, interceptar dados enviados para armazenamento e interceptar dados de decisão do sistema de

biometria. Mecanismos como encriptação dos dados antes de serem transmitidos e/ou prover canais de comunicação seguro podem mitigar este tipo de ataque.

1.1.4 Ataque de spoofing

O ataque de *spoofing* em sistemas de autenticação biométrica, é um tipo de ataque em que o atacante forja o trato biométrico alvo apresentando uma biometria falsa ao sensor, burlando o sistema de autenticação. Em sistemas de autenticação baseados em biometria há duas motivações para se forjar um trato biométrico. A **primeira** motivação é o atacante obter o trato biométrico de outra pessoa a fim de tomar sua identidade. Em sistemas de autenticação baseados na voz, o atacante pode gravar a voz da identidade alvo e usar esta gravação como entrada no sistema de autenticação. Em sistemas de autenticação baseados em impressões digitais o atacante pode obter alguma impressão latente da identidade alvo e gerar um dedo artificial contento a impressão digital roubada. A saber em (**?**), (**?**) e (**?**) são trabalhos relacionados a ataques de *spoofing* em sistemas de autenticação baseados em impressões digitais, em (**?**), (**?**) e (**?**) são trabalhos relacionados a ataques de *spoofing* baseados na biometria da iris e em (**?**) e (**?**) são trabalhos relacionados a ataques de *spoofing* em sistemas de autenticação baseados na biometria da voz humana. A **segunda** motivação é um atacante gerar um trato biométrico totalmente artificial (sem se basear em uma biometria real), a fim de enganar o sistema de cadastro e autenticação biométricos. Com isso, o atacante pode compartilhar esta biometria falsa com outros atacantes.

Melhores práticas de segurança orientam utilizar mecanismos como, criptografia de dados e criação de canais seguros para mitigar ataques para a maioria dos casos citados anteriormente (**?**). No caso dos ataques de *spoofing*, o sensor de biometría (ponto alvo deste tipo de ataque) é o único ponto do fluxograma da Figura 1.2 em que nenhum dos mecanismos são efetivos para mitigá-los, tornando-se assim o ponto mais frágil a ataques e este será o ponto central desta dissertação.

1.2 Objetivos, Hipóteses e Resultados esperados

Este projeto verifica duas hipóteses:

1.3 Organização do trabalho

Capítulo 2

Revisão da literatura

2.1 Reconhecimento facial

Por sua natureza não intrusiva, autenticação utilizando a biometria da face é uma das áreas mais ativas e desafiadoras no campo da biometria. Apesar do significativo progresso da tecnologia de reconhecimento facial nas últimas décadas em uma série de tópicos como o envelhecimento dos indivíduos e iluminação exterior complexo ainda são desafios de pesquisa na área. Avanços na área foram amplamente relatados em (**?**, **?**). No entanto, a tarefa de verificar se o rosto apresentado a uma câmera é realmente um rosto de uma pessoa real, e não uma tentativa de forjar uma identidade (*spoof*) tem sido quase sempre esquecido.

Recentemente, a mídia tem documentado algumas situações de ataques em sistemas de reconhecimento de face em produção. Usando fotografias simples, um grupo de pesquisa da Universidade de Hanói mostrou que com relativa facilidade é possível burlar os sistemas de autenticação face em produção nos laptops Lenovo, Asus e Toshiba (**?**). Desde o lançamento da versão *Ice Cream Sandwich*, o sistema operacional Android vêm com um sistema de autenticação de face embarcado com a finalidade de desbloquear o celular. Desde então, tem sido amplamente demonstrado em toda a web como é possível burlar esta barreira de autenticação. Como resposta, uma contramedida baseada no piscar de olhos, foi introduzida na versão mais recente do sistema operacional Android.

Deficiências contra ataques de spoofing não é exclusividade da biometría facial. Podem ser observados em (**?**, **?**, **?**) que sistemas biométricos baseados em impressões digitais sofrem de fraqueza semelhante. A mesma deficiência em sistemas de reconhecimento de íris foram diagnosticados em (**?**, **?**). Finalmente, em (**?**, **?**) ataques spoofing para reconhecimento de locutor são abordados.

Um sistema de autenticação baseado na biometria de face pode ser forjada de diversas maneiras (**?**) e são elas a apresentação para a câmera de:

- Fotos com a face do usuário alvo;
- Vídeos com a face do usuário alvo;
- Máscaras construídas a partir da face do usuário alvo;

- Maquiagem na tentativa de imitar a identidade do usuário alvo;
- Cirurgia plática na tentativa de imitar a identidade do usuário alvo.

Embora seja possível para falsificar um sistema de autenticação de face usando maquiagem, cirurgia plástica ou máscaras; fotografias e vídeos são provavelmente as ameaças mais comuns. Além disso, devido à crescente popularidade das redes sociais na WEB, (facebook, youtube, flickr, instagram e outros) uma grande quantidade de conteúdo multimídia, especialmente vídeos e fotografias, estão disponíveis e estes dados podem ser utilizados facilmente para atacar um sistema de autenticação de faces. Para mitigar os sucessos dos ataques dessa natureza, contramedidas eficazes deve ser pesquisadas e desenvolvidas.

Contramedidas para ataques de spoofing em reconhecimento de face podem ser classificados quanto à dependência da colaboração do usuário. Métodos que são ditos **colaborativos**, partem do princípio que a pessoa que está efetuando a autenticação deve favorecer o mesmo, executando alguma atividade do tipo desafio-resposta. Em (**?**) e (**?**) o usuário é orientado a falar um texto gerado automaticamente e os movimentos labiais são correlacionados com reconhecimento de fala a fim de gerar uma checagem forte acerca da presença de um usuário em frente à câmera.

Métodos que não são colaborativos, procuram trabalhar somente com imagens ou vídeos capturados por câmeras convencionais que operam no espectro de luz visível. Estes métodos possuem a vantagem de não necessitar de nenhum hardware específico para este fim, tornando-os naturalmente mais baratos. Outra vantagem é abrangência de tais métodos, já que câmeras que operam no espectro visível são encontrados em muitos equipamentos de nosso uso diário como *webcams* comuns, *laptops*, *smartphones* e *tablets*. Tais métodos podem ser aplicados a sistemas de autenticação de face em diversos equipamentos. Uma outra vantagem que merece ser destacada é que a usabilidade de sistemas de autenticação de face não é onerada com estes métodos, já que o usuário não toma ciência de que uma checagem de sua presença em frente a câmera está sendo efetuada. Dada as vantagens descritas, o método escolhido para estudo será este (Horrível).

Estratégias não colaborativas podem ser classificados em que exploram:

- Presença de vitalidade (liveness detection);
- Discrepância em padrões de movimento;
- Discrepância relativa a qualidade da imagem;
- Combinação.

2.1.1 Presença de vitalidade

Presença de vitalidade ou *liveness detection* consiste na seleção de características faciais que apenas pessoas vivas conseguem reproduzir.

O piscar de olhos é uma tarefa involutária que os seres humanos executam constantemente. Um ser humano comum pisca de forma involuntária a cada 2 ou 6 segundos para manter os olhos limpos e umidecídos. Este intervá-lo pode variar drasticamente em situações de stress

aumentando este intervalo para mais de 20 segundos. Porém, não importa a situação de stress que se está submetido; em algum momento este movimento irá ocorrer. Não há estabelecido um limite máximo estabelecido em que um ser humano consegue suportar sem piscar os olhos.

Apoiado na hipótese citada anteriormente, (**?**, **?**) desenvolveram uma contramedida baseada no piscar dos olhos com o objetivo de bloquear ataques efetuados com fotografia. O sistema desenvolvido modela a piscadela utilizando cadeias escondidade de markov (HMM) mapeando os estados de olho aberto para olho fechado e olho aberto novamente. Experimentos foram conduzidos utilizando uma base de dados criada pelos autores e livremente disponível para download[1] mostraram uma acurácia de 95,7% contra uma taxa de falsos positivos abaixo de 0,1%.

2.1.2 Discrepância em padrões de movimento

Contramedidas baseadas em padrões de movimento partem do princípio que tentativas de ataque possuem padrões de movimento distintos de acessos reais.

Apoiado na hipótese que em um ataque regiões distintas da face apresenta padrões de movimento altamente correlacionados se comparados com acessos reais, (**?**) desenvolveu uma heurística baseada em fluxo ótico para mensurar tal correlação. Comparando a direção entre o centro da face e a região das orelhas como pode-se observar na Figura X,

Fig X

O algoritmo pode ser sumarizado como segue:

1. Detectar a face;
2. Definir região facial está se movendo mais horizontalmente ou mais verticalmente;
3. Delimitar a região do centro da face e das orelhas; verticalmente;
4. Se a cabeça estiver se movendo mais horizontalmente, LIMIARES. TERMINAR
5. Será considerado ataque se a quantidade de movimento percebida na região do centro da face semelhante a quantidade de movimento na região das orelhas.

A performance foi avaliada utilizando uma base de dados construída sobre a base de dados de face XM2VTS. Os acessos reais foram gerados utilizando o subconjunto *Head Rotation Shot* desta base de dados e o ataques foram gerados com fotografias impressas em papel das mesmas imagens utilizadas para os acessos reais e regravados utilizando uma câmera de computador. Com esta base de dados criada um $EER = 0,5\%$ foi obtido. Esta base dados não foi disponibilizada publicamente pelos autores de modo que qualquer tentativa de reprodução dos resultados fica impossibilitada.

Ainda com fluxo ótico, (**?**) explora as diferenças de movimento entre um objeto plano e um objeto tridimensional. Ao contrário de (**?**), Bao et al não detecta regiões específicas da face, mas sim as divide em quatro partes (horizontalmente e verticalmente) e avalia se ...

Com a seguinte euristica..

[1] http://www.cs.zju.edu.cn/ gpan/database/db_blink.html

A contramedida proposta por (**?**) mede a correlação do movimento da região da face em relação ao seu fundo utilizando como métrica de movimento uma simples diferença das intensidades dos pixels em quadros sucessivos. O movimento acumulado entre esta diferença (M_D), para um dado uma Região de Interesse (ROI) e seu respectivo fundo, é calculado usando a seguinte equação:

$$M_D = \frac{1}{S_D} \sum_{(x,y)\in D} |I_t(D) - I_{t-1}(D)| \tag{2.1}$$

em que D é o RoI, S_D é a área do RoI e I_t é a intensidade de um pixel.

Para introduzir o coeficiente de movimento em um classificador, 5 quantidades são extraídos para cada janela de 20 quadros. As quantidades são as seguintes: o mínimo de M_D em que a janela de tempo, no máximo, a média, o desvio padrão e a proporção R espectral entre a soma de todos os componentes não-DC e DC do componente em si tomadas como base o N pontos transformada de Fourier do sinal (ver Equação ref eq: motionR). Estes 5 quantidades são alimentados em um Multi-camada classificador Perceptron (MLP), com 5 neurônios na camada escondida, que é treinado para detectar ataques de spoofing.

2.1.3 Discrepância relativa a qualidade da imagem

Capítulo 3

Metodologia

Capítulo 4

Conclusões e Perspectivas

Perspectivas

Publicações

Submissões