Brazilian Journal of Development

Continuous authentication of users based on network usage behavior at corporate wireless local networks

Autenticação continuada de usuários baseada em comportamento no uso de redes locais corporativas sem fio

DOI:10.34117/bjdv5n7-061

Recebimento dos originais:12/05/2019
Aceitação para publicação: 25/06/2019

Pedro Luiz Teixeira de Moura
Mestre em Engenharia da Computação pelo Instituto de Pesquisas Tecnológicas de São Paulo
Instituição: HCL Technologies
Endereço: Rua Olimpíadas, 205, Conjunto 12 – Vila Olímpia, São Paulo – SP, 04551-000
E-mail: pedro.moura@hcl.com pedro.z.moura@gmail.com

Eduardo Takeo Ueda
Doutor em Engenharia Elétrica pela Escola Politécnica da Universidade de São Paulo Mestre em Ciências da Computação pelo Instituto de Matemática e Estatísticas da Universidade de São Paulo
Instituição: Instituto de Pesquisas Tecnológicas de São Paulo, São Paulo, Brasil
Endereço: Av. Prof. Almeida Prado, 532 - Butantã, São Paulo - SP, 05508-901
E-mail: edutakeo@usp.br

RESUMO

A Autenticação Continuada (AC) de Dispositivos Móveis (DM) pessoais, em redes sem fio locais corporativas, baseia-se em geral em biometria, uso de teclado, toque em tela ou na análise do uso de aplicativos. Isso implica na instalação de um programa de monitoramento no próprio DM, impactando negativamente o sentimento de intrusão por parte do proprietário do DM, e sobrecarregando a equipe de apoio técnico. Além disso, a AC pode ser comprometida caso o aplicativo seja desativado por um intruso ou se o DM for clonado. Com a finalidade de evitar tais problemas, este artigo propõe e avalia, em ambiente de teste, o uso do tráfego de rede para autenticar usuários em seus DMs, aplicando métodos de Aprendizado de Máquina, Supervisionado e Não Supervisionado.

Palavras chave: Autenticação continuada, dispositivos móveis, rede local sem fio, aprendizado de máquina, autenticação baseada em comportamento.

ABSTRACT

The Continuous Authentication (CA), of personal Mobile Devices (MD) to access wireless networks at corporate locations, is based on biometrics or analysis of application usage in such devices. This generally entails the installation of a monitoring program in the MD, negatively impacting the feeling of intrusion on MD owners, as well as burdening the technical support team. In a worst scenario the CA can be compromised in case the application get disabled by

Brazilian Journal of Development

the intruder or if the MD be cloned. Seeking to avoid such problems, this paper proposes and evaluates, in a test environment, the use of network traffic to authenticate users in their DMs, applying Machine Learning, Supervised and Unsupervised methods.

Keywords: Continuous Authentication, machine learning, mobile devices, wireless local area network, behaviour based authentication.

Brazilian Journal of Development

1 INTRODUÇÃO

O uso de dispositivo móvel (DM) como ferramenta de produtividade em ambientes corporativos é ponto comum no atual cenário empresarial, impactando positivamente em agilidade e mobilidade nas operações diárias de um empreendimento. Contudo a mobilidade também demanda um controle maior em termos de autenticação. Para suprir tal necessidade é que se desenvolvem soluções de Autenticação Continuada (AC) que como complemento à autenticação tradicional no ingresso à sistemas e dispositivos, buscam manter a legitimidade do usuário como um processo constante, possibilitando ratificar ou não que, a pessoa em posse de um dispositivo móvel (DM) – um telefone celular "inteligente" (*smartphone*) ou *tablet* – é o usuário legítimo do respectivo aparelho.

Em Deutschmann (2013) e Wheeler (2013), AC possui também outra denominação: Autenticação Ativa (AA). Que apesar de menos usada, traduz o trabalho dinâmico que é realizado, com dados baseados, normalmente, em comportamento no uso de recursos, tais como, teclado e telas sensíveis ao toque (*touchscreen)*, como visto em Bours e Mondal (2013), Frank et al. (2013), Seo et al. (2012) e Wheeler et al. (2013), que instalam programas de monitoramento no DM, para coletar tais dados e, para analisá-los. Tais soluções, lançam mão do uso de técnicas de Aprendizado de Máquina (AM), extraindo padrões e comparando-os com novos conjuntos de dados.

Contudo essa abordagem apresenta quatro desvantagens: a primeira está relacionado à necessidade de se criar uma estrutura de apoio técnico para instalação, desinstalação (em caso de descomissionamento de dispositivo), reinstalação (em caso de troca de dispositivos) e manutenção dos aplicativos de coleta de dados nos DMs – *tablets* e *smartphones* – em uso na *WLAN*, não apenas para garantir a efetiva coleta de dados mas também para evitar que falhas nos programas relacionados possam comprometer o funcionamento dos DMs. A dimensão dessa estrutura, e portanto seu custo, deverão ser ainda maiores caso tais dispositivos não pertençam à empresa, ou seja em iniciativas de BYOD (*Bring Your Own Device*), devido à maior variedade de tipos e modelos de DMs. A segunda desvantagem, devido à abordagem invasiva, é dificultar a adesão de funcionários a programas de BYOD. Como terceira desvantagem, a instalação de tais aplicativos de coleta em equipamentos pessoais, poderia ser utilizada como possível prova de violação de privacidade em eventuais processos trabalhistas contra a empresa. E a quarta, é que a AC pode ser comprometida caso o aplicativo seja desativado pelo usuário ilegítimo em posse do dispositivo ou, em casos em que o DM seja clonado.

Como contraponto Moghaddam e Helmy (2011) trabalharam na determinação de correlações entre grupos de interesse – como por exemplo a maior tendência de usuários Apple acessarem determinados sites de notícias, em relação a usuários Microsoft – analisando apenas o tráfego de rede, sem implicar nenhuma instalação de programas em DM. Entretanto, como essa pesquisa não buscou identificar usuários, permanece, pois, o desafio de extrair, do tráfego de rede, dados que, analisados por técnicas de AM, possam diferenciar usuários, validando seu uso em AC.

Assim este artigo tem como propósito testar o uso de rede como abordagem de AC de usuários em Redes Sem fio Locais. A partir da captura do tráfego de rede, foram selecionadas e testadas características do fluxo gerado pelo usuário legítimo, em seu DM, que foram modeladas por AM Supervisionado e por agrupamentos (*clusters*) extraídos pelo AM Não Supervisionado, determinando assim um padrão legítimo para o DM, e posteriormente testando-o contra outros fluxos legítimos e ilegítimos.

Sem utilizar aplicativos instalados nos DMs e analisando apenas o comportamento do uso de rede, a aplicação desta abordagem de AC, não implica no aumento do sentimento de intrusão da empresa nos equipamentos (DM) de seus funcionários em cenários de BYOD; não cria evidências para processos trabalhistas e não implica na necessidade de criar uma estrutura de apoio técnico em função de aplicativos de monitoramento instalados nos DMs, podendo neste caso servir como alternativa de AC, aos trabalhos relacionados, em ambientes que suportem a diferença de precisão entre tais soluções e esta proposta. E como não é suscetível a ser desabilitada pelo intruso em uso do DM, e mesmo em caso de clonagem de DMs o seu tráfego é capturado, a abordagem desta pesquisa também pode ser aplicada em conjunto com as técnicas tradicionais de AC, como um complemento ou contingência.

Neste artigo são apresentados os trabalhos relacionados, o funcionamento da abordagem proposta, a implementação dos testes, a discussão dos resultados e a conclusão descritos a seguir:

Metodologia: Explanação sobre método escolhido e ambiente de testes.

Trabalhos Relacionados: Análise dos trabalhos relacionados e referencial teórico;

Autenticação Baseada no Uso de Rede: Apresentação do método proposto e sua estrutura para os testes;

Brazilian Journal of Development

Testes, Resultados e Discussão: Execução dos testes e discussão sobre os resultados em comparação aos trabalhos relacionados;

Conclusão: Avaliação dos resultados obtidos e propostas para trabalhos futuros.

2 METODOLOGIA

O método de trabalho escolhido é o experimental, composto da coleta de tráfego de rede como em um cenário de BYOD em uma empresa, uma vez que se trata da situação mais impactada pelos problemas das abordagens estudadas de AC. Do tráfego de rede serão criadas estruturas de dados, baseadas no fluxo gerado a partir do DM, por ser mais determinante de comportamento do que o fluxo recebido por ele.

Tais estruturas foram submetidas a técnicas de AM supervisionado e não supervisionado, para determinar quais algoritmos e estruturas de dados são mais efetivos para AC e também, para gerar modelos de padrão de uso, para classificar os fluxos de rede nas demais fases do trabalho, para fins de AC.

3 TRABALHOS RELACIONADOS

Os trabalhos relacionados, encontrados na literatura, foram agrupados em função da abordagem usada na coleta de dados. Assim foram definidas três categorias, conforme mostrado na Tabela 1.

Brazilian Journal of Development

Tabela 1. Categorias dos trabalhos relacionados

Categoria	Nome da Categoria	Descrição
1ª	Intrusivos no DM de conhecimento do usuário	Com a ciência do usuário, é instalado um programa que monitora o uso do equipamento para coleta de dados
2ª	Intrusivos no DM sem o conhecimento do usuário	Sem a ciência do usuário, é instalado um programa que monitora o uso do equipamento para coleta de dados
3ª	Não intrusivos no DM	Nenhum programa é instalado no dispositivo do usuário para monitorar o comportamento

Na primeira categoria, os trabalhos usam aplicativos instalados no DM para a coleta de dados, sendo que o usuário está ciente de que está sendo monitorado. Com exceção de Bours e Mondal (2013) – que fazem uma proposta de arquitetura – Frank et al. (2013), Seo et al. (2012), Deutschmann e Lindholm, (2013) e Orekondy et al. (2012) apresentam resultados práticos ao autenticar o usuário pelos padrões de digitação, toques em tela sensível, interação com aplicativos ou reconhecimento facial e corporal.

Na segunda categoria (Wheeler et al., 2013), é instalado um *software* no equipamento do usuário (sem seu conhecimento), que simula periodicamente um mau funcionamento de janelas do pacote MICROSOFT OFFICE (MO), assim, a forma como o usuário interage com as "falhas" é usada autenticá-lo.

A terceira categoria (Moghaddam e Helmy, 2011), caracteriza-se pela falta de intervenção no DM, com o objetivo de encontrar tendências de comportamento, em grupos de usuários, em redes móveis, a partir da investigação dos registros da *WLAN* por meio de Aprendizado de Máquina Não Supervisionado, usando a técnica de Mapas Auto-organizados (MAO); sem contudo, ter a autenticação de usuários como objetivo.

As soluções estudadas, relacionam-se com os seguintes tipos de ataques, apresentados na Tabela 2, a serem mitigados pela implementação de AC deste artigo.

Brazilian Journal of Development

Tabela 2. Cenários de ataque

Cenários de ataque
Uso do DM por um usuário ilegítimo, em posse física do dispositivo;
DM invadido remotamente;
DM comprometido por algum *malware*;
DM clonado e/ou ***MAC spoofing***;

Dos possíveis cenários de ataque apresentados na Tabela 2, os trabalhos relacionados, estão aptos a combater somente o *"uso do DM por um usuário ilegítimo, em posse física do dispositivo"*.

Assim, as duas primeiras categorias, em função de seus impactos negativos com relação à invasão de privacidade; necessidade de manutenção e, risco de serem desabilitadas ou ignoradas em clonagens, podem ser beneficiadas pela abordagem deste artigo, como alternativa e/ou complemento. Por sua vez a categoria "Não intrusiva no DM" (Moghaddam e Helmy, 2011), motivou esta pesquisa, na aplicação de Aprendizado Não Supervisionado contra fluxos de rede, apesar de não autenticar usuários, mas por preencher os requisitos em relação à: privacidade no âmbito do DM; não poder ser desabilitada no DM, e por não necessitar de mais recursos (equipe de apoio técnico) de manutenção. Devido aos resultados obtidos por Orekondy et al. (2012) com Máquinas de Vetor de Suporte (MVS) esta técnica também é incluída nos testes deste artigo, mas aqui, aplicada ao tráfego de rede.

4 AUTENTICAÇÃO BASEADA NO USO DE REDE

AC visa autenticar o comportamento do usuário, e no caso deste artigo, é na dinâmica do uso de rede que isto acontece. Não apenas analisando o consumo de rede por usuário e seus destinos em relação ao tempo, entre outras características, mas ao aplicar AM contra tais dados, explorando suas correlações.

Inicialmente foram utilizadas as técnicas de AM Supervisionado com o modelo MVS e Não Supervisionado com Mapas Auto-organizados (MAO), para classificar os padrões de uso de rede de cada usuário com seu respectivo DM. Criou-se, assim, um conjunto de treinamento (*data set*) que representou o uso legítimo de cada par "Usuário"+"DM". Assim, para que um novo fluxo de rede fosse considerado legítimo, ele deveria ter um grau de pertinência

(similaridade) ao comportamento aprendido anteriormente, verificado por Matriz de Confusão (MC) no caso de MVS e combinação com os agrupamentos para o caso de MAO.

4.1 ETAPAS DO MECANISMO DE AUTENTICAÇÃO

Foi criada uma estrutura, simulando um ambiente de trabalho, com rede sem fio local e acesso à Internet. Vinte e um voluntários, usando DMs participaram dos experimentos, realizando tarefas acessando a Internet para assuntos de interesse pessoal – como redes sociais, por exemplo. Para cada dia (tomada) de testes, foram realizadas cinco coletas de tráfego de rede, de acordo com a Tabela 3.

Tabela 3. Descrição das capturas de tráfego de rede

Nome do Teste/Coleta	Descrição
Teste1	Tráfego Legítimo – para Treinamento do modelo de AM
Teste2	Tráfego Legítimo – para Testar o modelo de AM gerado com os dados do Teste1
Teste3	Troca de DM entre usuários de grupos de interesse e gênero diferentes. Tráfego Ilegítimo
Teste4	Troca de DM entre usuários de grupos de interesse comum. Tráfego Ilegítimo
Teste5	Troca aleatória de DM entre os usuários. Tráfego Ilegítimo

O tratamento dos dados durante os testes respeitou as seguintes etapas: coleta e pré-processamento; aplicação de técnicas de AM; trocas de usuários e novos experimentos; e análise dos resultados.

Brazilian Journal of Development

4.2 CONJUNTO DE DADOS

Os cabeçalhos do fluxo de rede, oriundos dos DMs, foram capturados e, juntamente com os dados relativos à Rede *Wireless* sobre associação e desassociação aos Access Point (APs), foram armazenados em arquivos texto. A partir deles, foram extraídos, os seguintes dados, elencados na Tabela 4 a seguir, onde a proposta visou monitorar as relações de fluxo para os destinos mais (d_n) e menos (q_n) frequentes, computando o número de pacotes SYN e RST para tais destinos, bem como o somatório de pacotes com tais flags (SDT e SQT); o somatório de tempo de conexão (Td_n e Tq_n); o tamanho em bytes (Zd_n e Zq_n); a relação de próximo destino ($_nd_id_j$ e $_nq_iq_j$) e os pedidos de associação e desassociação ($_nA$ e $_nD$) aos APs.

Tabela 4. Características do fluxo de rede

Abreviação da Característica	Descrição
d_n	Redes de destino mais comuns (n = 1 a 10)
q_n	Redes de destino menos comuns (n = 1 a 10)
SDT	Total de pacotes *SYN* para os 10 destinos mais comuns na rede
SQT	Total de pacotes *SYN* para os 10 destinos menos comuns na rede
sd_n	quantidade de pacotes *SYN* para d_n (n = 1 a 10)
sq_n	quantidade de pacotes *SYN* para q_n (n = 1 a 10)
Rd_n	quantidade de pacotes *RST* para d_n (n = 1 a 10)
Rq_n	quantidade de pacotes *RST* para q_n (n = 1 a 10)
$_nA$	quantidade de pedidos de associação por AP
$_nD$	quantidade de pedidos de desassociação por AP
$_nd_id_j$	número de vezes que d_j (j = 1 a 10) é o próximo destino a receber um pacote *SYN* depois de d_i (i = 1 a 10)
$_nq_iq_j$	número de vezes que qj (j = 1 a 10) é o próximo destino a receber um pacote *SYN* depois de qi (i = 1 a

Abreviação da Característica	Descrição
	10)
Td_n	Tempo de conexão por dn (n = 1 a 10)
Tq_n	Tempo de conexão por qn (n = 1 a 10)
Zd_n	Tamanho do fluxo em Bytes do DM para o dn (n = 1 a 10)
Zq_n	Tamanho do fluxo em Bytes do DM para o qn (n = 1 a 10)

A definição do número de usuários, DMs, amostras, e tempo de coleta, considerou os valores usados nos trabalhos relacionados, implicando no plano de obter 60 amostras por DM. Dividindo o número de amostras pelos cinco testes previstos, o número de usuários e DMs necessário foi de doze. Tal quantidade é compatível, ainda com outros trabalhos como os de Montalvo Filho et al. (2006) e Muncaster et al. (2006), que testaram com 10 usuários. Com relação ao tempo de coleta, a base de cálculo foram os valores de Frank et al. (2013), com 25 a 50 minutos por amostra e Wheeler et al. (2013), com 120 minutos por tomada, de forma que o número proposto neste artigo foi de 30 minutos para a captura de cada amostra, totalizando 150 minutos por tomada (dia de teste).

4.3 APLICAÇÃO DAS TÉCNICAS DE APRENDIZADO SUPERVISIONADO E NÃO SUPERVISIONADO

Usando os dados coletados, foram aplicadas as técnicas de MVS e MAO, respectivamente, de AM Supervisionado para obter o aprendizado da classificação e Não Supervisionado para obter os agrupamentos. Para MAO foi usado o algoritmo de Kohonen devido aos resultados obtidos por Moghaddam e Helmy (2011), e para MVS foram testados várias possibilidades, sendo selecionados: Otimização Sequencial Mínima (OMS) e Árvore de Decisão J48. Os resultados subsequentes referentes a um DM foram comparados aos do primeiro experimento (Teste 1), visando encontrar graus de pertinência (similaridade) suficientes para diferenciar fluxos legítimos dos ilegítimos. Para analisar o impacto no aumento do tempo de treinamento, foi, ainda, realizada uma nova aplicação de AM contra um fluxo, com tempo de treinamento estendido.

Brazilian Journal of Development

5 TESTES, RESULTADOS E DISCUSSÃO

Os vinte e um voluntários foram organizados dentro de três dias de experimentos, conforme: suas disponibilidades, gênero, e grupos de interesse. Como mostra a Tabela 5, cada voluntário recebeu um codinome, em sequência, de Vol1 a Vol21, um hífen e o código do subgrupo ao qual pertenceu. Por exemplo, o voluntário Vol9-aB2, fez parte do subgrupo B2, relativo a mulheres, casadas e com filhos. O quadro ainda lista qual foi o DM original utilizado por cada voluntário, nos Testes 1 e 2, para a criação do modelo de treinamento, e para a autenticação do usuário legítimo, respectivamente; e, quais foram os DMs utilizados, nos demais Testes, 3, 4 e 5, para a avaliar a falha de autenticação com os fluxos ilegítimos.

Nas colunas relativo as trocas, na Tabela 5, as células destacadas em cinza, indicam que a permuta de DMs, cumpriu a meta esperada, ou seja: a primeira troca, deveria ocorrer entre voluntários de grupo de interesse e gênero diferentes; na segunda entre membros do mesmo grupo de interesse; e na terceira troca aleatória (sem repetição de troca). As células com valores, mas sem o destaque em cinza, indicam que não foi possível cumprir com tais metas, e as células vazias indicam que não houve captura para aquele voluntário no respectivo teste, devido a ausências temporárias dos mesmos.

Tabela 5. Distribuição real de voluntários e DMs nas tomadas

Codinome do Voluntário	DM original (testes 1 e 2)	trocas			Gênero	Grupos de Interesse			Tomada de testes
		Primeira (teste 3)	Segunda (teste 4)	Terceira (teste 5)		(I) Casados Sem Filho	(II) Casados com filhos	(III) Solteiros	
Vol1-aC2	ib2	ab7	ab11	ab9	Feminino			X	Primeira
Vol2-aC1	ab11	ab9	tb4	ib3	Masculino			X	Primeira
Vol3-bC1	tb4	ib3	ib2	ab11	Masculino			X	Primeira
Vol4-aA1	ab7	ib2	ib3	tb4	Masculino	X			Primeira
Vol5-aA2	ib3	tb4	ab9	ab7	Feminino	X			Primeira
Vol6-bA2	ab9	ab11	ab7	ib2	Feminino	X			Primeira
Vol7-aC2	ab62	ib51	ab61	tb41	Feminino			X	Segunda
Vol8-aC1	ab61	tb41	ab62	ib51	Masculino			X	Segunda
Vol9-aB2	tb41	ab61	ib51	ab61	Feminino		X		Segunda
Vol10-aB1	ib51	ab62	tb41	ab62	Masculino		X		Segunda
Vol11-aC2	tb1	ab15	tb6	ab8	Feminino			X	Terceira
Vol12-aA2	ib9	tb12	ab5		Feminino	X			Terceira
Vol13-aC1	tb12	ab5	tb14		Masculino			X	Terceira
Vol14-aB2	ab5	tb13	ab10	tb1	Feminino		X		Terceira
Vol15-bC2	tb6	ab8	tb1	tb13	Feminino			X	Terceira
Vol16-aA1	ab15	ib9			Masculino	X			Terceira
Vol17-aB1	ab8	tb1	ab14	tb6	Masculino		X		Terceira
Vol18-bB2	ab10	tb14	tb13	ab14	Feminino		X		Terceira
Vol19-bC1	tb13	ab10	ab15	tb14	Masculino			X	Terceira
Vol20-cC2	tb14	ab14	ib9	ab5	Feminino			X	Terceira
Vol21-bB1	ab14	tb6	ab8	ab10	Masculino		X		Terceira

Na Tabela 6, é mostrado a comparação da proposta implementada neste artigo com os trabalhos relacionados, com relação ao número de usuários, DMs, amostras por DM e o tempo de coleta.

Brazilian Journal of Development

Tabela 6. Comparação do número de usuários, amostras e tempo de coleta de dados

Trabalho	Número de usuários	Número de dispositivos	Número de amostras	Tempo de coleta
(Frank et al., 2013)	41	4	41	25 a 50 minutos por amostra
(Seo et al., 2012)	50	–	50 amostras	–
(Deutschmann e Lindholm, 2013)	99	99	99	200 horas
(Orekondy et al., 2012)	2	2	2	–
[Wheeler et al. 2013]	62	-	40 por usuário	120 minutos por coleta (2 coletas)
Este artigo	3 tomadas com 6, 4 e 11 usuários respectivamente	6 na tomada 4 na tomada II 11 na tomada III	30 na tomada 20 na tomada II 51 na tomada III	30 minutos por amostra, totalizando 150 minutos por tomada

5.1 REALIZAÇÃO DAS CAPTURAS DE TRÁFEGO

Em cada uma das três tomadas (dias de testes), foram realizadas 5 capturas de tráfego de rede, nomeadas como Testes 1, 2, 3, 4 e 5, como visto na Tabela 3. Cada teste teve a duração aproximada de 40 minutos, desconsiderados os 10 minutos finais de cada coleta, devido à dispersão dos participantes, ficando assim 30 minutos efetivos de tráfego por coleta, e um intervalo entre si de 20 minutos para gravação dos arquivos de tráfego, ajustes no ambiente de

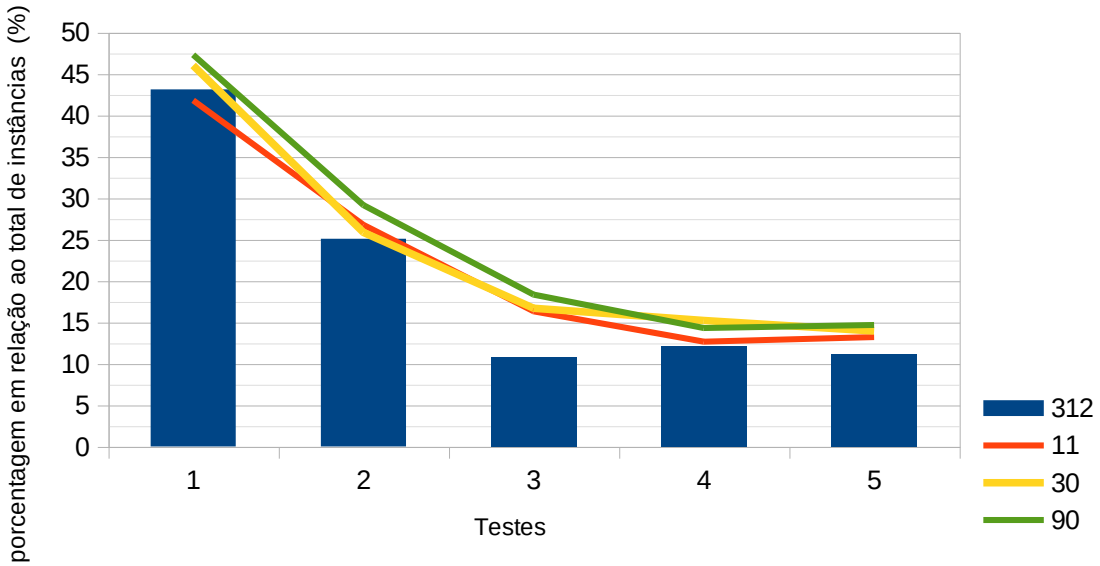
captura e para a troca de DMs entre os participantes, totalizando 5 horas por Tomada. Desta forma, o volume total de tráfego de rede coletado foi de 10,23 GB.

5.2 APLICAÇÃO DE APRENDIZADO DE MÁQUINA SUPERVISIONADO

Na média, os resultados dos Testes 1 e 2, foram mais precisos em relação aos testes ilegítimos, com diferença mais expressiva na Tomada III, cujos resultados são exibidos nas Figura 1 e 2, justamente onde o número de participantes foi maior, ou seja, 11 (5 a mais do que na Tomada I e 7 mais do que na Tomada II). O gráfico ainda evidencia a importância da análise com diferentes números de atributos selecionados a partir da Tabela 4, ou seja, 312, 11, 30 e 90 atributos, servindo de comprovação da tendência de queda de precisão dos fluxos ilegítimos, quando comparados com os legítimos.

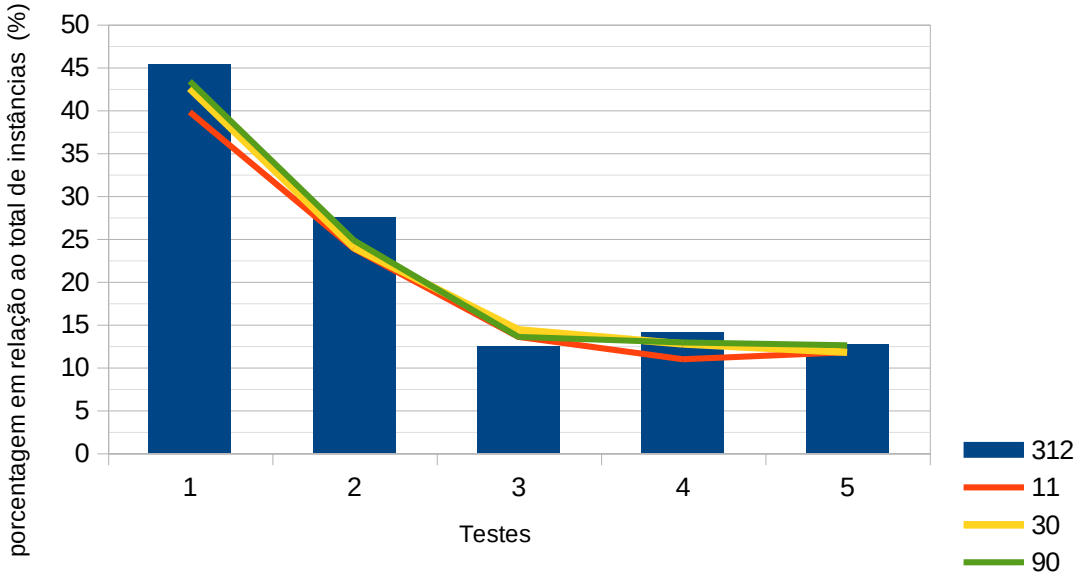
Brazilian Journal of Development

Figura 1. Resultados médios da Tomada III usando OMS



A Figura 2, exibe os resultados da Tomada III, com a aplicação do algoritmo J48, onde de forma similar aos testes com OMS, os Testes 1 e 2, foram mais precisos em relação aos testes ilegítimos.

Figura 2. Resultados médios da Tomada III usando J48



Brazilian Journal of Development

Na análise individual de cada DM, em AM Supervisionado, foi considerado que o dispositivo foi autenticado, se, o Teste 2 alcançou melhores resultados, dos que os do teste ilegítimo, em pelo menos 3 das 4 possíveis análises em função da quantidade de atributos. E assim, estendendo tal modelo para a comparação do Teste 2 com todos os testes ilegítimos, da mesma tomada, considerou-se que o DM foi: **a) corretamente autenticado**, quando o Teste 2 foi mais preciso que os 3 testes ilegítimos; **b) parcialmente autenticado**, quando o Teste 2 foi mais preciso que dois testes ilegítimos; **c) fracamente autenticado**, quando o Teste 2 foi superior somente a um dos testes ilegítimos; e, finalmente, **d) não autenticado**, quando todos os testes ilegítimos foram superiores ao Teste 2. A Tabela 7 mostra os resultados de autenticação por Tomada. Sendo a terceira a mais bem sucedida com 6 DMs corretamente autenticados.

Tabela 7. Resultados da aplicação de AM supervisionado

Tomada	Algoritmo	Quantidade de DMs que tiveram Teste 2 mais preciso que os testes ilegítimos				Total de DMs na Tomada
		3 de 3 (100%)	2 de 3 (67%)	1 de 3 (33%)	0 de 3 (0%)	
Tomada I	OMS	3	1	1	1	6
Tomada I	J48	2	1	1	2	6
Tomada II	OMS	1	2	1	0	4
Tomada II	J48	2	1	1	0	4
Tomada III	OMS	6	1	0	4	11

Tomada	Algoritmo	Quantidade de DMs que tiveram Teste 2 mais preciso que os testes ilegítimos				Total de DMs na Tomada
		3 de 3 (100%)	2 de 3 (67%)	1 de 3 (33%)	0 de 3 (0%)	
Tomada III	J48	6	0	0	5	11

Para analisar o impacto do acréscimo de uma hora no tempo de treinamento, foram extraídos, dos Testes 3 e 5 da Tomada I, os fluxos correspondentes a cada usuário legítimo do Teste 1. Por exemplo, Vol2-aC1 foi o usuário legítimo no Teste 1 com o DM ab11, e nos Testes 3 e 5 ele usou os DMs ab9 e ib3, respectivamente. Assim os arquivos dos testes ilegítimos, sofreram alterações de ab9 e ib3 para ab11, para estender o tempo de teste. O novo conjunto de treinamento estendido foi testado com fluxos dos Testes 2 (legítimo) e 4 (ilegítimo), da mesma tomada. A Tabela 8 mostra a comparação de quantidade de autenticações obtidas (i.e. quando o resultado do Teste 2 foi superior ao do Teste 4), em relação aos testes com tempo normal e estendido.

Brazilian Journal of Development

Tabela 8. Comparação dos resultados com tempo de treinamento normal e estendido

Resultados de autenticação com tempo normal e aplicação de OMS				
Quantidade de atributos	312	7	30	90
Quantidade de DMs Corretamente autenticados	3	3	2	3
Total de DMs do teste	6	6	6	6
Porcentagem de autenticação	50,00	50,00	33,33	50,00
Resultados de autenticação com tempo normal e aplicação de J48				
Quantidade de atributos	312	7	30	90
Quantidade de DMs Corretamente autenticados	3	2	2	2
Total de DMs do teste	6	6	6	6
Porcentagem de autenticação	50,00	33,33	33,33	33,33
Resultados de autenticação com tempo estendido e aplicação de OMS				
Quantidade de atributos	312	7	30	90
Quantidade de DMs Corretamente autenticados	6	4	3	4
Total de DMs do teste	6	6	6	6
Porcentagem de autenticação	100,00	66,67	50,00	66,67
Resultados de autenticação com tempo estendido e aplicação de J48				
Quantidade de atributos	312	7	30	90
Quantidade de DMs Corretamente autenticados	6	3	3	3
Total de DMs do teste	6	6	6	6
Porcentagem de autenticação	100,00	50,00	50,00	50,00

Com o treinamento (Teste1) estendido em uma hora, o Teste 2 (legítimo), quando comparado ao Teste 4, foi mais preciso em relação ao resultados obtidos com o tempo de treinamento normal, chegando a autenticar 100% dos DMs com 312 atributos.

5.3 APLICAÇÃO DE APRENDIZADO DE MÁQUINA NÃO SUPERVISIONADO

A técnica de AM Não Supervisionada, com MAO, foi aplicada separadamente contra o fluxo do Teste 1 (treinamento) de cada DM, não apenas determinando os *clusters*, mas distribuindo as instâncias do treinamento entre eles, de forma que, ao aplicar, posteriormente a este modelo, os demais fluxos (Testes 2, 3, 4 e 5), buscou-se o quanto tais fluxos se aproximaram dos resultados do Teste 1.

Foram selecionados 11 atributos com valores na ordem de dezena, para evitar uma dispersão nos pontos e, assim, não gerar agrupamentos sem aderência. Desta forma os atributos selecionados foram: INTERVALO, SDT, SQT, RDT, RQT, ZDT, ZQT, TDT, TQT, FDT e FQT, conforme indicado na Tabela 4. Durante os testes observou-se que, com quatro agrupamentos foi obtida uma melhor distribuição das instâncias.

Brazilian Journal of Development

Somente a Tomada III, obteve resultados superiores para o Teste 2 em relação aos ilegítimos. Na Tabela 9, são apresentados os valores desta tomada, onde a cor cinza claro indica que o valor da célula é o mais próximo do respectivo valor do Teste 1 para o agrupamento, apontando melhor aderência ao modelo; e a cor cinza escuro indica que o valor da célula é o mais distante do valor do Teste 1 para o agrupamento.

Como visto na Tabela 9, o Teste 2 na Tomada, obteve melhor resultado em 50% dos casos, estando aproximadamente a 18 pontos percentuais acima do segundo colocado (Teste 4), e o Teste 2 é ainda o que possui a menor ocorrência de ser o mais distante estando a 4 pontos percentuais do penúltimo (Teste4).

Brazilian Journal of Development

Tabela 9. Aplicação de MAO contra a Tomada III

DM	cluster	teste1*	teste2	teste3	teste4	teste5
tb1	*0*	45	39	40	35	40
tb1	*1*	0	26	29	27	24
tb1	*2*	23	11	8	8	10
tb1	*3*	32	24	23	29	26
ib9	0	41	15	0	3	0
ib9	1	5	44	48	52	52
ib9	2	5	35	52	42	48
ib9	3	50	6	0	3	0
tb12	*0*	50	34	26	26	26
tb12	*1*	5	6	3	0	0
tb12	*2*	45	24	40	44	44
tb12	*3*	0	35	31	31	31
ab5	0	14	37	37	37	34
ab5	1	45	0	0	5	10
ab5	2	0	39	31	24	27
ab5	3	41	34	32	34	29
tb6	*0*	41	44	45	45	44
tb6	*1*	27	0	0	0	0
tb6	*2*	18	48	50	55	50
tb6	*3*	14	8	5	0	6
ab15	0	50	6	2	0	0
ab15	1	27	19	0	11	0
ab15	2	23	35	44	42	44
ab15	3	0	39	55	47	56
ab8	0	23	3	6	5	3
ab8	1	36	32	27	34	31
ab8	2	5	34	34	34	32
ab8	3	36	31	32	27	34
ab10	0	41	37	34	40	23
ab10	1	5	44	48	44	48
ab10	2	55	8	11	6	13
ab10	3	0	11	6	10	16
tb13	0	0	31	3	23	19
tb13	1	27	23	29	26	31
tb13	2	36	23	35	29	32
tb13	3	36	24	32	23	18
tb14	0	9	39	0	3	29
tb14	1	36	31	40	37	29
tb14	2	32	16	31	31	18
tb14	3	23	15	29	29	24
ab14	0	9	48	52	52	52
ab14	1	32	5	0	0	0

Brazilian Journal of Development

Na Tabela 10, os resultados deste artigo são comparados com os dos trabalhos relacionados, em relação a precisão de autenticação, se é passível de ser desabilitado no DM, se consegue continuar analisando a autenticação mesmo se o dispositivo for clonado, invadido ou contaminado, e qual tipo de distinção é lograda.

Brazilian Journal of Development

Tabela 10. Comparação com os trabalhos relacionados

Trabalho	Precisão	Passível de ser desabilitado no DM	Continua autenticando o mesmo se o DM for clonado, invadido remotamente, ou contaminado por *malware*	Intrusivo no DM com ciência do usuário	Intrusivo no DM sem a ciência do usuário	Tipo de dados coletados	Nível de precisão distinção (ex: usuário ou grupo de interesse comum)
Bours e Mondal (2013)	_	Sim	Não	X	_	Biometria Comportamental (teclado /mouse/ aplicação)	Usuário, mas sem resultados (teórico)
Frank et al. (2013)	Taxa de erro entre 0 e 4%	Sim	Não	X	_	Biometria Comportamental (*touchscreen*)	Usuário
Seo et al. (2012)	~ 100%	Sim	Não	X	_	Biometria Comportamental	Usuário

Trabalho	Precisão	Passível de ser desabilitado no DM	Continua autenticando o mesmo se o DM for clonado, invadido remotamente, ou contaminado por *malware*	Intrusivo no DM com ciência do usuário	Intrusivo no DM sem a ciência do usuário	Tipo de dados coletados	Nível de precisão distinção (ex: usuário ou grupo de interesse comum)
						(*touchscreen*)	
Deutschmann e Lindholm (2013)	FR/u/dia RP/u/2.4 min**	Sim	Não	X	_	Biometria Comportamental (teclado/mouse)	Usuário
Orekondy et al. (2012)	60 a 80%	Sim	Não	X	_	Biometria Facial e Corporal	Usuário
Wheeler et al. (2013)	+ 80%	Sim	Não	_	X	Biometria Comportamental Aplicação	Usuário

Trabalho	Precisão	Passível de ser desabilitado no DM	Continua autenticando o mesmo se o DM for clonado, invadido remotamente, ou contaminado por *malware*	Intrusivo no DM com ciência do usuário	Intrusivo no DM sem a ciência do usuário	Tipo de dados coletados	Nível de precisão distinção (ex: usuário ou grupo de interesse comum)
[Moghaddam and Helmy 2011]	_	Não	_	_	_	Comporta-mento na Rede	Grupo de interesse comum
Este artigo	**54,45 % (tempo normal)**	**Não**	**Sim**	**_**	**_**	**Comporta-mento na Rede**	**Usuário**

A Tabela 10 denota vantagem deste artigo em relação à: não intrusão nos DMs; inviabilidade de ser desabilitada no DM; ser capaz de detectar a ilegitimidade do tráfego de um DM clonado, caso o seu comportamento seja diferente do original; e a efetividade de autenticar usuários e não somente tendências de grupo em redes.

6 CONCLUSÃO

Em Aprendizado de Máquina Supervisionado, os resultados médios, foram mais precisos para o Teste 2 (legítimo), em comparação com todos os ilegítimos. Já na análise individual, a quantidade máxima de DMs autenticados, foi de 54,45%. Contudo, ao ampliar o tempo de treinamento em uma hora, os resultados do Teste 2 (legítimo) em relação ao 4

(ilegítimo), ficaram mais precisos em 50 e 16,67 pontos percentuais, para as respectivas quantidades de atributos, 312, 7, 30 e 90 (estes três últimos com o mesmo valor de acréscimo). Com relação aos grupos de interesse, em média, o Teste 3 obteve resultados inferiores aos do Teste 4, validando a hipótese que o grupo de interesse influencia no tráfego de rede. Já com a técnica de Mapas Auto-organizados, em Aprendizado de Máquina Não Supervisionado, os resultados relevantes para a autenticação ocorreram apenas na Tomada III, chegando a autenticar 50% dos DMs.

Assim, abordagem proposta neste artigo contribui como: **a) alternativa de AC**, em referência aos trabalhos relacionados, em ambientes onde a menor precisão, seja compensada pelo fato de: não aumentar o sentimento de intrusão nos DMs; não criar evidências para processos trabalhistas; e não requerer apoio técnico para aplicativo de monitoramento instalado nos DMs; **b) complemento ou contingência de AC**, para os trabalhos relacionados, em caso de falha, ou comprometimento do monitoramento por parte de um invasor em posse do DM.

Para trabalhos futuros, pretende-se: **a)** aplicar a mesma técnica para o fluxo de entrada em Servidores, como uma alternativa sobre alertas para ataques novos (*Zero day*); **b)** utilizar a mesma abordagem em redes cabeadas para a proteção de dispositivos fixos, usando computadores *desktop* para os testes e alternando usuários no uso deles; **c)** testar tal abordagem de AC em ambientes de comunicação *device-to-device*, visto que nestes casos não há como aplicar AC baseado em ações contra *touch-screen* e/ou teclado virtual, para dispositivos que não possuem humanos interagindo com eles, mas que de qualquer forma são passíveis de invasões via rede; **d)** avaliar a abordagem de AC pelo uso de rede para analisar a eficiência em ataques de *tailgate*, ou seja, verificar em quanto tempo pode ser detectado a troca do usuário legítimo com o impostor, de forma similar ao que Orekondy et al. (2012), realizaram com imagens.

REFERÊNCIAS

A. L. Samuel. Some Studies in Machine Learning Using the Game of Checkers. IBM Journal of Research and Development. Year: 1959, Volume: 3, Issue: 3. Pages: 210 – 229, DOI: 10.1147/rd.33.0210. IBM Journals & Magazines.

Ahmed, A. System and method for determining a computer user profile from a motion-based input device, 5 out. 2006. Disponível em: http://www.google.com/patents/US20060224898. Acessado em: 18/12/2017.

Apache Mahout: Scalable machine learning and data mining. Disponível em https://mahout.apache.org. Acessado em: 01/09/2017.

Bours, Patrick ; Mondal, Soumik. Continuous Authentication using Behavioural Biometrics SIMEM's 2013 JED, 2013. Disponível em: http://jedsimem2013.sciencesconf.org/conference/jedsimem2013/pages/Book_of_abstract_JED2013.pdf#page=13. Acessado em: 07/06/2017.

Class Balancer. Disponível em: http://weka.sourceforge.net/doc.dev/weka/filters/supervised/instance/ClassBalancer.html. Acessado em: 15/10/2017.

Class ConditionalProbatilities. Disponível em: http://weka.sourceforge.net/doc.dev/weka/filters/supervised/attribute/ClassConditionalProbabilities.html. Acessado em: 15/10/2017.

Class J48,weka.classifiers.trees. Disponível em: http://weka.sourceforge.net/doc.dev/weka/classifiers/trees/J48.html. Acessado em: 15/10/2017.

Class Puk. Disponível em: http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/supportVector/Puk.html. Acessado em: 15/10/2017.

Class SMO. Disponível em: http://weka.sourceforge.net/doc.dev/weka/classifiers/functions/SMO.html. Acessado em: 15/10/2017.

Class ZeroR. http://weka.sourceforge.net/doc.dev/weka/classifiers/rules/ZeroR.html. Acessado em: 01/10/2017.

Deutschmann, I.; Lindholm, J. (2013). Behavioral biometrics for DARPA's Active Authentication program. Disponível em: http://cs.emis.de/LNI/Proceedings/Proceedings212/225.pdf. Acessado em: 25/02/2017.

Frank, M. et al. Touchalytics: On the Applicability of Touchscreen Input as a Behavioral Biometric for Continuous Authentication. IEEE Transactions on Information Forensics and Security, v. 8, n. 1, p. 136–148, 1 jan. 2013.

Ghosh, S.; Dubey, S. K. (2013). Comparative Analysis of K-Means and Fuzzy C-Means Algorithms. Ijacsa, 4(4), 35–39. doi:10.14569/IJACSA.2013.040406. Disponível em: http://thesai.org/downloads/volume4no4/paper_6-comparative_analysis_of_k-means_and_fuzzy_c_means_algorithms.pdf. Acessado em: 30/04/2017.

GNU Octave. Disponível em: https://www.gnu.org/software/octave/. Acessado em: 01/10/2017.

GNU General Public License. Disponível em: https://www.gnu.org/licenses/gpl-3.0.en.html. Acessado em: 15/10/2017.

Kashi, R. Habit-based authentication, 8 maio 2012. Disponível em: http://www.google.com/patents/US8176159. Acessado em: 12/10/2017.

KNIME | KNIME Analytics Platform. Disponível em: https://www.knime.org/knime. Acessado em: 12/10/2017.

Kohavi, Ron And Provost, Foster. Glossary of Terms. Special Issue on Applications of Machine Learning and the Knowledge Discovery Process. ©1998 Kluwer Academic Publishers, Boston, Manufactured in The Netherlands. Disponível em: http://robotics.stanford.edu/~ronnyk/glossary.html. Acessado em: 08/08/2017.

Lantz, Brett 2013. Machine Learning with R. Packt Publishing. ISBN:1782162143 9781782162148.

Mitchell, Thomas M. Machine Learning McGraw-Hill, Inc. New York, NY, USA. 1997. ISBN:0070428077 9780070428072.

Moghaddam, S.; Helmy, A. Multidimensional modeling and analysis of wireless users online activity and mobility: a neural-networks map approach. Proceedings of the 14th ACM international (2011).

Montalvao Filho, J.R. and Freire, E.O. (2006), "Multimodal biometric fusion - joint typist (keystroke) and speaker verification", Proceedings of the International Telecommunications Symposium , pp. 609-614.

Muncaster, J. and Turk, M. (2006), "Continuous Multimodal Authentication Using Dynamic Bayesian Networks", Proceedings of the 2nd Workshop on Multimodal User Authentication, pp. 1-8.

OpenCV. Disponível em: http://opencv.org. Acesso em: 15 de Novembro 2017.

OREKONDY, Tribhuvanesh, et al. (2012). Application of Support Vector Machine in Continuos Authentication. In Information and Communication Technologies (WICT), 2012 World Congress on (pp. 608–613).

Partition membership. Disponível em: http://weka.sourceforge.net/doc.dev/weka/filters/supervised/attribute/PartitionMembership.html. Acessado em: 15/10/2017.

Quinlan, J. R. "Induction of decision trees", Machine Learning Journal., vol. 1, no. 1, pp.81 - 106 1986. Kluwer Academic Publishers. DOI 10.1007/BF00116251. Disponível em: http://link.springer.com/content/pdf/10.1007%2FBF00116251.pdf. Acessado em: 11/08/2017.

R: The R Project for Statistical Computing. Disponível em http://www.r-project.org. Acessado em: 02/10/2017.

Rabiner, L., "A tutorial on hidden Markov models and selected applications in speech recognition," Proceedings of the IEEE , vol.77, no.2, pp.257,286, Feb 1989. doi: 10.1109/5.18626. Disponível em: http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=18626&isnumber=698. Acessado em: 07/06/2017.

Rennie, J. D. M. et al. (2003). Tackling the Poor Assumptions of Naive Bayes Text Classifiers. Proceedings of the Twentieth International Conference on Machine Learning (ICML)-2003), 616–623. doi:10.1186/1477-3155-8-16. Disponível em: https://people.csail.mit.edu/jrennie/papers/icml03-nb.pdf. Acessado em: 28/01/2018.

Resample. Disponível em: http://weka.sourceforge.net/doc.dev/weka/filters/supervised/instance/Resample.html. Acessado em: 07/06/2017.

Salatas, John. SelfOrganizingMap: Cluster data using the Kohonen's Self-Organizing Map algorithm. Disponível em: http://weka.sourceforge.net/packageMetaData/SelfOrganizingMap/index.html. Acessado em: 01/06/2017.

Seo, Hojin et al . A Novel Biometric Identification Based on a User's Input Pattern Analysis for Intelligent Mobile Devices. 2012. doi:10.5772/51319. Disponível em: http://cdn.intechopen.com/pdfs-wm/38080.pdf. Acessado em: 04/03/2017.

Wang, Y.; Wei, J.; Vangury, K. Bring your own device security issues and challenges, 2014 IEEE 11th Consumer Communications and Networking Conference (CCNC). Anais...IEEE, jan. 2014. Disponível em: http://ieeexplore.ieee.org/articleDetails.jsp?arnumber=6866552. Acessado em: 05/06/2017.

Weka 3 – Data Mining with Open Source Machine Learning Software in Java. Disponível em: http://www.cs.waikato.ac.nz/ml/weka. Acessado em: 10/02/2017.

Wheeler, J. Et Al. ACTIVE AUTHENTICATION USING COVERT COGNITIVE INTERROGATION GAMES. Disponível em: http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA581055. Acessado em: 12/10/2017.

Wu, Junjie. 2012. Advances in K-Means Clustering: A Data Mining Thinking. Springer Publishing Company, Incorporated. ISBN:3642298060 9783642298066.