Uma análise de características de uso de aplicações em dispositivos móveis para autenticação ativa

Felipe Alberto Keller[1], **Daniel Ribeiro da Rosa**[1], **Luciano Ignaczak**[1]

[1]Universidade do Vale do Rio dos Sinos (UNISINOS)
Caixa Postal 950 – 93022-000 – São Leopoldo – RS – Brasil

{felipeakeller,xdanielribeiro}@gmail.com, lignaczak@unisinos.br

Abstract. *In the face of technological advances that insert mobile devices increasingly in everyday life, the use of methods is needed to ensure the security of personal information. In order to address this, new forms of authentication have been studied to find methods that can identify the user actively, through behavioral biometrics. This paper presents an analysis of features of the use on mobile devices that allow user identification. The results demonstrate the feasibility of applying the characteristics of the use on mobile devices as a way to perform active authentication.*

Resumo. *Diante dos avanços tecnológicos que inserem os dispositivos móveis cada vez mais no cotidiano, é necessário o uso de métodos para garantir a segurança das informações pessoais. A fim de resolver esta questão, novas formas de autenticação têm sido estudadas para encontrar métodos que possam identificar o usuário de forma ativa, através de biometria comportamental. Este artigo apresenta uma análise das características de uso em dispositivos móveis que permitem a identificação do usuário. Os resultados demonstram a viabilidade em aplicar as características de utilização dos dispositivos móveis como uma forma de realizar a autenticação ativa.*

1. Introdução

A denominação de biometria como conhecemos pode ser considerada relativamente nova, porém seu conceito na forma mais branda surgiu no final do século XIX. Francis Galton foi um dos primeiros a realizar pesquisas científicas na área, classificando traços de identificação dentro de cada impressão digital. O seu trabalho deu origem a primeira aplicação prática, utilizada pelas forças policiais, como forma de identificação [Ashbourn 2014]. O conceito de biometria está ligado diretamente à individualização, como definido por [Jain et al. 2011] com "a determinação da identidade de um indivíduo a partir de suas características físicas e/ou comportamentais".

Biometria comportamental é descrita por [Clarke 2011] como "o processo de autenticação de uma pessoa utilizando suas interações específicas em aplicações e/ou serviços". Ainda conforme o autor, a autenticidade do usuário depende de como é feito uso destas aplicações, quando elas são usadas, por quanto tempo, entre outras características. Esta técnica é aplicada de forma não invasiva, o que possibilita um monitoramento contínuo na identificação dos usuários [Modi 2011].

A utilização de métodos biométricos pode oferecer uma gama de vantagens em relação aos métodos tradicionais. [Jain et al. 2011] cita como vantagens da biometria,

algo que não pode ser facilmente transferido, esquecido, perdido ou copiado e considera também um aumento na conveniência do usuário. Em contrapartida, dificilmente um sistema biométrico garante 100% de precisão, devido à variabilidade inerente no processo de aquisição da amostra [Ashbourn 2014].

De acordo com [Alzubaidi and Kalita 2015], devido a popularização dos dispositivos móveis, problemas relacionados a privacidade e segurança tornaram-se primordiais. A grande maioria dos usuários tendem a armazenar informações sensíveis nos *smartphones*, onde uma vez perdido ou acessado por pessoas mal intencionadas, pode acarretar em sérios problemas. Técnicas de autenticação amplamente utilizadas em dispositivos móveis, incluindo senhas, PIN e padrões secretos, demonstram deficiências em lidar com ataques. Diante deste cenário, uma das soluções mais adequadas para esses problemas de segurança é a utilização de autenticação baseada em biometria comportamental.

[Fridman et al. 2015] definem a autenticação ativa como um acompanhamento contínuo das características biométricas comportamentais de um usuário com base na sua interação com um dispositivo. A partir dessa definição, este trabalho buscará responder a seguinte questão de pesquisa: é possível estabelecer a autenticação ativa de um usuário de um dispositivo móvel com base apenas nas características de uso de aplicações?

Para responder essa pergunta, este trabalho possui como objetivo principal analisar se a biometria comportamental, baseada nas características de uso de aplicações em um dispositivo móvel, pode ser usada para autenticação ativa. Para realizar a análise proposta foi necessário desenvolver uma aplicação que permitiu capturar os dados de uso do dispositivo móvel de uma amostra de usuários. Os dados capturados foram processados para verificar a possibilidade de identificar um usuário com base em características de uso de aplicações em dispositivos móveis, as quais foram definidas durante este trabalho.

O trabalho está dividido e organizado em seis seções distintas. Dentre estas, a primeira faz uma breve introdução do assunto e contextualiza a proposta. A segunda seção descreve os trabalhos relacionados. A terceira seção descreve a implementação do aplicativo de captura dos dados. A quarta apresenta a metodologia. Na quinta seção serão apresentados os resultados obtidos com a análise dos dados coletados. Por fim, serão expostas as conclusões.

2. Trabalhos Relacionados

Trabalhos anteriores já avaliaram a viabilidade da autenticação ativa considerando diferentes formas de implementação. [Shen et al. 2016] demonstram a autenticação ativa utilizando características comportamentais, considerando as operações em telas sensíveis ao toque. Importante ressaltar que a análise foi realizada utilizando dados de 134.900 operações de toque de 71 participantes em um cenário real, atingindo taxas de erros entre 1,72% e 9,01% para os diferentes tipos de operações de toque. Seguindo a mesma abordagem [Darabseh and Siami Namin 2015] descrevem uma autenticação ativa utilizando o modo de digitação do usuário, avaliando um conjunto de palavras utilizadas com frequência. O experimento foi realizado avaliando 28 usuários e os resultados demonstram boas perspectivas no uso de características de digitação para uma autenticação ativa.

Considerando os trabalhos para criação de *datasets*, podem ser destacadas as pesquisas realizadas por [Eagle and Pentland 2006]. Esta pesquisa, realizada no instituto

de pesquisa de Massachusetts (MIT), envolveu 100 *smartphones* Nokia 6600 e permitiu a coleta de mais de 450 mil horas de informações do comportamento de seus usuários. Apesar deste conjunto de dados ainda ser utilizado para estudo, o MIT também possui um trabalho mais recente, no qual [Fridman et al. 2015] criaram um conjunto de dados a partir de dispositivos móveis, com informações do texto digitado, aplicativos e sites visitados e a localização com base no GPS ou WiFi.

Diferentes formas de autenticação envolvem diferentes algoritmos que auxiliam na tomada de decisão, [Anjum and Ilyas 2013] demonstram que a utilização da árvore de decisão C4.5 é mais eficaz para o reconhecimento de padrões em um conjunto de dados, com traços de uso capturado por sensores de um smartphone. Contudo, outras pesquisas definem diferentes algoritmos, por exemplo, para [Li et al. 2014] a opção de classificação mais eficaz foi a abordagem com base em regras, considerando a relação entre o poder computacional empregado e a acuracidade de sua aplicação.

O objetivo deste projeto está diretamente relacionado com os trabalhos de [Fridman et al. 2015] e [Li et al. 2014], nos quais, o modo de uso nas aplicações de um smartphone é utilizado para uma autenticação ativa. Para definir os melhores resultados diferentes algoritmos devem ser implementados, assim como realizado por [Anjum and Ilyas 2013]. O *dataset* desenvolvido por este projeto foi elaborado com base nas características do trabalho de [Eagle and Pentland 2006], considerando a sua relevância na área. O diferencial desta pesquisa perante os trabalhos relacionados é o direcionamento da análise para características associadas ao comportamento do usuário no manuseio de aplicações, que não foi abordado por nenhum dos artigos pesquisados.

3. Implementação

Para atingir o objetivo proposto neste trabalho foi necessário o desenvolvimento de uma aplicação capaz de capturar informações que permitam analisar o comportamento de um usuário a partir de um dispositivo móvel. Assim como no trabalho de [Fridman et al. 2015] a aplicação foi desenvolvida para o sistema operacional Android, sendo necessário, definir as características de uso, capturar e armazenar os dados e, por fim, extrair os dados para análise.

Com intuito de reduzir as diferenças entre diferentes sistemas operacionais, optou-se pelo desenvolvimento exclusivo para o Android Lollipop. A escolha desta versão de desenvolvimento ocorreu através de uma pesquisa entre possíveis usuários de testes, considerando os seus dispositivos e as disponibilidades para participar da pesquisa. Para criação do aplicativo de captura dos dados foi necessário utilizar um ambiente de desenvolvimento Android. Desta forma foi utilizada a ferramenta Android Studio 1.3.2 em conjunto com um dispositivo móvel Nexus 4, rodando Android 5.1.1. Após a configuração do ambiente de desenvolvimento foram definidas três etapas para criação da aplicação, sendo elas: o processo de captura, o processo de registro e, por fim, a extração dos dados.

Para realizar a captura dos dados foi desenvolvido um serviço que mantém sua execução em segundo plano, de forma a verificar constantemente eventos de mudança nas aplicações. Com o intuito de interferir o mínimo possível no uso padrão do usuário, o serviço de captura não depende de nenhuma ação do usuário após sua inicialização. Para reduzir o consumo de bateria do *smartphone*, foi desenvolvido uma funcionalidade capaz de verificar o estado dos aplicativos somente enquanto a tela do dispositivo estiver

ativa. Considerando outros trabalhos que obtiveram problemas com alto processamento no processo de captura, optou-se por capturar as informações da forma simplificada, sem que houvesse um filtro ou uma etapa de refinamento dos dados capturados. Importante ressaltar que os dados foram armazenados em um banco de dados SQLite, devido a sua simplicidade e o suporte padrão em dispositivos com sistema operacional Android. A última etapa do desenvolvimento da aplicação corresponde a extração dos dados capturados. Deste modo, optou-se por extrair o próprio banco de dados do dispositivo, sem a necessidade de um processo intermediário. Para facilitar essa extração foi criada uma funcionalidade na aplicação capaz de ler o arquivo de banco de dados, realizar uma cópia e envia-la como anexo no serviço de e-mail.

4. Metodologia

O experimento realizado por este trabalho visa realizar uma análise dos dados capturados. Para isso foi necessário selecionar a amostra, processar os dados e, ao final, realizar a análise. Um ponto que deve ser destacado neste trabalho é a realização do experimento em um cenário real, definindo três semanas para coleta dos dados. Considerando a disponibilidade dos usuários, o experimento completo foi executado ao longo de dois meses, permitindo coletar mais de cem mil registros que representam as informações de comportamento dos usuários.

A seleção dos usuários participantes da pesquisa foi realizada com base em um levantamento, considerando o dispositivo utilizado e a disponibilidade para participar do experimento. Os usuários selecionados foram informados quanto ao projeto proposto e os passos necessários para participação da pesquisa. Seguindo as etapas necessárias cada usuário acessou um link que leva ao serviço de armazenamento de arquivos Google Drive contendo a APK (*Android Package*) e um documento do guia de instalação da aplicação. Ao total foram selecionados 20 usuários, entretanto, apenas 15 se disponibilizaram a seguir os procedimentos e manter o serviço de captura em execução durante três semanas. Decorrido o tempo necessário para captura das informações, os usuários enviaram o banco de dados com todos os dados através do serviço de e-mail do dispositivo. Em virtude do tamanho do banco de dados não houve necessidade de acesso físico aos dispositivos, permitindo que o envio ocorresse pela internet. Com base nas informações coletadas, foi possível extrair um total de nove variáveis distintas para representar o perfil do usuário, representadas na Tabela 1.

Tabela 1. Informações que podem ser extraídas a partir dos dados capturados

Nome	Descrição
Processo	Nome do processo que foi executado.
Hora	A hora em que este processo foi executado.
Tempo	Tempo em que o processo permaneceu ativo.
Kilobytes recebidos	Total de kilobytes recebidos enquanto a aplicação está ativa.
Kilobytes transmitidos	Total de kilobytes transmitidos enquanto a aplicação está ativa.
Dia da semana	O número do dia da semana.
Fim de semana	Determina se a instância foi coletada durante o fim de semana.
Tempo de tela	Tempo total em que a tela ficou ativa em relação à hora da amostra.
Contador de apps	Contador do total de apps executados em relação à hora da amostra.

Para realizar a extração das informações apresentadas na Tabela 1, foi necessário realizar uma etapa de processamento dos dados. Esta etapa contou com o desenvolvimento de uma aplicação em linguagem de programação Java, que tem como entrada os arquivos de banco de dados SQLite e como saída arquivos estruturados com formatos CSV (*Comma-separated values*) e ARFF (*Attribute-Relation File Format*). Estes arquivos são gerados para cada conjunto de dados, de forma que o CSV é utilizado para melhor visualização dos dados, enquanto que o ARFF é utilizado pela ferramenta Weka na etapa de análise dos dados.

Visando as necessidades da etapa de análise de dados, a aplicação foi desenvolvida com a finalidade de gerar dois conjuntos de dados, divididos em dados de treinamento e dados de teste. Considerando as três semanas de dados capturados foi realizado uma segmentação em duas partes, onde as duas primeiras semanas foram utilizadas para caracterização do usuário (treinamento) e a terceira semana para aferição da amostra (testes). A Figura 1 demonstra esta divisão, onde 67% dos dados foram utilizados para o treinamento e criação do perfil do usuário e o restante dos dados foram utilizados para extração de amostras de teste. Dado o conjunto de testes criado foi preciso selecionar algumas amostras de cada usuário, para facilitar a análise posteriormente. Visto isso, foi desenvolvida uma funcionalidade capaz de extrair a partir do conjunto de testes, três amostras arbitrárias de cada usuário, contendo uma hora de dados. A segmentação realizada no conjunto de testes foi realizada porque em um cenário real é desejável detectar o uso do dispositivo móvel por um usuário não autorizado, mesmo que por um período pequeno de tempo. Além disso, os autores consideraram as limitações na capacidade de processamento dos dispositivos móveis.

A última etapa do experimento foi a análise dos dados gerados pela etapa de processamento, onde foram selecionadas as amostras de teste correspondente a uma hora de dados, e, consequentemente efetuado a comparação com o arquivo de treinamento. Para isso foi selecionado o módulo Explorer da ferramenta Weka (*Waikato Environment for Knowledge Analysis*), na versão 3.6.2. A seleção desta ferramenta se deve ao fato dela permitir executar algoritmos de aprendizado de máquina para realizar tarefas de mineração de dados.

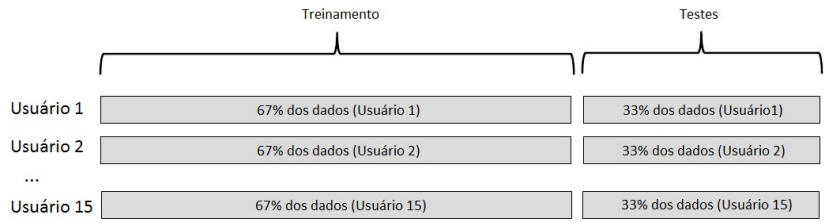

Figura 1. Divisão do conjunto de dados capturado

Sendo assim, foi selecionado o arquivo de treinamento através da interface gráfica da ferramenta, contabilizando 103.918 registros (instâncias) referentes ao conjunto de treinamento. Em seguida foi selecionado o arquivo de testes a ser avaliado. Tendo os conjuntos de treinamento e teste definidos, foi necessário escolher os algoritmos para a classificação dos dados. Com base nas pesquisas de [Li et al. 2014] e

[Anjum and Ilyas 2013] foram selecionados os algoritmos de classificação C4.5 *decison tree*, denominado J48 na ferramenta Weka, e *Naïve Bayes* devido a facilidade de aplicação e suporte padrão pela ferramenta Weka.

Para analisar se as características de uso das aplicações em um dispositivo móvel podem ser utilizadas para autenticar um usuário, o experimento realizou duas análises no conjunto de dados capturado. Na primeira análise, os autores selecionaram três amostras para cada usuário, referentes a uma hora de uso do dispositivo. Neste primeiro momento não foi realizado qualquer filtro em relação à qualidade das amostras selecionadas, se comparadas com o conjunto de treinamento. Já em uma segunda análise, os autores definiram que a seleção das amostras deve ocorrer com base no critério de variação entre o conjunto de treinamento e o número de instâncias na amostra

Os resultados da etapa de análise são obtidos através da ferramenta Weka, que disponibiliza após a realização de cada teste uma taxa de *True Positives* (TP), que corresponde ao percentual da amostra corretamente classificado.

5. Resultados

Os resultados apresentados nesta seção são provenientes dos testes realizados durante o experimento proposto e contabilizam um total de 180 testes. Conforme apresentado, os testes foram realizados em duas etapas, de forma que cada etapa engloba 45 amostras de testes, três por usuário, aplicado a dois algoritmos distintos a fim de extrair as taxas de TP de cada resultado. Importante ressaltar que este trabalho não visa comparar a eficácia entre algoritmos, mas utiliza-los de forma a gerar resultados de classificação.

Como pode ser observado na Tabela 2, os algoritmos *Naïve Bayes* e J48 foram usados. Os conjuntos de testes foram selecionados de forma arbitrária através do próprio sistema de processamento desenvolvido, permitindo que qualquer amostra com ao menos uma instância fosse selecionada.

Nesta etapa, de um total de 90 testes realizados, 15 obtiveram resultado com pontuação superior a 60%. A pontuação para comparação é definido por [Jain et al. 2011] como a medida da similaridade entre o *template* (conjunto de treinamento) e a amostra, sendo que uma pontuação maior indica maior similaridade. Com base nos resultados, foi feita uma análise a fim de interpretar a relação existente entre a taxa de TP encontrada e a variação da amostra, desta forma foi identificado que as amostras com melhores resultados, possuem geralmente uma variação menor.

Para a segunda etapa de testes foi considerada uma variação máxima permitida. Utilizando os recursos disponíveis no sistema de processamento desenvolvido, foi selecionado um novo conjunto de amostras que, individualmente, possuam variação de no máximo 50% em relação ao conjunto de treinamento. Dado as novas amostras selecionadas foi executado os mesmos procedimentos da primeira etapa para realização dos testes, os quais estão ilustrados na Tabela 2.

É perceptível a diferença entre a primeira e a segunda etapa de testes, onde todas as pontuações possuem resultados superiores na segunda etapa, isso demonstra que o uso de critérios para seleção das amostras de teste produz melhores resultados. Podem ser destacadas as amostras com pontuação acima de 80%, que obtiveram uma melhora de 180% em relação à primeira etapa, demonstrando que a utilização de um critério máximo

para a variação da amostra pode melhorar a similaridade entre o conjunto de treinamento e o conjunto de testes.

A partir dos resultados da segunda etapa, é possível afirmar que para utilizar características de uso de aplicações para autenticação ativa, faz-se necessário estabelecer pré-requisitos na seleção dos dados de teste. Os resultados produzidos neste artigo não permitem que os autores estabeleçam todos os pré-requisitos, necessitando novos testes e uma amostra com tamanho superior.

Tabela 2. Comparação entre a primeira e a segunda etapa de testes

Algoritmo	Pontuação			
	60%	70%	80%	90%
Etapa 1				
Naïve Bayes	8	7	5	3
J48	7	7	5	3
Total da Etapa 1	15	14	10	6
Etapa 2				
Naïve Bayes	18	18	17	8
J48	18	15	11	8
Total da Etapa 2	36	33	28	16
Diferença entre etapas	140%	135%	180%	166%

6. Considerações Finais

Este trabalho avaliou se as características de uso de aplicações em dispositivos móveis podem ser utilizadas para autenticação ativa. Para isso, foram coletados os dados, através de uma aplicação desenvolvida pelos autores, e realizado uma análise em duas etapas, demonstrando a necessidade de estabelecer critérios para a seleção de amostra.

De acordo com o experimento realizado, a primeira etapa demonstra que as amostras com variação alta apresentam taxas menores de TP, visto que apenas 15 amostras possuíram taxas superiores a 60%. Já a segunda etapa que faz uso de critérios para a seleção da amostra, apresenta melhores resultados devido à redução da variação em relação ao conjunto de treinamento. Desta forma a segunda etapa apresenta 36 amostras com taxas superiores a 60%.

Importante ressaltar que a utilização de métodos alternativos como forma de identificação do usuário não substitui outros métodos de autenticação tradicional [Clarke 2011]. Entretanto a utilização deste método é possível quando o conjunto de dados de teste atende a determinado critério específico. Nestes casos pode ser aplicada a autenticação ativa com base nas características de uso em dispositivos móveis.

Com base na metodologia utilizada para realizar este trabalho foi possível identificar que apenas usando as características de uso de aplicações em dispositivos móveis, não é possível realizar uma autenticação ativa. Porém, este estudo demonstrou uma melhora relevante ao utilizar diferentes critérios para seleção da amostra. Os autores darão continuidade neste projeto permitindo que sejam identificados melhores critérios, ou ainda, analisando a contribuição de cada uma das características de uso coletadas individualmente. Como trabalho futuro é proposta uma análise mais ampla, aumentando a

quantidade de usuários participantes do projeto e também utilizando diferentes critérios para seleção da amostra, ou ainda, extraindo outras características de uso de aplicações em dispositivos móveis.

Referências

Alzubaidi, A. and Kalita, J. (2015). Authentication of smartphone users using behavioral biometrics. *Journal of IEEE Communications Surveys and Tutorials*.

Anjum, A. and Ilyas, M. U. (2013). Activity recognition using smartphone sensors. In *Consumer Communications and Networking Conference (CCNC), 2013 IEEE*.

Ashbourn, J. (2014). *Biometrics in the New World: The Cloud, Mobile Technology and Pervasive Identity*. Springer Science & Business Media.

Clarke, N. (2011). *Transparent user authentication: biometrics, RFID and behavioural profiling*. Springer Science & Business Media.

Darabseh, A. and Siami Namin, A. (2015). Keystroke active authentications based on most frequently used words. In *Proceedings of the 2015 ACM International Workshop on International Workshop on Security and Privacy Analytics*. ACM.

Eagle, N. and Pentland, A. (2006). Reality mining: sensing complex social systems. *Personal and ubiquitous computing*.

Fridman, L., Weber, S., Greenstadt, R., and Kam, M. (2015). Active Authentication on Mobile Devices via Stylometry, Application Usage, Web Browsing, and GPS Location. *IEEE Systems Journal*.

Jain, A., Ross, A. A., and Nandakumar, K. (2011). *Introduction to biometrics*. Springer Science & Business Media.

Li, F., Clarke, N., Papadaki, M., and Dowland, P. (2014). Active authentication for mobile devices utilising behaviour profiling. *International journal of information security*.

Modi, S. K. (2011). *Biometrics in identity management: Concepts to applications*. Artech House.

Shen, C., Zhang, Y., Guan, X., and Maxion, R. A. (2016). Performance analysis of touch-interaction behavior for active smartphone authentication. *IEEE Transactions on Information Forensics and Security*.