GUSTAVO VALENTE NUNES

COLETA DE DADOS DE USO DE APLICAÇÕES ANDROID PARA ANÁLISE COMPORTAMENTAL DE USUÁRIOS

(*versão pré-defesa, compilada em 11 de dezembro de 2023*)

Trabalho apresentado como requisito parcial à conclusão do Curso de Bacharelado em Ciência da Computação, Setor de Ciências Exatas, da Universidade Federal do Paraná.

Área de concentração: *Ciência da Computação*.

Orientador: André Ricardo Abed Grégio.

CURITIBA PR

2023

RESUMO

Este trabalho apresenta uma análise abrangente dos sistemas de autenticação, abordando métodos tradicionais e biométricos, destacando os fatores de autenticação. Os fatores comuns – conhecimento, posse e identidade pessoal – são explorados, enfatizando a autenticação baseada em senhas, tokens físicos (software e hardware) e características físicas ou comportamentais únicas. Adicionalmente, são discutidos fatores de autenticação adicionais, como localização, horário e comportamento do usuário, demonstrando a busca pela autenticação contínua baseada nos padrões de interação e comportamento do usuário. A autenticação biométrica é examinada de forma detalhada, diferenciando características fisiológicas e comportamentais. A análise destaca as características fisiológicas, como impressões digitais e reconhecimento facial, devido à sua singularidade e precisão na identificação de usuários, embora alguns métodos, como reconhecimento de íris, enfrentem desafios de implementação devido à necessidade de equipamentos específicos. O capítulo sobre falhas de segurança destaca as vulnerabilidades em sistemas tradicionais e biométricos. No caso dos sistemas tradicionais, ataques de força bruta, enumeração de usuário, "shoulder-surfing"e phishing são mencionados como ameaças comuns. Já nos sistemas biométricos, ataques de Spoofing, como Artificial Synthesis e Replay Attack, como o Wolf Attack, são explicados, destacando a vulnerabilidade desses sistemas à imitação de características fisiológicas ou comportamentais. Além disso, apresenta informações sobre a autenticação contínua, enfatizando os dados usados para análises contínuas ao longo do tempo e a relevância da UEBA (User and Entity Behavior Analytics). Finaliza mencionando um framework de coleta de dados em dispositivos Android, projetado para pesquisas futuras em autenticação contínua. Esse framework é baseado na coleta de dados dos pacotes Android, capturando informações das janelas ativas e em segundo plano. Busca adquirir uma variedade de comportamentos dos usuários para analisar padrões de interação e comportamentos específicos. Esse conjunto diversificado de dados é fundamental para viabilizar a autenticação contínua, permitindo a avaliação precisa e constante dos hábitos dos usuários para futuras autenticações.

Palavras-chave: Palavra-chave 1. Palavra-chave 2. Palavra-chave 3.

ABSTRACT

This paper presents a comprehensive analysis of authentication systems, covering both traditional and biometric methods, highlighting authentication factors. Common factors - knowledge, possession, and personal identity - are explored, emphasizing password-based authentication, physical tokens (software and hardware), and unique physical or behavioral characteristics. Additionally, additional authentication factors such as location, time, and user behavior are discussed, showcasing the pursuit of continuous authentication based on interaction and user behavior patterns. Biometric authentication is scrutinized by distinguishing physiological and behavioral characteristics. The analysis emphasizes physiological features like fingerprints and facial recognition for their uniqueness and precision in user identification, though some methods, such as iris recognition, face implementation challenges due to specific equipment needs. The section on security flaws underscores vulnerabilities in both traditional and biometric systems. In traditional systems, brute force attacks, user enumeration, shoulder-surfing, and phishing are mentioned as common threats. In biometric systems, spoofing attacks like artificial synthesis and replay attacks, such as the Wolf Attack, are explained, highlighting the susceptibility of these systems to imitate physiological or behavioral traits. Furthermore, it provides insights into continuous authentication, emphasizing the data used for ongoing analyses and the relevance of UEBA (User and Entity Behavior Analytics) in bolstering corporate security. It concludes by mentioning a data collection framework in Android devices designed for future continuous authentication research. This framework is based on collecting Android package data, capturing information from active and background windows, aiming to acquire a range of user behaviors to analyze interaction patterns and specific behaviors. This diverse dataset is crucial for enabling continuous authentication, allowing precise and constant evaluation of user habits for future authentications.

Keywords: Keyword 1. Keyword 2. Keyword 3.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

1 INTRODUÇÃO

A autenticação é um elemento essencial da segurança digital, onde a validação da identidade de um usuário é fundamental para garantir o acesso seguro a sistemas e dados sensíveis. Dentro desse contexto, a autenticação básica é amplamente reconhecida como um método fundamental, envolvendo a associação de um nome de usuário (ou e-mail) e senha. Além disso, outras formas de autenticação, como o uso de PINs, perguntas de segurança e a evolução para autenticação de dois fatores (2FA), também desempenham um papel crucial na garantia da identidade do usuário. A diferenciação entre autenticação e autorização é crucial: enquanto a autenticação verifica a identidade do usuário, a autorização define as permissões de acesso a recursos protegidos.

O uso de biometria tem sido uma faceta emergente dos métodos de autenticação, que se baseia em características fisiológicas ou comportamentais únicas para identificar indivíduos. Ela abrange diversos fatores, como reconhecimento facial, de íris e de impressão digital, cada um com suas próprias vantagens e desafios. Apesar de sua importância, tanto os métodos tradicionais quanto os biométricos enfrentam desafios de segurança. Desde vulnerabilidades como senhas fracas, reutilização de credenciais e ataques de phishing até questões específicas de ataques aos sistemas biométricos, como spoofing attacks, há uma constante necessidade de defesas mais robustas.

É nesse cenário que a autenticação contínua se destaca, indo além da validação única da identidade do usuário e monitorando constantemente suas atividades. Por meio de técnicas como análise de comportamento do usuário, dados contextuais e o uso de aprendizado de máquina, a autenticação contínua busca verificar a identidade do usuário de maneira persistente e adaptativa. Compreender os desafios, vetores de ataque e estratégias de defesa é crucial para avançar o estado da arte em segurança considerando os métodos de autenticação. Neste projeto, propõe-se a implementação de um *framework* de coleta de dados em dispositivos Android, visando utilizar informações comportamentais dos usuários para futuras tentativas de autenticação baseadas em seus padrões de comportamento ao longo do tempo. Essa abordagem visa fomentar a avaliação da viabilidade do uso deste tipo de informação do usuário a fim de explorar novos horizontes na segurança da autenticação, levando em consideração os benefícios e desafios da autenticação contínua e das tecnologias biométricas (comportamentais) para criar sistemas mais robusto e adaptáveis em face de ataques sofisticados contra a identidade do usuário.

1.1 OBJETIVO

O objetivo do presente trabalho é desenvolver uma ferramenta que ajude na coleta de dados para que no futuro esses dados sejam utilizados para realizar trabalhos em cima de sistemas de autenticação contínua. Este trabalho faz parte de um novo projeto de pesquisa em cooperação

com a Texas A&M University, o qual passou por aprovação departamental e no Comitê de Ética em Pesquisa do Setor de Ciências Sociais, uma vez que se lida com dados comportamentais do uso de sistemas computacionais nos laboratórios do Departamento de Informática.

Para tanto, foi modelado e prototipado um *framework* para a plataforma de celulares Android que obtém informações sobre os aplicativos que o usuário está utilizando no momento, como o título da janela que está aberta, quais janelas estão em background e qual janela está ativa e o nome do pacote que está sendo utilizado, que seria o nome do pacote dado na hora que foi desenvolvido, um exemplo seria algo como *org.example.package*. O protótipo se utiliza do paradigma cliente servidor e anonimiza os usuários de forma a não ser possível identificá-los, mas ao mesmo tempo agregando dados de um mesmo usuário para permitir que se verifique se os comportamentos são únicos a ponto de mapear seu comportamento e autenticá-lo em sistemas homogêneos.

Dessa forma, o projeto foi divulgado para alunos das disciplinas de Programação 1 e Algoritmos e Estruturas de Dados 2, ministradas no segundo semestre de 2023 pelo orientador desta monografia, para fins de testes preliminares com aqueles que se voluntariaram.

1.2 ORGANIZAÇÃO DO TEXTO

Este trabalhos está dividido da seguinte forma: no Capítulo 2 está toda a fundamentação teórica do trabalho onde se fala sobre os métodos de autenticação, tanto o básico que é o conhecido email+senha como alguns métodos mais recentes como os biométricos; O Capítulo 3 se fala sobre métodos de autenticação contínua em dispositivos móveis e quais são os tipos de dados que são coletados para poder fazer essa análise; No capítulo 4 é apresentado a proposta do *framework* e explicação de seu funcionamento interno para a coleta de dados; E o capítulo 4.6 é a conclusão do trabalho e falamos sobre propostas e caminhos futuros nessa linha de pesquisa.

2 SISTEMAS DE AUTENTICAÇÃO

Um aspecto importante da segurança cibernética é a identidade. O processo de autenticação envolve provar que a pessoa (ou usuário de um sistema computacional) que está tentando acessar determinado recurso é ela mesmo, o que geralmente ocorre por meio da verificação uma senha, um *token* de hardware ou alguma outra informação que comprove sua identidade (15).

Da mesma forma que um funcionário de uma companhia aérea precisa confirmar o passaporte ou a carteira de identidade de uma pessoa ao embarcar em um avião, é necessário que os sistemas e aplicativos de computadores verifiquem a identidade do usuário a fim de se confirmar suas permissões de acesso e uso (15). Considerando o exemplo do sistema aéreo, a autenticação é válida depois que o passaporte é considerado válido e a pessoa tem acesso ao seu voo, enquanto que nos sistemas e aplicativos esse processo é validado depois de digitar a senha e ter acesso a sua página pessoal dentro desse sistema. O processo de autenticação não é necessariamente apenas para seres humanos, também é necessário para robôs, APIs e outros que precisem ter acesso a determinado recurso (Devoteam).

Enquanto a autenticação se concentra na verificação da identidade, a autorização lida com permissões para ações em sistemas ou recursos protegidos. A autenticação envolve confirmar a identidade do usuário, pelo meio de senhas, biometria ou tokens, com mais de um fator e até mesmo por localização e hora (12). A autorização determina o que um usuário autenticado está autorizado a fazer, limitando seu acesso com base no que se tem permissão para se fazer, um exemplo seria: Um usuário comum não possui autorização de acessar dados da parte administrativa do sistema.

Neste capítulo serão discutidos alguns dos métodos de autenticação que são utilizados e estudados por pesquisadores, desde os sistemas de autenticação tradicionais como sistemas de autenticação mais modernos como os biométricos. Além de levantar alguns conceitos sobre a diferença de fatores de autenticação, discussões sobre problemas nos sistemas de autenticação e um overview sobre as possibilidades de ataques e defesas em cima desses sistemas.

2.1 AUTENTICAÇÃO BÁSICA

Nesse formato de autenticação, envolve está referenciando a algum tipo de conhecimento básico que somente essa pessoa tem, está muito atrelado a combinação de nome de usuário (ou email) e senha (20). Outros exemplos desse método é a utilização de pins de segurança, que são amplamentes utilizados em celulares e nos computadores como um método mais fácil de autenticar e também temos as perguntas de segurança, que é utilizado em último caso, geralmente utilizado quando o usuário esqueceu sua senha e email de recuperação de senha (15). Esse é o método de autenticação mais conhecido pelas pessoas, utilizado na maioria das aplicações

diárias da sociedade, como nas redes sociais, qualquer aplicação Web, nos sistemas Desktops e nos Sistemas Operacionais.

2.1.1 Vantagens e Desvantagens

Como é uma forma de autenticação muito utilizada, todo mundo já entende o motivo de ser importante e a forma correta de se usar. O custo de se manter uma aplicação dessa é menor se comparado a outros formatos de autenticação, são compatíveis com uma ampla gama de sistemas e dispositivos e é versátil para ser utilizado na maioria das situações. Como já existe a bastante tempo, é um método testado e validado, já foram feitos vários testes de segurança em cima desse formato, além de já existirem várias tecnologias que facilitam a implementação, desde frameworks como ferramentas que te permitem fazer o controle de usuários de um sistema.

Embora seja um ótimo método de autenticação, existem problemas que o circulam. A pessoa pode acabar esquecendo a senha e não sabe trocar com algum email de recuperação, existe a possibilidade do usuário criar uma senha fraca de se usar e acabar tendo sua conta invadida por algum ataque de força bruta ou dicionário (20). Ao criar uma conta, em alguma aplicação é possível que essa aplicação colete seus dados para fins de pesquisa sem a pessoa saber, já que ela é um usuário autenticado, então é possível criar uma identidade para essa pessoa.

2.2 AUTENTICAÇÃO MULTIFATOR

A autenticação multifator (MFA) é a capacidade de verificar a identidade de uma pessoa, utilizando dois ou mais fatores de autenticação, em vez de apenas um. É um método de autenticação mais seguro do que a autenticação de fator único, porque é muito mais difícil fazer o bypass de mais de um fator de autenticação (13). Supondo que uma pessoa chamada Bob, seja alvo de um ataque e tenha sua senha e nome de usuário roubados, através de um ataque de phishing se ele só tivesse esse mecanismo de autenticação, esse tipo de roubo pode afetar muito a integridade do usuário.

Porém, vamos supor também, que Bob possui um segundo método de autenticação, através de um código que chega a partir de um aplicativo do seu celular, toda vez que precisa fazer login em uma aplicação. Se o indivíduo que roubou sua senha e seu nome de usuário tentar acessar, não vai conseguir, já que Bob possui outro fator de autenticação. Além dessa segurança a mais, cria um alerta para Bob, já que ele vai saber que alguém tentou acessar sua conta. Para ser considerado autenticação de vários fatores, fatores separados precisam ser verificados (27).

Um exemplo, que não é considerado multi-fator seria algo como um aplicativo pedindo para o usuário digitar uma senha e logo em seguida responda perguntas de segurança para autenticar, isso é considerado autenticação de fator único. Isso porque, a entrada da senha seguida pelas perguntas de segurança, entra no fator “algo que você sabe” (27).

2.2.1 Tipos de fatores

As características que um sistema de autenticação verificará são chamados de “fatores”. Existem três fatores que são comuns e são amplamente utilizados nos dias de hoje, eles são:

- **Algo que a pessoa sabe:** Esse fator de autenticação é um tipo de conhecimento que somente a pessoa sabe sobre, que é muito conhecido como a senha do usuário. É um fator clássico que é utilizado na maioria dos sistemas de autenticação, outros exemplos que é considerado parte desse fator, é um pin do usuário ou perguntas de segurança que somente o usuário sabe (27).

- **Algo que a pessoa tem:** Esse fator verifica se o usuário possui algum item físico que comprova sua identidade, em sistemas digitais esse fator de autenticação depende de um token físico e é possível classificar de duas formas, Tokens de Software e Tokens de Hardware (13). Os tokens de software são ligados a dispositivos, como celulares, assim é possível que o sistema de autenticação envie alguma mensagem por mensagem ou por algum aplicativo de celular, que possa validar a identidade do usuário. E existem os Tokens de Hardware, que é um item físico que se conecta a um computador ou dispositivo móvel via Bluetooth, USB, ou de alguma outra forma, sendo possível validar sua identidade ao conectar esse token ao dispositivo (27).

- **Algo que a pessoa é:** Esse fator de autenticação avalia os fatores físicos de uma pessoa. Muito utilizado pelos dispositivos móveis para autenticar o usuário utilizando impressões digitais dos dedos, biométrica fácil. Também existe a possibilidade da autenticação por voz e outros métodos mais avançados porém não convencionais. (27).

- **Fatores adicionais:** É possível a autenticação do usuário baseado na localização (Onde está) e horário (Quando está acessado) e além desses dois, também é muito pesquisado sobre autenticar o usuário baseado em comportamento, sobre como ele digita no computador, o que acessa, como se comporta ao utilizar um dispositivo (13).

2.2.2 Autenticação em dois fatores

A autenticação em dois fatores (2FA) é uma forma de autenticação que requer dois tipos distintos de fatores, para verificar a autenticidade da identidade do usuário. Isso significa que, antes do usuário ter acesso ao recurso que foi solicitado, ele precisa fornecer duas formas diferentes de comprovar sua identidade (28). Geralmente, 2FA envolve “algo que você sabe” e “algo que você possui”, esses dois também é conhecido como uma senha que o usuário sabe mais uma confirmação no celular (o mais popular) (11), com dois fatores torna mais difícil para um invasor conseguir acesso não permitido impossibilitando o comprometimento do usuário. Existem diferentes fatores de autenticação, desde o básico como senhas até outros utilizando dados biométricos (como impressões digitais) e localização (como GPS).

Assim como no MFA, para ser considerado autenticação em dois fatores, é necessário ter mais de um fator de autenticação. É muito comum a utilização de dois fatores e é super importante devido aos problemas atuais de segurança, utilizando o 2FA aumenta muito a segurança contra ataques como phishing, roubos de credenciais e outros. Muito mais eficaz do que apenas autenticação baseada em senhas que acabou se tornando algo "fácil" de se roubar pelos criminosos (3).

Porém, mesmo assim existem diferentes níveis de segurança nas autenticações de dois fatores, um exemplo seria se um usuário tivesse a sua senha mais a autenticação por SMS (Serviço de mensagens curtas), ainda existe o perigo do SMS ser interceptado por invasores. É muito mais seguro utilizar uma senha mais alguma senha (ou PIN) de uso único e limitado. Embora seja muito mais seguro utilizar 2FA, é bem desencorajador para usuários que não veem segurança como algo muito importante ou por preguiça de ter que utilizar o celular toda vez que for entrar em determinada aplicação (11).

2.3 AUTENTICAÇÃO BIOMÉTRICA

A biometria pode ser explicada como um atributo fisiológico e/ou comportamental quantificável que pode ser encapsulado e análogo a uma ocorrência no momento da verificação (19). Autenticação biométrica lida com o problema de identificar indivíduos baseado na fisiologia ou em características comportamentais. Como muitas das características físicas, como o rosto, íris, impressão digital, etc., são únicas do indivíduo, a análise biométrica oferece uma capacidade própria de resolver o problema de identificar o usuário.

É visto que as características fisiológicas apresentam um melhor conjunto de maneiras de resolver o problema de identificação do usuário do que as características comportamentais (1). Isso porque as características fisiológicas são sempre as mesmas ou permanecem as mesmas durante um longo período de tempo de outro lado, as características comportamentais são influenciadas por fatores emocionais, estresse e até mesmo se a pessoa está com algum tipo de problema de saúde ou não, podendo transformar a identificação e as características do usuário não linear (1).

2.3.1 Características de tecnologias Biométrica e Desafios

Quando falamos de características biométricas, especificamente para características fisiológicas, existem várias formas de poder utilizá-las para fazer a identificação do usuário. Elas podem ser o rosto, a mão, a íris, o DNA e a impressão digital, por exemplo (1). Todos são métodos válidos, porém tem alguns que são melhores que outros. A seguir, iremos explicar as qualidade e defeitos de cada característica fisiológica para a autenticação de um usuário.

- **Impressão Digital:** O reconhecimento de impressões digitais tem sido amplamente utilizado, com um desempenho notável em termos de precisão, alcançando até 99.2% de

precisão (33). Essa tecnologia, que se baseia em características geométricas, desempenha um papel crucial na identificação e nas medidas de segurança. A excepcional precisão desse método o tornou a escolha preferencial para diversas aplicações, incluindo a identificação de indivíduos nos setores governamentais e privados. A combinação única de círculos, linhas, curvas e outros padrões distintivos encontrados nas impressões digitais fornece uma base sólida para garantir a confiabilidade na identificação de usuários em uma variedade de aplicações (2).

- **Reconhecimento Facial:**Reconhecimento facial baseado em Redes Neurais (Deep Neural Networks) (29), possui uma acurácia de reconhecimento que atinge mais de 98%. A estrutura facial que é utilizada para o reconhecimento facial é baseada nos olhos, nariz, lábios e queixo e como essas características faciais variam de pessoa para pessoa é possível identificar as pessoas por essas propriedades de forma única. Existem algumas formas de utilizar o rosto como um processo de identificação única para a autenticação, o primeiro método seria um cenário clássico que é considerado como a forma "estática" que é a partir da coleta de um conjunto de imagens e o segundo seria um reconhecimento em tempo real utilizando o formato de vídeo para o processo. As imagens capturadas são utilizadas para criar um template de exemplo de cada indivíduo que então é feito uma assimilação com o processo de reconhecimento (5).

- **Reconhecimento de Íris:** O princípio de reconhecimento de íris é parecido com o reconhecimento facial, é feito baseado no reconhecimento baseado na imagem da íris (8). Podendo chegar a uma acurácia de até 97%. Porém não é tão utilizada devido aos equipamentos necessários para poder fazer esse reconhecimento. A íris pode ser definida como um diafragma redondo com localização entre a lente e a córnea do olho humano (7). Como cada íris é única de cada indivíduo é possível utilizá-la como forma de reconhecimento único. Com base em sua composição biológica, ela possui uma rede específica de tecidos que podem ser reconhecidos e identificados. A íris de um indivíduo se forma em seu primeiro ano de vida, quando as características da íris não podem ser alteradas geneticamente. O processo de reconhecimento de íris é parecido com o reconhecimento facial, é baseado na coleta de imagens dos indivíduos e então essas imagens são passadas por um extrator de dados para fazer a identificação das características de cada amostra para então preparar para o processo de autenticação (5).

Agora que temos uma breve sobre autenticação biométrica e como funciona, a tabela a seguir demonstra as vantagens e desvantagens sobre cada método de autenticação:

Biométricas	Vantagens	Desvantagens
Impressão Digital	- Um método muito seguro - Muito utilizado para o bloqueio e desbloqueio de dispositivos sem esperar ter que lembrar da senha. - Um método simples, seguro, barato e rápido de configurar para uso. - É quase como um padrão entre os métodos de autenticação biométrica. - Baixo uso de memória.	- Impressão digital pode ser considerada um método facilmente roubável. - O equipamento pode não ter uma ótima estrutura, podendo acontecer de esse componente não autorizar uma pessoa autorizada a acessar seu dispositivo devido ao dedo estar com pouco umido, suado, com um pequeno corte ou por causa de algum outro pequeno problema. - É quase como um padrão entre os métodos de autenticação biométrica.
Reconhecimento Facial	- Sensores disponíveis - Fácil de usar. - Mais confiável e preciso quando comparados a outros recursos biométricos, como a impressão digital.	- Pode ter influências da chapéus, óculos, cabelo. - Mudança no rosto devido a idade. - Influência da luz ou expressão.
Reconhecimento de Íris	- Reconhecimento estável durante toda a vida. - Mais confiável e preciso quando comparados a outros recursos biométricos, como a impressão digital. - Muito seguro (Díficil de ser roubado).	- Reconhecimento de íris é um modelo de autenticação que requer mais treinamento das máquinas. - Díficil a captura dos dados de alguns indivíduos.

Tabela 2.1: Comparativo entre características fisiológicas para autenticação (1)

2.4 OUTRAS FERRAMENTAS DE SEGURANÇA

Além dos sistemas de autenticação, existem ferramentas que ajudam a validar se um determinado produto e a comunicação com serviços estão seguros. A lista de conceitos abaixos são exemplos disso:

- Um certificado digital é um pequeno arquivo digital que contém informações para verificação de identidade, assim como um documento que contém os dados para verificar a identidade de uma pessoa. Esses certificados recebem uma assinatura digital para provar a autenticidade da autoridade que os emite, como um passaporte, carteira de identidade, papel-moeda, podendo ter uma marca da água para provar sua autenticidade. Um certificado também contém uma sequência de valores aleatórios denominada de chave pública, a qual corresponde a uma chave privada armazenada de forma separada. A posse desse certificado capacita uma entidade a autenticar-se por meio de assinaturas digitais em dados, validando assim a titularidade da chave privada e, consequentemente, a autenticidade do usuário (14).

  Embora, atualmente, os certificados digitais sejam frequentemente utilizados para verificar a identidade de indivíduos, é importante notar que a maioria das pessoas faz uso de certificados digitais em suas atividades diárias, muitas vezes sem perceber. Um exemplo seria um site que carrega um site que utiliza HTTPS, a versão com criptografia do protocolo HTTP. O protocolo TLS usa um certificado digital do site, conhecido como certificado SSL ou certificado TLS (Serasa).

- Na criptografia, uma chave é um grupo de caracteres disponibilizado em uma determinada ordem, que é utilizada em algoritmos de criptografia para alterar os dados de uma forma, que pareçam aleatórios. Da mesma forma que uma chave física tranca com um cadeado alguma coisa dentro de um cofre, e destranca caso queira pegar o que está dentro, a função dessa chave de criptografia é a mesma coisa. Com essa chave, é possível bloquear e desbloquear dados dados e somente com a chave correta é possível fazer esse processo. Sabendo a função das chaves, é possível utilizar duas formas de manter informações seguras na internet, a criptografia simétrica e criptografia assimétrica, também conhecida como criptografia de chave pública (Serasa).

  Na criptografia simétrica, as duas pessoas ou partes que desejam se comunicar utilizam a mesma chave secreta para comunicação, então as duas partes conhecem a chave secreta de cada um e essa mesma chave serve tanto para criptografar quanto para descriptografar. Por outro lado, na criptografia assimétrica, cada parte possui um par de chaves (pública e privada), onde somente cada indivíduo conhece sua chave privada e a chave pública que é possível compartilhar com qualquer pessoa. Quando alguém quer enviar uma mensagem segura para outra pessoa, usa a chave pública do destinatário para criptografar

a mensagem. Apenas o dono da chave privada correspondente pode descriptografá-la (Serasa).

- O protocolo SSL (Secure Sockets Layer), agora chamado de TLS (Transport Layer Security), é um método que é utilizado para tornar a comunicação que acontece na internet segura. Um site que utiliza HTTPS ao invés de HTTP está empregando esse tipo de segurança. Nesse contexto, o site ou a aplicação web possui uma chave pública(conhecida por todos e presente no seu certificado SSL) e uma chave privada (mantida em segredo no servidor). Quando duas partes desejam estabelecer uma conexão segura, é iniciado um processo chamado de "handshake". Durante esse procedimento, tanto o site quanto o cliente utilizam a chave pública e a chave privada para criar novas chaves de sessão que são utilizadas para criptografar as mensagens trocadas durante a sessão, mantendo a comunicação segura. De certa forma, TLS começa como criptografia assimétrica (chaves públicas e privadas) para estabelecer a conexão, e em seguida passa para criptografia simétrica(chave única) para proteger a comunicação. Cada vez que uma nova comunicação é iniciada com um cliente, novas chaves são inicializadas para garantir a segurança contínua (15).

2.5 PROBLEMAS DE SEGURANÇA RELACIONADOS A AUTENTICAÇÃO

Os sistemas de autenticação tradicionais como email e senha com dois fatores ou multi-fatores de autenticação possuem problemas de segurança e isso pode variar desde senhas fracas até falhas ocasionadas por humanos. Analisando os sistemas tradicionais, os problemas podem variar de senhas fracas, reutilização de senhas em vários serviços diferentes, vazamento de dados devido aos serviços que o usuário possui acesso e até mesmo é possível ter suas credenciais roubadas devido a algum ataque de phishing (10), (37). Já nas formas de autenticação com mais de um fator, os problemas diminuem, porém ainda existem. Assim como no método de autenticação tradicional, se o usuário utiliza o email e senha como um primeiro fator de autenticação e utiliza um dispositivo celular como segunda fator de autenticação por exemplo, podem existir problemas de o usuário ter o dispositivo roubado e ter seu segundo fator de autenticação comprometido, o usuário pode sofrer ataque de engenharia social e passar dados sensíveis ao atacante, se o usuário utiliza o SMS como uma forma de autenticar, pode sofrer ataque de Sim Swapping (36) e ter uma mensagem de SMS roubada e isso é considerado apenas o dispositivo de hardware como fator de autenticação. Quando olhamos mais a frente e olhamos para formas biométricas de autenticação, como impressão digital, os problemas crescem ainda mais, porque é possível o roubo de credenciais, com o avanço no estudo das inteligências artificiais é possível ter a voz e a face roubada.

2.5.1 Ataques nesses métodos de autenticação

Em qualquer tipo de sistema, sempre pode haver falhas de segurança ou falhas humanas. Neste capítulo iremos tratar sobre os possíveis ataques que sistemas de autenticação tradicionais quanto biométricos podem sofrer.

2.5.1.1 Sistemas Tradicionais:

- Força Bruta: Sistemas de senhas tradicionais são vulneráveis a ataques de força bruta. Força bruta e ataques de dicionários são ataques comuns contra sistemas de autenticação baseados em senhas. Força bruta é um tipo de ataque que o atacante precisa colocar todas as possíveis possibilidades de senha (palavras ou frases) dentro de um dicionário. Existem dois tipos de ataques de Força Bruta, o primeiro seria um tipo de ataque que testa todas as combinações de senhas baseado em um espaço de senhas específico para cada usuário e o outro método de ataque tenta encontrar um usuário que usa uma senha escolhida por um atacante. Esse tipo de ataque explora uma falha humana, já que está ligado a usuários que escolhem senhas simples e fáceis de lembrar (20).

- Adivinhação (Guessing Attack): Uma tentativa de adivinhação de senha pode identificar mais eficientemente a senha de um usuário do que um ataque de força bruta, já que as senhas dependem das preferências, conhecimento e experiências dos usuários. A taxa de sucesso de um ataque de adivinhação depende do conhecimento do atacante sobre o usuário-alvo. É mais fácil para o atacante adivinhar a senha do usuário quando possuem muitas informações sobre ele. Os invasores tentam adivinhar a possível senha (palavra ou frase) no dicionário definido pelas informações que se tem do usuário. A diferença entre ataques de força bruta e ataques de dicionário é que o primeiro precisa testar todas as combinações possíveis de senhas, enquanto o segundo utiliza uma lista de palavras pré-definida. Em outras palavras, os ataques de adivinhação reduzem o número de tentativas de senha (20).

- Shoulder-surfing Attack: Esse tipo de ataque é muito comum no cotidiano das pessoas. Quando estamos colocando informações nos celulares ou computadores outras pessoas podem olhar as nossas ações por trás ou ao nosso lado (pelos ombros), e podem até mesmo filmar as nossas ações (20).

- Phishing Attack: induz o usuário a visitar sites falsos através de emails ou mensagens. Dessa forma , os atacantes podem obter dados sensíveis desse usuário, como nome, senha, códigos PIN e até mesmo dados bancários dependendo do tipo de phishing que for utilizado (20).

2.5.1.2 Sistemas Biométricos:

Spoofing attack: São divididos em duas categorias principais, Artificial Synthesis (síntese artificial) e Replay Attack (Ataque de Reprodução)

- Artificial Synthesis: (4) Faz uso da biometria original para criar uma versão artificial. Não é tão difícil para uma máquina imitar características fisiológicas, impressões digitais em portas, fotos pessoais postadas nas redes sociais, imagens de câmeras em lugares públicos são exemplos de informações que acabam sendo acessíveis para um possível atacante. Esses atacantes podem utilizar modelos de facemasks e rubber fingerprint para falsificar essas características (20).

- Replay Attack: Nesse modelo, ao invés do atacante treinar alguma máquina para imitar alguma características fisiológica, é feito diretamente. O atacante pode pegar uma foto postada nas redes sociais, ou a partir da gravação da voz tentar a autenticação no sistema (20).

- Wolf Atack: Parecido com ataques de força bruta, mas voltados para sistemas biométricos. O Wolf Attack pega amostras biométricas que são similares às registradas de dados humanos, e forma um template, tentando assim bater esse template com o de algum usuário (35).

2.5.2 Defesas nesses métodos de autenticação

Devido aos problemas de segurança nos sistemas baseados em senha, alguns mecanismos de defesa foram propostos para possibilitar uma maior segurança. Um desses mecanismos é um contador que limita a quantidade de tentativas na hora de se fazer o login em alguma aplicação. E nesse questão entra um problema de quantas tentativas deve se dar ao usuário, de no paper de (21) três tentativas serial necessárias para se ter uma segurança, porém como o número de tentativas é baixo os usuários podem ter suas contas bloqueadas, porém quanto maior o número de tentativas maior o risco. E um outro método proposto foi chamado de “Automated Turing Test (ATT)” que muito utilizado nos dias de hoje, ele é basicamente um mecanismo que diferencia humanos e máquinas. Hoje em dia é muito utilizado no formato de CAPTCHAs e SMS são exemplos de ATTs.

- Shoulder-surfing Attack: Foram propostas diversos modelos de autenticação para previnir problemas de surfing attacks. É introduzido teclados baseado no qwerty, no ABC, Touch e Slide de teclado seguro (20). Porém, a conscientização do usuário sobre o assunto, que existe esse tipo de problema ao utilizar o celular ou notebooks em lugares públicos.

- Phishing: primeiro passo para se defender contra ataques de phishing, é a conscientização do usuário sobre o assunto. Monitorar a rede e enviar notificações a usuários quando

uma possível tentativa de phishing acontecer (emails), usar um filtro de spam nos emails, autenticação em dois fatores, instalar software de anti-phishing e outras opções (20).

- Spoofing Attacks: Um bom mecanismo de defesa é um que consiga mitigar todos os tipos de spoofing attacks, incluindo o de "artificial synthesis" e "replay attacks". De uma perspetiva geral, existem três categorias de defesas, são os baseados em hardware, software e uma combinação dos dois.
  - Baseado em Hardware: Reforçar a segurança física dos dispositivos utilizando autenticação biométrica, chips de segurança ou hardware dedicado para impedir acessos não autorizados (20).
  - Baseados em Software: Implementar criptografia avançada, atualizações regulares de software e firewalls robustos para proteger contra invasões e ataques cibernéticos, mantendo sistemas e dados seguros contra a manipulação por meio de inteligência artificial (20).

3 DESAFIOS: AUTENTICAÇÃO CONTÍNUA E ANÁLISE COMPORTAMENTAL

A autenticação contínua e a análise de comportamento de usuário e entidade (UEBA) representam abordagens distintas na segurança digital, cada uma com seus pontos fortes e limitações. Enquanto a autenticação contínua foca na verificação contínua e em tempo real do usuário para identificar padrões suspeitos, a UEBA analisa comportamentos posteriores em busca de anomalias, atuando em um momento subsequente ao evento.

Este capítulo fala sobre a autenticação contínua, destacando como ela lida com os hábitos e comportamentos dos usuários, especialmente em dispositivos móveis. Exploraremos métodos como a análise da forma de andar, toques na tela, forma de digitação, entre outros, e como esses podem ser combinados para uma autenticação mais robusta. Contudo, a autenticação contínua enfrenta desafios, como vulnerabilidades a diferentes tipos de ataques, como ataques de imitação (mimic attack).

Enquanto isso, a análise comportamental de usuário e entidade (UEBA) é apresentada como uma solução mais holística, utilizando algoritmos e aprendizado de máquina para detectar anomalias não apenas nos comportamentos dos usuários, mas também em dispositivos e redes. Abordaremos os benefícios, a evolução da UEBA ao longo do tempo e os algoritmos internos que a sustentam, como a Distância de Mahalanobis.

Embora ambas as abordagens sejam promissoras, este trabalho busca explorar a viabilidade e as limitações de cada uma delas. Antes de mergulhar em métodos de autenticação contínua, é crucial compreender a eficácia e os desafios enfrentados pela UEBA. A compreensão desses aspectos é essencial para desenvolver estratégias de autenticação mais seguras e eficazes, considerando os cenários de ameaças e a complexidade das interações digitais modernas.

3.1 MÉTODOS DE AUTENTICAÇÃO CONTÍNUA

Esse método de autenticação é importante porque ela pode detectar padrões comuns a partir do momento de autenticação do usuário, permitindo uma resposta rápida a ameaças. Além disso, esse método de autenticação é valiosa na mitigação de ameaças internas, podendo identificar traços incoerências realizados por usuários autenticados (6). Outro aspecto importante da autenticação contínua é a proteção contra acesso não autorizado, mesmo se as credenciais de um usuário forem comprometidas a autenticação contínua verifica continuamente se o comportamento corresponde ao do usuário legítimo. Além disso, é sempre considerado o contexto em que a atividade ocorre, incluindo fatores como localização geográfica, dispositivo utilizado e outros fatores. Neste trabalho será abordado autenticação contínua em cima de comportamentos e hábitos que um usuário possui que podem ser analisados e dentro dos sistemas biométricos, como que um sistema desse pode coletar informações do usuário para manter ele autenticado durante um período de tempo (6). Como é um campo muito vasto e existem diversas pesquisas sobre o

tema em diversos campos, neste trabalho terá um foco em métodos de autenticação contínua em celulares.

Podemos classificar de duas formas, quando se fala de autenticação baseada em métodos biométricos e autenticação baseada nos sistemas tradicionais. Nos métodos biométricos está muito relacionado a face, impressão digital e ao rosto da pessoa no decorrer do tempo e nos modelos tradicionais é baseada na localização, forma de tecla, redes e outros. Porém, podem existir uma combinação desses métodos quando se fala de celulares ou outros dispositivos, no celular por exemplo podemos ter a impressão digital para autenticar de forma normal e depois ter a autenticação contínua a partir da forma que a pessoa tecla, ou a partir de sua localização. E é sobre isso que esse capítulo será tratado, será falado sobre as formas em ambientes separados e como é feita uma junção de várias características de autenticação contínua.

- Forma de Andar: É uma técnica que se baseia na medição e análise da maneira de como o indivíduo caminha ou corre, é utilizado sinais de aceleração gerados pelo dispositivo móvel. É possível de se coletar esses dados, porque os smartphones são integrados a alguns sensores, como o acelerômetro (coleta a velocidade que a pessoa está andando), giroscópio (coleta a orientação do smartphone) e magnetômetro. Essa técnica é interessante, porque não é necessário que o usuário tenha um papel ativo para que ela funcione. Porém, mudança de direção (25), terrenos irregulares, lesões, fadiga, e outras coisas possam afetar a sua precisão (26).

- Toques: Os toques se referem às formas desenhadas manualmente na tela dos celulares, feitas a partir de um ou vários traços diferentes, consistindo de uma série de coordenadas numéricas. Esses gestos são examinados quanto à direção do toque, duração, velocidade e aceleração do movimento, tanto de maneira independente quanto combinada. Para a coleta de dados, é utilizado sensores de tela sensível ao toque do smartphone, capturando as ações de entrada relacionadas a parâmetros como velocidade, tamanho e comprimento. Direção ou pressão, transformando-os em um modelo de gestos único para cada usuário. Esses parâmetros, que variam entre usuários, indicam o comportamento individual e formam a base dos sistemas de autenticação por gestos de toques (26).

- Forma de teclar: O processo de registrar os inputs do teclado digitados por um indivíduo em um dispositivo móvel e a tentativa de identific-alo por meio de uma análise baseada em seus hábitos de digitação, é conhecida como dinâmica de digitação (Keystroke dynamics) (24). Alguns pesquisasdores fazem a coletam de dados a partir de textos pré-definidos, como durante a digitação de uma mensagem de texto ou durante o login ao inserir senhas. Em ambos os casos, os resultados possuem alta precisão. Algumas das características que são verificadas neste processo, são: Duração, latência, pressão e localização (34).

  - Duração: Corresponde ao período de tempo entre o ato de pressionar e o ato de soltar uma tecla.
  - Latência: Tempo entre teclas, é o período de tempo decorrido entre soltar uma tecla pressionada e pressionar a próxima tecla.
  - Pressão: É a força exercida em uma tecla.
  - Localização: São os pontos de localização do dedo, ou a área do dedo na tela.

- Perfil Comportamental: Dados de uso em dispositivos móveis são explorados para autenticar indivíduos por meio de padrões comportamentais estabelecidos ao interagirem com aplicativos e serviços digitais. A coleta de dados varia desde a interação com redes Wi-Fi até o uso de aplicativos, incluindo informações de localização, duração de sessões e frequência de uso (34). A extração de características considera aspectos como tempo de pressão, latência entre pressões, dados de localização, Bluetooth, Wi-Fi, uso de aplicativos, gestos de toque e dinâmica de digitação. Os estudos analisam uma ampla gama de informações, desde coordenadas de GPS até padrões de caminhada registrados por sensores, com o intuito de identificar padrões únicos de uso associados a cada usuário, para fins de autenticação comportamental (34).

- Mover das mãos: O interesse recente na identificação de usuários pelo padrão de movimento do pulso ao interagir ou segurar smartphones dispensa ações extras do usuário além de segurar o dispositivo. Características como torção do pulso, velocidade, amplitude e frequência são exploradas para distinguir indivíduos pela diferença no movimento de suas mãos (34). É utilizado dos sensores de acelerômetro, giroscópio e magnetômetro para capturar Movimento da Mão, Orientação e Agarrar (HMOG), observando a orientação do dispositivo e movimentos sutis derivados da forma como o usuário toca no smartphone após segurá-lo (9).

- Bateria: A correlação entre os padrões de comportamento do usuário e os padrões de consumo de energia é a base desse método. Até o momento, o consumo de energia não é utilizado como um identificador único em sistemas biométricos, mas sim em combinação com outras características biométricas (34).

- Combinação de métodos: Fusão de dados biométricos por meio de um único sensor utilizando múltiplas instâncias de uma biometria e a fusão alcançada por múltiplos sensores, com foco específico na fusão multimodal de biometria comportamental. A principal meta dos sistemas multimodais é reunir informações de diversas fontes e modalidades distintas para aprimorar a precisão da autenticação (34).

3.1.1 Problemas da Autenticação Contínua

Esses tipos de sistemas enfrentam diversas ameaças, como ataques de malware, vigilância por ombro (shoulder surfing attack), imitação (mimic), falsificação (impersonation), repetição (replay), estatísticos, algorítmicos, entre outros. Esses ataques podem ser divididos em duas categorias principais: ataques de zero esforço (zero-effort), nos quais não há ações sofisticadas pelo adversário e se baseiam na similaridade suficiente entre os modelos do atacante e do usuário legítimo, e os ataques adversários (advsarial attacks), nos quais ações sofisticadas são empreendidas para a imitação bem-sucedida. Os pesquisadores biométricos concentram-se especialmente nos ataques de imitação, pois não requerem modificações no sistema de autenticação e podem ser lançados facilmente em sistemas biométricos (34). Há três tipos de ataques de imitação nos sistemas de autenticação baseados em biometria comportamental: zero esforço, mínimo esforço (minimal-effort) e alto esforço (high-effort). No ataque de zero esforço, os imitadores são selecionados aleatoriamente pelos invasores, enquanto no mínimo esforço são escolhidos de acordo com critérios específicos, e no alto esforço, são intensivamente treinados para imitar o usuário legítimo (22).

- Formas de andar: Vários estudos abordam a vulnerabilidade da autenticação baseada em padrões da forma de andar de um indivíduo. Existem estudos que examinam ataques de impostores passivos de zero esforço, ataques de imitação de impostores ativos, incluindo ataques de impersonação de esforço mínimo e ataques após treinamento do adversário. Esses ataques são divididos em dois tipos: passivos, que não envolvem ações sofisticadas do adversário, e ativos, onde os invasores tentam ser aceitos pelo sistema de autenticação (34).

- Toques: Estudos recentes, como (18), exploram os riscos em sistemas de autenticação baseados em toque, destacando ameaças como observação não autorizada e ataques de malware. Embora existam evidências de resistência desses sistemas a ataques simples, outras ameaças preocupam: a variabilidade das biometrias comportamentais entre usuários, permitindo a geração de falsificações a partir de bancos de dados populacionais, e a possibilidade de implementação de gestos usando dispositivos robóticos comerciais (32). Estudos como os de Serwadda et al. (32) mostram a eficácia de robôs "Lego"na criação de falsificações, elevando as taxas de erro dos sistemas de autenticação em até 1004%. Essas pesquisas questionam a eficácia das abordagens de avaliação, especialmente diante de ataques sofisticados.

- Forma de Teclar: Vários estudos investigam a vulnerabilidade de sistemas de autenticação baseados em digitação. Existêm alguns ataques em relação a forma de teclar, eles são o ataque Frog-Boiling, o ataque Algorítmico, ataque de imitação (mimic), e o ataque Snoop-forge-replay (34). Esses ataques funcionam tanto em teclados virtuais quanto físicos e apresentam descobertas relevantes e similares.

3.2 ANÁLISE COMPORTAMENTAL DE ENTIDADE E USUÁRIO (UEBA)

A UEBA (*User and Entity Behavior Analytics*) é uma solução de cibersegurança que utiliza algoritmos e aprendizado de máquina para detectar anomalias no comportamento não apenas dos usuários, mas também dos roteadores, servidores e dispositivos finais em uma rede corporativa (17). Ela visa reconhecer comportamentos peculiares ou suspeitos, identificando irregularidades em padrões de uso normais. A UEBA não se limita à monitoração do comportamento humano, estendendo-se à análise do comportamento de máquinas. Por exemplo, ela pode detectar um aumento incomum de solicitações a um servidor, indicando um possível ataque de negação de serviço distribuído (DDoS) (17).

Para funcionar efetivamente, a UEBA deve ser instalada em todos os dispositivos usados por cada funcionário na organização, incluindo dispositivos pessoais. Os três principais componentes da solução são a coleta e organização de dados, a integração com outros sistemas de segurança, e a apresentação de descobertas e respostas apropriadas (17). Os benefícios da UEBA incluem a detecção de uma ampla gama de ameaças cibernéticas, redução da necessidade de analistas de TI, diminuição de custos, e a redução do risco ao identificar comportamentos anômalos. Ela se destaca em detectar ameaças mais sofisticadas que podem passar despercebidas por outras soluções de segurança (17).

A UEBA vem como uma resposta aos desafios crescentes na segurança, inicialmente introduzida em 2015 como User Behavioral Analytics (UBA), a abordagem evoluiu para incluir o monitoramento e análise do comportamento não apenas de usuários individuais, mas também de entidades, como dispositivos e redes. Essa expansão reflete a compreensão de que as ameaças modernas vão além das interações de usuários isolados e frequentemente envolvem comportamentos complexos que podem ser detectados por meio do monitoramento abrangente de entidades inter-relacionadas (30). A UEBA incorpora o aprendizado de máquina não supervisionado para identificar mudanças significativas nas atividades de usuários, dispositivos e redes, indicando assim a presença de ataques em tempo real. A capacidade da UEBA de construir perfis de comportamento ao longo do tempo, adaptando-se dinamicamente a alterações nas funções dos usuários e nas condições do ambiente, destaca-se como uma evolução crucial. Essa abordagem permite uma detecção mais precisa de anomalias, contribuindo para a segurança proativa das redes. Além disso, a inclusão de automação na resposta a alertas de segurança reforça a eficiência da UEBA, minimizando o tempo de resposta a incidentes de segurança e fortalecendo a postura global de defesa cibernética das organizações (30).

Os algoritmos internos da UEBA baseiam-se na Distância de Mahalanobis, uma medida de anomalia que avalia o quão longe uma observação está da média de todas as observações em um conjunto multidimensional (23). Essa distância é útil para detectar eventos improváveis, considerando a distribuição empírica da variável e utilizando o desvio padrão para comparar observações. Na prática, são usadas estimativas empíricas da média e matriz de covariância, sendo

necessário precaução com a computação desses valores para evitar problemas de estabilidade numérica.

Para contornar tais limitações, o algoritmo proposto na UEBA utiliza uma abordagem alternativa que evita o cálculo da matriz de covariância inversa. Em vez disso, normaliza as variáveis, realiza uma Decomposição de Valores Singulares (SVD) para identificar componentes importantes e, em seguida, calcula a distância Mahalanobis utilizando esses componentes reduzidos (23).

Além disso, para tornar a abordagem mais flexível, são propostas extensões que permitem desvios unilaterais, ponderação de variáveis e robustez a outliers. Essas extensões visam atender a diferentes requisitos de segurança, como identificar desvios apenas em uma direção específica, atribuir diferentes pesos às variáveis e lidar com outliers de forma mais eficaz. A metodologia visa também tornar os resultados mais explicáveis, além de mapear a distância de Mahalanobis para uma pontuação de confiança entre 0 e 100, facilitando a interpretação dos resultados (23).

3.2.1 Implantação da UEBA em um ambiente real

A implantação da UEBA em uma empresa é uma opção a se analisar devido à sua capacidade singular de analisar não apenas o comportamento dos usuários, mas também dos dispositivos e sistemas em uma rede corporativa (23). Essa abordagem abrangente oferece uma visão holística das atividades na rede, permitindo a detecção precoce de ameaças sofisticadas que podem passar despercebidas por outras soluções de segurança. A adaptabilidade da UEBA às mudanças no ambiente e a capacidade de identificar padrões de comportamento incomuns são fundamentais para reduzir os falsos positivos, permitindo uma resposta mais rápida e precisa a possíveis incidentes de segurança (30).

Além disso, sua capacidade de aprendizado contínuo, combinada com a automação de respostas a alertas de segurança, fortalece significativamente a postura de defesa cibernética da organização (23). Com a UEBA, é possível não apenas identificar anomalias de forma eficaz, mas também personalizar suas configurações para atender às necessidades específicas de segurança de cada empresa, oferecendo uma segurança mais proativa e abrangente. Essa flexibilidade e adaptabilidade fazem da UEBA uma solução altamente valiosa para proteger os ativos digitais de uma organização e manter suas operações seguras diante de ameaças cada vez mais sofisticadas (17).

3.2.2 Problemas da UEBA

O User Entity Behavior Analytics (UEBA) enfrenta uma série de desafios consideráveis na detecção eficaz de ameaças. A natureza do aprendizado de máquina subjacente ao UEBA pode encontrar dificuldades ao lidar com usuários privilegiados e insiders, cujas atividades muitas vezes fogem dos padrões estabelecidos, tornando a construção de perfis precisos um desafio. Além disso, o UEBA pode não ser ágil o suficiente para identificar ataques que se desenvolvem

lentamente ao longo do tempo, como as ameaças de "baixo e lento", apresentando obstáculos na detecção de ações maliciosas que não impactam imediatamente as operações cotidianas das organizações (30). A capacidade limitada de se adaptar rapidamente a ataques zero-day, que exploram vulnerabilidades desconhecidas, é outro desafio crítico (30). Esses ataques evasivos, que não seguem padrões pré-definidos, representam um grande risco para a eficácia do UEBA, exigindo constantes atualizações e refinamentos para manter a precisão na identificação de comportamentos anômalos e ameaças cibernéticas.

4 PROPOSTA DO PROJETO

Assim como foi visto no decorrer deste trabalho, os métodos tradicionais de autenticação tanto os baseados em senhas quanto os biométricos possuem falhas de segurança e até mesmo os sistemas de autenticação contínua. O projeto proposto neste trabalho é a implementação de um *framework* que realiza a coleta de dados em dispositivos Android com a ideia de que no futuro todos esses dados sejam utilizados para a tentativa de autenticar usuário baseado em seu comportamento ao longo do tempo.

Esse projeto de coleta de dados em dispositivos Android tem um papel fundamental ao fornecer informações valiosas para testes na UEBA. A capacidade de coletar dados detalhados dos dispositivos permitirá uma análise mais ampla do comportamento dos usuários. Além disso, ao reunir esses dados, será possível alimentar e fortalecer os mecanismos de autenticação contínua.

4.1 MOTIVAÇÃO

A autenticação contínua surge como uma solução promissora, pois vai além da verificação única da identidade do usuário e monitora as atividades do usuário de forma constante, verificando o comportamento e dados contextuais para verificar a sua identidade de forma constante. Assim como foi visto no artigo - ONLINE BINARY MODELS ARE PROMISING FOR DISTINGUISHING TEMPORALLY CONSISTENT COMPUTER USAGE PROFILES - nesse artigo é feita a coleta de dados de processos, dados de rede, mouse, teclado e outros eventos. Com esses dados, é feito uma análise utilizando um conjunto de séries temporais e modelos de machine learning tentando resolver três perguntas, se os hábitos desses usuários se repetem ? Se os usuários possuem perfis únicos e/ou distintos ? e o último ponto a ser levantado seria quais dados são importante para se construir um “profiling” (perfil único) ? Sabendo disso, o Framework proposto neste projeto visa a coleta de dados para dispositivos android com a intenção de se responder essas perguntas e se é um método viável de autenticação em um trabalho futuro.

4.2 EXPLICAÇÃO DETALHADA DO TRABALHO

Para ser possível fazer esse trabalho funcionar, é necessário que o usuário ative a função de acessibilidade no celular e permita que esse aplicativo requisite todas as permissões necessárias para a coleta de dados. Para demonstrar as funcionalidades do aplicativo, será feito via diagrama de classes sobre como as coisas acontecem e se conectam. O ponto principal de partida do código é um arquivo chamado “MainActivity.java” que é o ponto inicial da aplicação, que é criado logo que o aplicativo é instalado e a principal lógica, que faz o serviço de acessibilidade funcionar é o arquivo “WindowEnumerationService.java”.

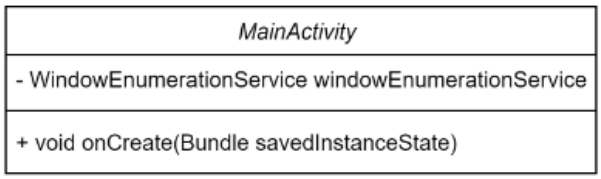

Figura 4.1: MainActivity do Framework

O arquivo “MainActivity.java” possui apenas duas coisas, preciso de uma referência do meu serviço de acessibilidade que começa a rodar no momento que uma instância do aplicativo é criada. No método onCreate() é criada uma instância do “WindowEnumerationService”, sendo possível assim criar meu serviço de acessibilidade que irá fazer a coleta dos dados. A classe abaixo é uma representação de como foi implementado.

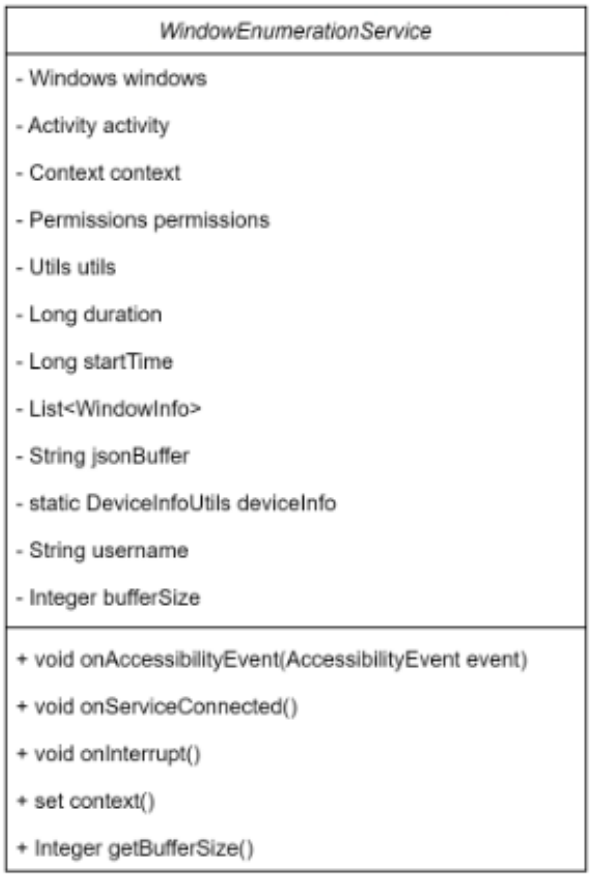

Figura 4.2: WindowsEnumerationService classe do Framework

Acaba sendo uma classe bem grande, já que ela é a parte que faz todo o controle do projeto. Mas, começando do básico, temos algumas informações que são importantes que acontecem assim que o serviço é trigado. Algumas coisas no android, acontecem por evento e o projeto desenvolvido é baseado nisso. Por exemplo, temos o método onServiceConnected() que é a primeira coisa que vai ser chamada assim que o serviço de acessibilidade rodar pela primeira vez, nele eu coleto algumas informações, configuro o contexto de aplicação e algumas classes que são inicializadas nessa parte, como a classe Windows e Utils.

Agora que temos esse serviço acionado no celular do usuário, é necessário utilizar o método “onAccessibilityEvent”, nesse método eu consigo selecionar qual evento eu quero estar monitorando. No caso deste projeto, é feita a monitoração do evento “TYPE_WINDOW_STATE_CHANGED” que indica quando o usuário selecionou uma janela ou quando está mudando de janela e os dados coletados para a análise futura, segue nessa linha que é sobre como o usuário utilizar o seu celular baseado no aplicativos que é utilizado, por isso que é feito o monitoramento desse evento, sendo possível fazer a coleta das janelas que está selecionada e as que estão ativas.

Esse é o ponto principal, agora é sobre como pegar essas informações para fazer a coleta de dados e enviar para um servidor. Para facilitar a forma de organizar esses dados para serem enviados, é feita uma classe chamada “WindowInfo“, que é demonstrada no diagrama abaixo:

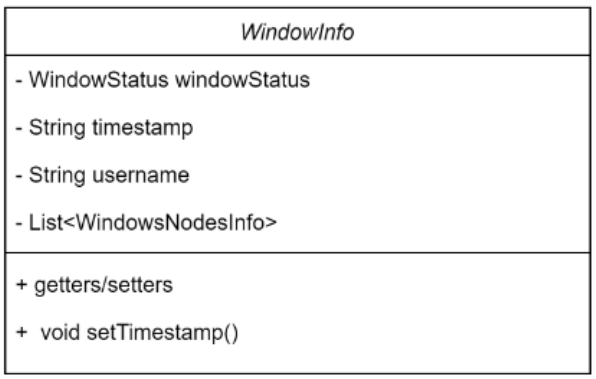

Figura 4.3: WindowsInfo classe do Framework

Essa classe coleta quatro informações, elas são:

- Status da Janela: Indica se o usuário está utilizando a janela no momento ou se está em background.
- Timestamp: Indica o momento atual.
- String: Indica o nome do usuário, é coletado a partir de uma informação que o Android disponibiliza e é misturado com um script interno, para garantir que o dado coletado fique anônimo.
- Lista de Janelas: Coleta o título e o nome do pacote de cada janela que está aberta no dispositivo do usuário.

  É sempre importante verificar na coleta do título da janela, porque existem casos que a janela não tem título e para isso, foi coletado o nome do pacote e transformado em título, com uma flag indicando os casos em que isso aconteceu, como na classe abaixo:

A partir do momento em que os dados estão todos coletados e no formato correto é necessário criar uma requisição para enviar esses dados ao servidor. Para isso, é feito uma

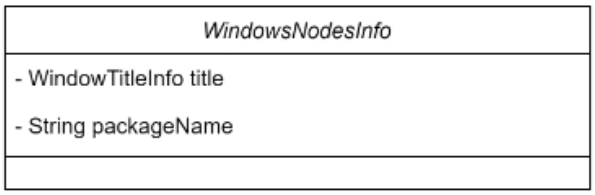

Figura 4.4: WindowsNodesInfo classe do Framework

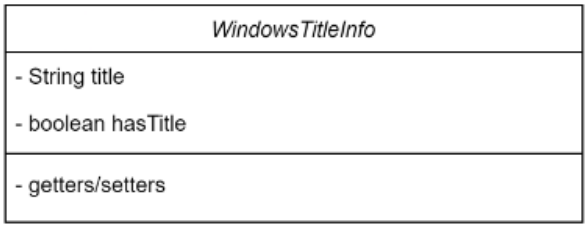

Figura 4.5: WindowsTitleInfo classe do Framework

serialização desse objeto para “Json” e então é criado uma requisição desses dados que serão enviados ao servidor. Para fazer a transição de uma classe para Json, é feito utilizando uma feature do java chamada de “objectMapper “. Um exemplo da classe de um request é mostrado abaixo:

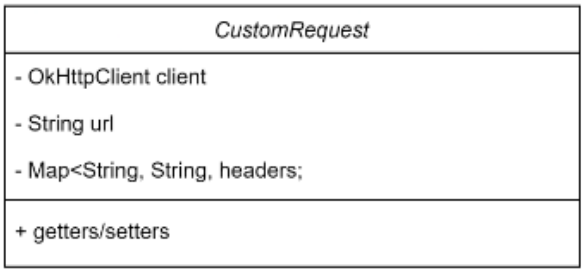

Figura 4.6: CustomRequest classe do Framework

Depois que todo esse processo fica pronto, eu é necessário verificar se o tamanho do buffer mostrado no diagrama “WindowsEnumerationService” (List<Windows>) está cheio ou não, para evitar ser necessário enviar requisições a todo momento, é enviado a requisição assim que o buffer dos dados ficar cheio, enviando assim várias informações de uma vez só. Isso é importante, porque ao fazer isso é necessário realizar de forma assíncrona e toda vez que isso é feito, é necessário avisar para o Android que está sendo enviado um request e que esse request não será feito a partir do fluxo original de execução, mas sim a partir de uma outra thread.

Agora que se tem a verificação do buffer cheio é necessário verificar se o tempo que esse buffer está cheio, ainda não se esgotou. Isso é necessário, porque pode ser que o usuário pegou o dispositivo, mexeu um pouquinho e deixou ele de canto. Esse tempo verifica se deu o tempo base necessário, se o tempo se esgotou ou se por algum motivo o serviço foi interrompido, significa que a requisição precisa ser enviada, evitando assim uma possibilidade de perder esses

dados. E o Android provisiona um outro método chamado “onInterrupt()” que indica o que o código deve fazer para o caso de o serviço parar de funcionar.

Agora é só enviar esses dados para o servidor, onde os dados serão armazenados para análise futura. Já no lado do servidor, é um código bem mais simples, existe apenas uma única rota que serve para receber os dados enviados pelos clientes, e duas outras funções, uma serve para fazer o log do buffer dos dados que são recebidos dos clientes e uma função que é um contador, que serve para resetar o contador do buffer. É feito isso, para limitar o buffer do servidor e para economizar I/O. O servidor foi feito utilizando Python em junto ao aiohttp de forma assíncrona para possibilitar uma maior capacidade de receber os dados dos diversos clientes. O código abaixo é um exemplo de como é feito o lado do servidor para receber os dados.

```
async def save_windows(request):
    global counter

    try:
        data = await request.json()
        data_str = json.dumps(data).replace("'", '"')
        data_str = data_str.replace('false', '"False"')
                           .replace('true', '"True"')



        async with lock:
            await buffer.put(data_str)
            counter += 1

            if counter == 10:
                await asyncio.gather(log_buffer(), reset_counter())

        response_text = "Ok"
        return web.Response(text=response_text, status=200,
                            content_type="text/plain")

    except Exception as e:
        return web.Response(
            text=f"Error: {str(e)}",
            status=500,
            content_type="text/plain",
        )
```

```
async def log_buffer():
    while not buffer.empty():
        elem = await buffer.get()
        print(elem)
        logger.info(elem)
    logger.rotate_logs()


async def reset_counter():
    global counter
    counter = 0
```

Para fazer o deploy desse servidor, foi considerado preocupações como a quantidade de requisições e segurança, já que a versão de desenvolvimento do python não é o ideal. Para isso, foram utilizadas três tecnologias: o Nginx, Supervisord e Gunicorn. O supervisord serve para casos em que o processo está rodando possa cair, o Gunicorn é um WSGI HTTP e o Nginx é um proxy que vai filtrar as requisições. Com isso tudo, o exemplo abaixo é um exemplo de um caso de log que é enviado do dispostivo (cliente) que é recebido pelo servidor:

```
{
    "timestamp": "2023-07-10T20:06:42.538",
    "windowStatus": "ACTIVE",
    "windowsNodes": [
        {
            "packageName": "com.android.settings",
            "title": {
                "hasTitle": "False",
                "title": "Settings"
            }
        }
    ]
}
```

4.3 DISCUSSÃO

Desenvolver um projeto na universidade frequentemente enfrenta desafios significativos, especialmente quando se trata de coletar dados para pesquisas. A necessidade de voluntários

dispostos a ceder informações e permitir a coleta de dados pode ser um obstáculo, já que nem todos estão dispostos ou têm tempo para participar ativamente. Além disso, existem limitações inerentes à quantidade e qualidade dos dados coletados, pois a diversidade dos voluntários pode influenciar diretamente na representatividade dos dados.

Após o desenvolvimento do framework, embora tenha havido um interesse inicial considerável, apenas uma entre as dez utilizações permanece ativa de forma constante. Isso destaca um desafio adicional na manutenção do engajamento e na garantia de uma participação contínua ao longo do tempo, o que é essencial para a validade e relevância contínua do projeto. O *framework* desenvolvido está disponível para download e uma breve explicação sobre a pesquisa está na página do Prof. Marcus Botacin (https://www.inf.ufpr.br/mfbotacin/pesquisa/).

5 CONCLUSÃO

O panorama da autenticação continua a evoluir, impulsionado pela necessidade crescente de segurança digital e pela busca por métodos mais eficazes de validação de identidade. Ao longo deste trabalho, exploramos a autenticação básica, biométrica e contínua, destacando suas vantagens, desafios e a constante necessidade de defesas mais robustas. A autenticação contínua em dispositivos móveis emerge como uma abordagem promissora, indo além da validação pontual e acompanhando os padrões de comportamento dos usuários ao longo do tempo. Esta técnica, baseada em dados contextuais, comportamentais e biométricos, oferece uma camada adicional de segurança, adaptando-se dinamicamente às nuances individuais do usuário.

No contexto de trabalhos futuros, aprimorar a autenticação contínua em dispositivos móveis torna-se crucial. Explorar técnicas mais avançadas de aprendizado de máquina e inteligência artificial para análise de comportamento do usuário pode levar a sistemas mais precisos e adaptáveis. Além disso, integrar métodos de autenticação contínua entre dispositivos móveis, desktops e ambientes web apresenta um desafio e uma oportunidade interessante. Pensando em um ecossistema em que os usuários se autentiquem de forma contínua e transparente, independentemente do dispositivo ou plataforma que estão usando. Isso exigirá um esforço significativo na padronização, sincronização e interoperabilidade entre sistemas, trazendo beneficios em termos de segurança e experiência do usuário.

A abordagem futura pode envolver a criação de um framework unificado de autenticação contínua, onde os dados comportamentais e biométricos dos usuários são coletados e processados de maneira consistente em diferentes dispositivos. Isso exigirá um foco em protocolos de comunicação seguros, algoritmos de análise de dados adaptáveis e uma forte ênfase na privacidade e proteção de dados dos usuários. Em suma, a autenticação contínua representa um ponto crucial no avanço da segurança digital. Ao integrar e aprimorar essa abordagem em dispositivos móveis e além, podemos construir sistemas mais confiáveis, adaptáveis e centrados no usuário, estabelecendo novos padrões de autenticação segura em um cenário cada vez mais conectado e diversificado de tecnologias.

REFERÊNCIAS

[1] Abdulrahman, S. A. e Alhayani, B. (2023). A comprehensive survey on the biometric systems based on physiological and behavioural characteristics. *Materials Today: Proceedings*, 80:2642–2646. SI:5 NANO 2021.

[2] Abdulrahman, S. A., Khalifa, W., Roushdy, M. e Salem, A.-B. M. (2020). Comparative study for 8 computational intelligence algorithms for human identification. *Computer Science Review*, 36:100237.

[3] Acemyan, C. Z., Kortum, P., Xiong, J. e Wallach, D. S. (2018). 2fa might be secure, but it's not usable: A summative usability assessment of google's two-factor authentication (2fa) methods. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 62(1):1141–1145.

[4] Adler, A. e Schuckers, S. (2009). *Biometric Vulnerabilities, Overview*, páginas 160–168. Springer US, Boston, MA.

[5] Alsaadi, I. (2015). Physiological biometric authentication systems, advantages, disadvantages and future development: A review. *International Journal of Scientific Technology Research*, Volume 4.

[6] Baig, A. F. e Eskeland, S. (2021). Security, privacy, and usability in continuous authentication: A survey. *Sensors*, 21(17).

[7] Bapat, A. e Kanhangad, V. (2017). Segmentation of hand from cluttered backgrounds for hand geometry biometrics. *2017 IEEE Region 10 Symposium (TENSYMP)*, páginas 1–4.

[8] Bazrafkan, S. e Corcoran, P. (2018). Enhancing iris authentication on handheld devices using deep learning derived segmentation techniques. Em *2018 IEEE International Conference on Consumer Electronics (ICCE)*, páginas 1–2.

[9] Buriro, A., Crispo, B. e Zhauniarovich, Y. (2017). Please hold on: Unobtrusive user authentication using smartphone's built-in sensors. Em *2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA)*, páginas 1–8.

[10] Cisco (2023). O que é sim swap. `https://www.cisco.com/c/en/us/products/security/email-security/what-is-phishing.html#~ai-and-phishing`. Acessado em 19/11/2023.

[11] Cloudflare. Certificado digital. `https://www.cloudflare.com/pt-br/learning/access-management/what-is-two-factor-authentication/`. Acessado em 19/11/2023.

[12] Cloudflare. Diferença de autenticação e autorização. https://www.cloudflare.com/pt-br/learning/access-management/authn-vs-authz/. Acessado em 19/11/2023.

[13] Cloudflare. Mfa. https://www.cloudflare.com/pt-br/learning/access-management/what-is-multi-factor-authentication/. Acessado em 19/11/2023.

[14] Cloudflare. Mfa. https://serasa.certificadodigital.com.br/blog/certificado-digital/o-que-e-certificado-digital-e-para-que-serve/. Acessado em 19/11/2023.

[15] Cloudflare. O que é autenticação? https://www.cloudflare.com/pt-br/learning/access-management/what-is-authentication/. Acessado em 19/11/2023.

[Devoteam] Devoteam. Authentication and authorization to secure api's. https://nl.devoteam.com/expert-view/authentication-and-authorization-to-secure-apis/. Acessado em 19/11/2023.

[17] Fortinet (2023). O que é a ueba. https://www.fortinet.com/resources/cyberglossary/what-is-ueba. Acessado em 19/11/2023.

[18] Frank, M., Biedert, R., Ma, E., Martinovic, I. e Song, D. (2013). Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication. *IEEE Transactions on Information Forensics and Security*, 8(1):136–148.

[19] Jain, A. K., Nandakumar, K. e Ross, A. (2016). 50 years of biometric research: Accomplishments, challenges, and opportunities. *Pattern Recognition Letters*, 79:80–105.

[20] Jeyaraman, S. e Topkara, U. (2005). Have the cake and eat it too - infusing usability into text-password based authentication systems. Em *21st Annual Computer Security Applications Conference (ACSAC'05)*, páginas 10 pp.–482.

[21] Kirushnaamoni, R. (2013). Defenses to curb online password guessing attacks. páginas 317–322.

[22] Kumar, R., Phoha, V. V. e Raina, R. (2016). Authenticating users through their arm movement patterns.

[23] Madhu Shashanka, Min-Yi Shen, J. W. (1985). User and entity behavior analytics for enterprise security.

[24] Mahfouz, A., Mahmoud, T. M. e Eldin, A. S. (2017). A survey on behavioral biometric authentication on smartphones. *Journal of Information Security and Applications*, 37:28–37.

[25] Muaaz, M. e Mayrhofer, R. (2014). Orientation independent cell phone based gait authentication. Em *Proceedings of the 12th International Conference on Advances in Mobile Computing and Multimedia*, MoMM '14, página 161–164, New York, NY, USA. Association for Computing Machinery.

[26] Murmuria, R., Stavrou, A., Barbará, D. e Fleck, D. (2015). Continuous authentication on mobile devices using power consumption, touch gestures and physical movement of users. Em Bos, H., Monrose, F. e Blanc, G., editores, *Research in Attacks, Intrusions, and Defenses*, páginas 405–424, Cham. Springer International Publishing.

[27] Ometov, A., Bezzateev, S., Mäkitalo, N., Andreev, S., Mikkonen, T. e Koucheryavy, Y. (2018). Multi-factor authentication: A survey. *Cryptography*, 2(1).

[28] Petsas, T., Tsirantonakis, G., Athanasopoulos, E. e Ioannidis, S. (2015). Two-factor authentication: Is the world ready? quantifying 2fa adoption.

[29] Ranjan, R., Bansal, A., Zheng, J., Xu, H., Gleason, J., Lu, B., Nanduri, A., Chen, J.-C., Castillo, C. D. e Chellappa, R. (2019). A fast and accurate system for face detection, identification, and verification. *IEEE Transactions on Biometrics, Behavior, and Identity Science*, 1(2):82–96.

[30] Salitin, M. e Zolait, A. (2018). The role of user entity behavior analytics to detect network attacks in real time. páginas 1–5.

[Serasa] Serasa. Certificado digital. `https://www.cloudflare.com/pt-br/learning/ssl/what-is-an-ssl-certificate/`. Acessado em 19/11/2023.

[32] Serwadda, A. e Phoha, V. V. (2013). When kids' toys breach mobile phone security. Em *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, CCS '13, página 599–610, New York, NY, USA. Association for Computing Machinery.

[33] Shreyas, K. K., Rajeev, S., Panetta, K. e Agaian, S. S. (2017). Fingerprint authentication using geometric features. Em *2017 IEEE International Symposium on Technologies for Homeland Security (HST)*, páginas 1–7.

[34] Stylios, I., Kokolakis, S., Thanou, O. e Chatzis, S. (2021). Behavioral biometrics continuous user authentication on mobile devices: A survey. *Information Fusion*, 66:76–99.

[35] Une, M., Otsuka, A. e Imai, H. (2007). Wolf attack probability: A new security measure in biometric authentication systems. páginas 396–406.

[36] Wikipedia (2023a). O que é sim swap. `https://en.wikipedia.org/wiki/SIM_swap_scam`. Acessado em 19/11/2023.

[37] Wikipedia (2023b). O que é social engineer. `https://en.wikipedia.org/wiki/Social_engineering_(security)`. Acessado em 19/11/2023.