MINISTÉRIO DA DEFESA
EXÉRCITO BRASILEIRO
DEPARTAMENTO DE CIÊNCIA E TECNOLOGIA
INSTITUTO MILITAR DE ENGENHARIA

Victor Villas Bôas Chaves
Lucas Sousa Meireles
Cláudio Cavalcante Bomfim

Identification by Keystroke Dynamics

Rio de Janeiro
Dezembro 2017

Instituto Militar de Engenharia

Victor Villas Bôas Chaves

Identification by Keystroke Dynamics

Relatório Final do Programa Institucional de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação do CNPq / Instituto Militar de Engenharia.
Orientador: Prof. Ronaldo Goldschmidt - D.C.

Rio de Janeiro
Dezembro 2017

INSTITUTO MILITAR DE ENGENHARIA
Praça General Tibúrcio, 80 - Praia Vermelha
Rio de Janeiro-RJ CEP 22290-270

/ Victor Villas Bôas Chaves; orientados por Prof. Ronaldo Goldschmidt- Rio de Janeiro: Instituto Militar de Engenharia, Dezembro 2017.

13p. : il.

- Instituto Militar de Engenharia - Rio de Janeiro, Dezembro 2017.

Resumo

Nosso resumão **Palavras-chaves**: dinâmica da digitação, reconhecimento de usuário

Abstract

Our big abstract **Keywords**: our keywords

Sumário

1 INTRODUÇÃO 7
1.1 Contexto e Motivação 7
1.2 Objetivos 7
1.3 Contribuições Esperadas 7
1.4 Método 7
1.5 Introdução Teórica 7
1.6 Cronograma 8
1.7 Viabilidade 8
1.8 Organização do Texto 8

2 FUNDAMENTAÇÃO 10
2.1 Trabalhos Relacionados 10

3 SOLUÇÃO PROPOSTA 11
3.1 Modelo conceitual 11
3.2 Protótipo 11

4 EXPERIMENTOS E RESULTADOS 12

5 CONCLUSÃO 13

1 Introdução

1.1 Contexto e Motivação

Sistemas de Segurança da Informação modernos não se baseiam e um único método de autenticação, mas incrementalmente adicionam mecanismos com múltiplos fatores. Quanto mais e melhores fatores, maior a certeza da identidade ser autenticada corretamente.

Dentre as alternativas mais promissoras estão os fatores biométricos, valorizados por sua natureza individual e difícil falsificação. Os fatores biométricos frequentemente mencionados são os fisiológicos, mas seu emprego traz diversos fatores complicantes como a necessidade de amostragem prévia e diminuição da usabilidade do sistema. Uma alternativa é o uso de fatores biométricos de comportamento, como padrões comportamentais expressos naturalmente pelo usuário.

As vantagens dos fatores comportamentais incluem a possibilidade de amostragem silenciosa, maior variabilidade do grau de confiança e a transparência do mecanismo para o usuário. Em particular, sistemas providos pela *Web* em geral possuem uma uniformidade de interface que permite a coleta de vários padrões comportamentais durante todo o uso do sistema.

Alguns exemplos de comportamentos de interesse coletáveis incluem padrões de digitação, cliques de *mouse* ou áreas do sistema e recursos acessadas pelo usuário. A pesquisa sobre como utilizar fatores dessa natureza pode impulsionar sistemas mais seguros e menos impactantes na experiência do usuário.

1.2 Objetivos

Neste trabalho serão investigados os processos necessários para se utilizar os padrões de digitação como fator de autenticação biométrica comportamental. Tais processos incluem a coleta de dados, extração de informação, algoritmos de decisão e arquiteturas de sistema que tornem possível a implantação deste método.

- Analisar os tipos de informação que se podem extrair a partir dos padrões de digitação de um indivíduo;
- Modelar a combinação das informaçoes extraídas utilizando algoritmos de aprendizado de máquina;
- Sistematizar um mecanismo de coleta de amostras que permita o treinamento dos modelos escolhidos;
- Definir uma arquitetura de sistema para implantação dos mecanismos de coleta e autenticação definidos;

1.3 Contribuições Esperadas

1.4 Método

1.5 Introdução Teórica

O problema que o ramo de aprendizado por maquinas se propõe a resolver se baseia na busca pela aproximação funcional matemática de um problema real, onde $F : C- > R$ é a função alvo desconhecida, então para resolver o problema é suposto uma função $G : H- > R$, onde G é a função estimada que pretende-se aproximar de F, $H \subset C$ é o conjunto das amostras que pretende-se expandir para o conjunto real expandido

C, os valores de H e R são conhecidos; porém, para um problema não interpolável, não há um mapeamento conhecido de $F : C- > R$, seja por falta parâmetros de difícil análise, ou pelo caráter indeterminado do problema, em ambos os casos o tratamento é similar atravéz de estimação e tratamento de erros na abordagem.

Para o problema clássico de classificação pode-se usar, no domínio discreto, a seguinte abordagem: seja $f : x- > y$ a função alvo onde $x = [x_1, x_2, ..., x_d]^t$ é o vetor das amostras de H, seja o somatório $\sum_{i=1}^{d} \omega_i * x_i$, onde ω_i é o peso do seu respectivo elemento do vetor x, de tal forma que:

$$\sum_{i=1}^{d} \omega_i * x_i > t; y = 1. \sum_{i=1}^{d} \omega_i * x_i < t; y = 0. \tag{1.1}$$

De tal forma os pesos tem suas características definidas por seus respectivos elementos de x da seguinte forma:
$|\omega_i|$ é alto quando x_i for importante.
$\omega > 0$ quando x_i for benéfico.
$\omega < 0$ quando x_i for maléfico.
Por fim teremos uma aproximação da função alvo no problema de classificação para $h(x) = sign((\sum_{i=1}^{d} \omega_i * x_i) + \omega_0)$, onde $\omega_0 = t$.
Vale notar que o vetor x apenas representa as amostras colhidas assim embora ele represente o conjunto C como uma aproximação de C para H existe uma divergência conhecida como erro dentro da amostra E_in essa diferença é representada pela diferença entre x e sua projeção em C, há também o erro referente aos valores extrapolados pela função $G : H- > R$ como esse erro não se refere à amostra colhida ele é conhecido como erro fora da amostra E_out.
Uma ferramenta de verificação de erros que é utilizada afim de reduzi-los é desigualdade de Hoeffding Chernoff como meio de verificação sobre os erros dentro e fora da amostra, a desigualdade em questão pode ser expressa da seguinte forma:
$P[|\upsilon - \mu| > \epsilon] \leq 2 * e^{-2*\epsilon^2*N}$, para $\epsilon > 0$
$P[|\upsilon - \mu| \leq \epsilon] > 2 * e^{-2*\epsilon^2*N}$, para $\epsilon > 0$
A fim de refinar os valores e diminuir o erro escolhe-se o valor de ϵ de tal forma que $\upsilon + \epsilon \geq \mu \geq \upsilon - \epsilon$, nota-se que a fronteira denotada por $2 * e^{-2*\epsilon^2*N}$ não depende de μ, nem do tamanho do domínio.
Para o problema de classificação a desigualdade pode ser aplicada tomando $|E_in - E_out| = |\upsilon - \mu|$ assim tem-se as equações $P[|E_in - E_out| > \epsilon] \leq 2 * |H| * e^{-2*\epsilon^2*N}$, ou $P[|E_in - E_out| \leq \epsilon] > 2 * |H| * e^{-2*\epsilon^2*N}$, para $\epsilon > 0$.

1.6 Cronograma

1.7 Viabilidade

O que eu escrevo aqui?

1.8 Organização do Texto

No capítulo 2 são introduzidos os conceitos necessários para a modelagem conceitual de um sistema de autenticação por dinâmica de digitação. Na seção 2.1 são discutidos trabalhos relacionados.

	Fev	Mar	Abr	Mai	Jun	Jul	Ago	Set	Out
Revisão bibliográfica									
Modelagem conceitual									
Prototipagem									
Coleta de dados									
Entrega de relatório parcial									
Experimentação									
Entrega do relatório final									
Apresentação									

Tabela 1 – Cronograma mensal de trabalho

No capítulo 3 é apresentado uma arquitetura de sistema de autenticação isolado, para fácil implantação do método apresentado. Na seção 3.1 é definido o modelo de autenticação, especificando o fluxo de informações desde a coleta até a decisão de um grau de confiança de identidade, enquanto em 3.2 é demonstrada uma possível implementação da solução proposta, servindo como prova de conceito para o modelo.

No capítulo 4 são analisados os resultados do experimento proposto com o protótipo criado, analisando o sucesso da solução.

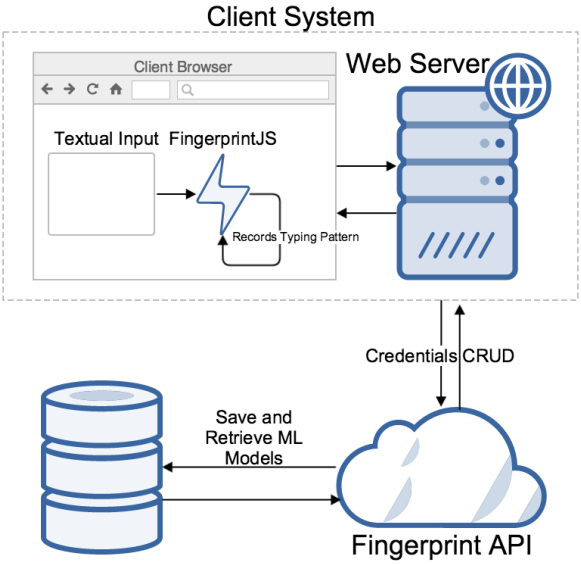
No capítulo 5 termina-se por sumarizar o conceito, a solução e os resultados obtidos pelo sistema apresentado.

2 Fundamentação

2.1 Trabalhos Relacionados

3 Solução Proposta

Aqui entra a arquitetura do sistema, ver com Meireles.



3.1 Modelo conceitual

Aqui entra a modelagem do processo de decisão.

3.2 Protótipo

Aqui entra o nosso POC, sendo feito no GitHub.

4 Experimentos e Resultados

5 Conclusão

Excelente trabalho, time!