**Instituto de Pesquisas Tecnológicas do Estado de São Paulo**

**Raphael Hungaro Moretti**

# Revogação de Acesso por meio de Análise Comportamental Utilizando o Padrão 802.1X

**São Paulo**

**2019**

Raphael Hungaro Moretti

Revogação de Acesso por meio de Análise Comportamental utilizando o Padrão 802.1X

Dissertação de Mestrado apresentada ao instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT, como parte dos requisitos para a obtenção do título de mestre em Engenharia da Computação.

Área de Concentração: Infraestrutura Computacional.

Data da aprovação ____/_____/_____

_____

Prof. Dr. Anderson Aparecido Alves da Silva (Orientador)

Mestrado em Engenharia de Computação

Membros da Banca Examinadora:

Prof Dr. Anderson Aparecido Alves da Silva (Orientador)
Mestrado Engenharia de Computação

Prof Dr. Fábio Dacêncio Pereira (Membro)
UNIVEM – Centro Universitário Eurípides de Marília

Prof Dr. Eduardo Takeo Ueda (Membro)
Mestrado Engenharia de Computação

Raphael Hungaro Moretti

# Revogação de Acesso por meio de Análise Comportamental utilizando o Padrão 802.1X

Exame de Defesa apresentado ao instituto de Pesquisas Tecnológicas do Estado de São Paulo – IPT, como parte dos requisitos para a obtenção do título de mestre em Engenharia da Computação.

Área de Concentração: Infraestrutura Computacional.

Orientador: Prof. Dr. Anderson Aparecido Alves da Silva
Co-orientador: Prof. Dr. Adilson Eduardo Guelfi.

**São Paulo**
**Março/2019**

Ficha Catalográfica
Elaborada pelo Departamento de Acervo e Informação Tecnológica – DAIT
do Instituto de Pesquisas Tecnológicas do Estado de São Paulo - IPT

M845r **Moretti, Raphael Hungaro**
Revogação de acesso por meio de análise comportamental utilizando o Padrão 802.1X. / Raphael Hungaro Moretti. São Paulo, 2019.
74p.

Dissertação (Mestrado em Engenharia de Computação) - Instituto de Pesquisas Tecnológicas do Estado de São Paulo. Área de concentração: Infraestrutura Computacional.

Orientador: Prof. Dr. Anderson Aparecido Alves da Silva
Coorientador: Prof. Dr. Adilson Eduardo Guelfi.

1. Análise comportamental 2. Padrão de comportamento de acesso 3. Internet (redes de computadores) 4. Tese I. Silva, Anderson Aparecido Alves da , orient. II. Guelfi, Adilson Eduardo, coorient. III. IPT. Coordenadoria de Ensino Tecnológico IV. Título

19-25 CDU 004.738.5(043)

Bibliotecário responsável: Maria Darci Cornellas Narciso – CRB 8/3569

**Agradecimentos**

*Agradeço a Deus pela grande força e a possibilidade de realizar este trabalho.*

*Agradeço a minha família, especialmente a minha mãe Roseli Hungaro, meu pai Ricardo Moretti e minha irmã Rebecca Hungaro Moretti pela paciência, apoio e compreensão nesse delicado momento que é a realização de um mestrado.*

*Agradeço ao recente companheiro canino Bola, que me diverte nos intervalos de dedicação a este trabalho e me faz companhia no momento de sua elaboração.*

*Agradeço ao gerente do setor de redes da CETESB Sr. Emerson Eikiti Matsukawa pela compreensão e grande apoio na elaboração deste trabalho, assim como todo corpo profissional envolvido da Companhia e a própria Companhia em si, pelo incentivo e aderência ao seu programa de formação profissional.*

*A todos meus amigos que me apoiaram neste processo, especialmente ao Leandro Conte, Diogo Carvalho e Paulo Testa.*

*Ao meu orientador Dr. Anderson Aparecido Alves da Silva, que me auxiliou e orientou na elaboração deste trabalho, me mostrando os caminhos acadêmicos da área de TI.*

*Ao querido e eterno "Mestre" Dr. Sandro Melo, uma das minhas referências profissionais que me indicou este curso e me introduziu ao caminho profissional e acadêmico.*

**Resumo**

Em redes onde é desejado um nível adequado de segurança, constantes monitoramentos e análises são chaves principais para detecção de anomalias. Verificações do meio físico, assinaturas pré-definidas e análises comportamentais são formas para a obtenção de informações para detecção de anomalias e possíveis intrusos que possam prejudicar ou até mesmo furtar informações sensíveis do perímetro da rede. Além da detecção de intrusos, é também importante o isolamento e, quando possível, a remoção destes elementos nocivos. Para isso, soluções que garantem o controle de acesso são também necessárias. O padrão 802.1X, utilizado para controle de acesso à rede, é um grande aliado na análise de problemas de segurança da informação. Sua forma de autenticação provê acesso seguro e auditável em redes cabeadas e sem fio. Seu papel é permitir ou negar acesso a determinados *hosts* de uma rede. Outro grande aliado no auxílio de segurança de uma rede é o Sistema de Detecção e Prevenção de Intrusões (SDPI), responsável por detectar comportamentos anômalos na rede e alertar o administrador sobre possíveis ações maliciosas. Apesar de serem elementos de grande importância no âmbito de segurança, o SDPI e o padrão 802.1X trabalham separadamente, cada um exercendo uma função. Por outro lado, é possível relacionar perfis e padrões de comportamento de usuários com a análise de tráfegos oriundos de servidores *proxy,* auxiliando na identificação de padrões de comportamento similar a ação do SDPI. Ainda pode-se contar também com dados de associação e desassociação de Access Points (AP) para definir padrões de comportamento dos usuários relativo a seus respectivos deslocamentos. Este trabalho busca identificar padrões anômalos de comportamento de usuários por meio da análise comportamental de uma rede, a fim de autenticar usuários, cujo acesso possa ser revogado em tempo real com o uso do padrão 802.1X em conjunto com uma proposta de lógica de solução aliado aos dados de captura de *proxy* e associações e desassociações de Access Points, similar a um SDPI. Ao final, verifica-se que há diferenças entre os perfis de funcionários e a identificação para autenticidade, é possível ser visualizada com as informações capturadas.

**Palavras-chave:** *Perfil Comportamental, anomalias, 802.1X, Análise comportamental por meio de Proxy, Análise do perímetro de Rede.*

**Abstract**

**Access Revocation by behavioral analysis using the 802.1X Standard**

In networks where an adequate level of security is desired, constant monitoring and analysis are important keys to detect anomalies. Physical checks, pre-defined signatures, and behavioral analysis are some ways to obtain information for detecting anomalies and possible intruders that could harm or even steal sensitive information from the network perimeter. In addition to intruder detection, it is also important to isolate and, when it is possible, remove these harmful elements. To do this, solutions that guarantee access control are also necessary. The 802.1X standard, used for network access control, is a great ally when analyzing information security issues. The authentication's form provides secure and auditable access to wired and wireless networks. Your function is to allow or deny access to certain hosts on a network. Another great ally in the security assistance of a network is the Intrusion Detection and Prevention System (IDPS), responsible for detecting anomalous behavior on the network and alerting the administrator to possible malicious actions. Although they are important elements in the security environment, the SDPI and the 802.1X standard work separately, each one performing a different function. On the other hand, it is possible to relate profiles and user behavior patterns with the traffic analysis that origins from Proxy servers, helping to identify behavior patterns similar to the SDPI action. We can also count on Access Points association and disassociation data to define user behavior patterns related to their respective movements. This work seeks to identify anomalous patterns through the behavioral analysis of a network in order to authenticate users whose access can be revoked in real time using the 802.1X standard in conjunction with a proposed IDPS solution logic. At the end, it is possible check that there are differences between employees and, it is also possible identify and authenticate them with the data captured.

**Keywords:** *Behavioral Profile, anomalies, 802.1X, Proxy Behavioral Analysis , Network Perimeter, Network Perimeter Analysis.*

Lista de Ilustrações

Lista de Tabelas

## Lista de Abreviaturas e Siglas

| | |
|---|---|
| 802.1X | Padrão definido pelo IEEE para mecanismos de autenticação e acesso |
| AP | Access Point |
| DDOS | Distributed Denial of Service |
| DN-AP | Dados de Navegação AP |
| DN-Proxy | Dados de Navegação de Proxy |
| EAP | Extensive Authentication Protocol |
| FP | Falso Positivo |
| HCA | Horário Comum de Acesso |
| Host | Elemento conectado a rede |
| IA | Intervalo de Acesso entre URL |
| IEEE | Institute of Electrical and Electronics Engineers |
| KVM | Kernel Based Virtual Machine |
| MAC | Media Access Control |
| QA | Quantidade de Acessos a uma URL |
| RFC | Request for Comment |
| RJ 45 | Registered Jack 45 |
| SDPI | Sistema de Detecção e Prevenção de Intrusões |
| TI | Tecnologia da Informação |
| URL | Uniform Resource Locator |
| Wireless | Redes sem fio |

**Sumário**

# 1 INTRODUÇÃO

Segurança sempre foi uma das grandes preocupações da humanidade. Desde o início da evolução humana o homem se preocupa em delimitar suas fronteiras. De acordo com Freitas e Tasinafo (2011), na idade média, a humanidade construiu muros ao redor de seus castelos e propriedades para evitar que invasores efetuassem saques ou destruíssem suas propriedades.

Já na era moderna, o conceito de delimitação de fronteiras ultrapassa os limites físicos, sendo que as proteções não se limitam mais a muros de pedra e concreto, mas sim a barreiras virtuais que devem impedir acessos indevidos e ataques de agentes maliciosos.

Mesmo delimitando as fronteiras digitais, ainda é possível encontrar situações onde possíveis componentes possam efetuar ações maliciosas internamente, já tendo passado pelas barreiras externas que o validaram como componente confiável daquele espaço delimitado (Segurança de Redes de Computadores na Internet, 2012).

Novamente na idade média, guardas dentro destas fronteiras detectavam possíveis bandidos e componentes suspeitos que, mesmo passando pela fronteira de forma válida, começavam a agir de forma maliciosa. Estes guardas trabalhavam como um segundo elemento de segurança, dando uma proteção extra a todo sistema de segurança do castelo, aprisionando ou expulsando o elemento malicioso da propriedade.

No mundo digital um possível atacante pode fornecer digitais válidas, entregando um certificado válido para o autenticador. Após seu ingresso na rede de forma legítima, o atacante pode começar seu ataque. Com o monitoramento do Sistema de Detecção/Prevenção de Intrusões (SDPI) o comportamento malicioso do usuário é detectado e, por meio de integração da plataforma com o autenticador, o certificado é revogado, eliminando de forma física a comunicação do *host*, similar a expulsão da propriedade antes realizada pelos guardas na idade média (Segurança de Redes de Computadores na Internet, 2012).

Este tipo de comportamento inibe quaisquer ações de contorno que o elemento malicioso possa efetuar para prejudicar ou até mesmo ingressar na rede com suas

credenciais, já que seu acesso foi completamente revogado e sua eliminação de forma física foi realizada, seja ela em portas do tipo *Registered Jack* (RJ45) com a impossibilidade de comunicação com os demais elementos de rede como também em redes sem fio, onde sua associação com a rede em questão também não é mais possível. Essas ações são possíveis graças ao constante monitoramento e comunicação conjunta do SDPI com o autenticador e o servidor de autenticação.

## 1.1 Motivação

Em redes onde se deseja um nível adequado de segurança, constantes monitoramentos e análises são componentes chave para a detecção de anomalias. Segundo Lackner *et al.* (2008), entre as formas de detecção de anomalias e identificação de *hosts* em redes sem fio, está a realização de análise de comportamento dos *hosts* por meio da captura de sinais do meio físico. Porém, esta é uma técnica que gera elevados índices de alertas Falso Positivos (FP), como mostra o trabalho de Cho, Lou e Tan (2008), que ao adicionar variáveis como distância dos dispositivos ao ponto de acesso alcança índices de acerto de 68% e de alertas FP em torno de 32%.

Ao efetuar tratativas de análise comportamental somente na camada física o índice de acerto não se mostra eficiente. Além disso, também é possível forjar ruídos no meio físico, que simplesmente podem tornar impossível a validação de *hosts*. Uma análise em camadas superiores, onde há uma disponibilidade maior de dados é uma forma de reduzir e tornar mais eficiente a detecção de usuários e *hosts* autênticos ou maliciosos.

Segundo Hwank *et al.* (2008), implementações do protocolo 8021.X tem como objetivo permitir ou negar acessos a redes cabeadas ou sem fio no momento de ingresso de determinado *host* ou usuário, permitindo que informações sejam validadas no momento de conexão. Ocorre que esta checagem é gerada somente uma vez, permitindo ou negando o acesso do *host* a rede. *Hosts* com intenções maliciosas podem entrar na rede fornecendo credenciais válidas, podendo posteriormente roubar informações e executar ataques. Como o 802.1X faz esta checagem somente no

ingresso do *host* à rede, futuros problemas e ataques não fazem parte do escopo de verificação do padrão do protocolo. Dessa forma, o 802.1X em sua função básica não consegue evitar problemas de ataques após a validação com sucesso de *hosts* na rede.

## 1.2 Hipótese principal

O desenvolvimento de uma solução de segurança que consiga determinar a legitimidade ou não de um determinado usuário por meio de seu comportamento pode diminuir drasticamente casos de acesso por roubo de identidade, além de diminuir os riscos que uma rede exposta a invasores pode sofrer. Tudo isso revogando o acesso do invasor a rede em si. No momento de sua detecção, o SDPI efetua verificação por meio de assinaturas e aprendizado de rede. Quando anomalias de comportamento são detectadas o SDPI efetua ações específicas para mitigação do problema. Segundo Kedma *et al.* (2013), cada usuário possui um padrão de comportamento único, que pode identificar a legitimidade de sua credencial ou não. O SDPI não efetua esta verificação comportamental para cada usuário. Ao criar um SDPI capaz de identificar os perfis de cada usuário, determinando sua legitimidade ou não, aliado ao 802.1X, responsável pela entrada e saída de usuários, é possível obter uma solução com alto índice de segurança de identificação, diminuindo a possibilidade de falsos positivos utilizando somente análises de camada 1 e 2.

## 1.3 Objetivo

Desenvolver a partir de análise comportamental, uma proposta lógica para detectar padrões de comportamentos de usuários de uma rede e, junto com o protocolo 802.1X, remover usuários com comportamentos considerados suspeitos em uma infraestrutura de rede. Este trabalho realiza uma prova de conceito, trazendo ao final um estudo de caso da solução proposta.

## 1.4 Método do Trabalho

O trabalho de pesquisa busca contribuir para a detecção de anomalias de comportamento e com a ação diante da detecção de invasores ou usuários maliciosos presentes na rede por meio de credenciais válidas. Para a realização deste trabalho, as seguintes atividades são realizadas:

1. Revisão Bibliográfica - Pesquisa de trabalhos relacionados que fazem a implantação e uso das tecnologias 802.1X (Coleman, Westcott e Harkins, 2017), efetuando a verificação de usuários autenticados. Também são realizadas pesquisas em trabalhos referentes à soluções SDPI, como *Snort* (Kim *et al.*, 2015), sua utilização, implantação (Trabelsi e Alketbi, 2013) e a forma de detecção de determinadas assinaturas em sua base de dados (Kumar e Sangwan, 2012). Trabalhando em conjunto com as tecnologias, são também relacionados artigos de análise comportamental de dispositivos em redes sem fio por meio do envio e recepção de pacotes (Alipour *et al.*, 2015), além da própria análise comportamental de usuários, analisando os padrões de acesso e comportamento na rede (Kedma *et al.*, 2013);
2. Criação de ambiente para testes e coleta de informações - Por meio de ambiente virtualizado;
3. Efetuar a correlação de padrões de acesso dos usuários e as decisões sobre os padrões de comportamento são realizadas junto ao SDPI para informar a autenticidade positiva ou negativa dos usuários da rede. Os dados de navegação e coleta de dados são fornecidos de forma manual, uma vez que os dados adquiridos são oriundos de uma estrutura à parte com auditoria por meio de *proxy e* informações de coleta de dados de *Access Points* (AP) de dispositivos conectados;
4. Caracterização e informações coletadas – Efetuar a coleta dos dados por meio do fornecimento das informações do servidor *proxy*. Após a coleta, os dados são filtrados de forma compreensível para manipulação e informações podem ser retiradas de seu resultado;
5. Geração de *ranking*, identificação de usuário e ação a ser tomada – Após a análise dos dados, um *ranking* é gerado e o resultado obtido é usado pelo sistema para a tomada de decisões: excluir ou o permitir o usuário na rede.

## 1.5 Organização do Trabalho

Este trabalho encontra-se organizado nas seguintes seções:

Seção 2 - ESTADO DA ARTE: Apresenta e explica conceitos aplicados no trabalho, que são essenciais para aplicação e compreensão do estudo realizado;

Seção 3 - TRABALHOS RELACIONADOS: Estudo e comparação de trabalhos que enfatizam pesquisas em análises comportamentais humanas diante da interação com a Internet, além do estudo de padrões de comportamentos em redes sem fio corporativas para identificação do perfil de usuários, com o objetivo de prover a autenticação binária;

Seção 4 - PROPOSTA DE TRABALHO: Apresenta o plano de trabalho utilizado nesta dissertação;

Seção 5 – DETECÇÃO POR MEIO DE ANÁLISE COMPORTAMENTAL: Mostra a análise dos dados capturados, a filtragem destes e sua categorização, classificando-os de forma que seja possível denominar os comportamentos de cada usuário, mostrando se estes são os comportamentos padronizados ou detectando padrões anômalos na observação. Com a detecção de padrões anômalos, é demonstrada nessa seção a classificação nos grupos de identificação adequados e a comunicação com o protocolo 802.1X informando a necessidade de revogação de credencial. Nesta seção também são analisados os resultados e a eficácia deste método na detecção de intrusões realizadas por meio de furto de credenciais;

Seção 6 - CONCLUSÃO: descreve um resumo do trabalho, com as contribuições e considerações finais sobre os principais resultados. Também apresenta propostas para pesquisas futuras.

# 2. ESTADO DA ARTE

Neste capítulo são apresentadas as principais referências verificadas para o desenvolvimento desta pesquisa e os conhecimentos necessários que sustentam este trabalho, bem como sua posição diante das tecnologias e outros trabalhos desenvolvidos.

## 2.1 O SDPI SNORT

De acordo com Lopez *et al.* (2014), um SDPI é responsável pelo processo de monitoramento e análise de eventos de um sistema em busca de sinais de possíveis incidentes de segurança, além de exercer uma atuação em casos onde seja detectada uma anomalia classificada como incidente de segurança.

Já o SNORT, segundo Trabelsi e Alketbi (2013), e Norton (2002), é uma ferramenta baseada em software livre capaz de efetuar análise de tráfego em tempo real. Por meio de análise, a ferramenta consegue identificar anomalias na rede, identificando se há algum comportamento que não se adequa ao funcionamento habitual daquele segmento. A ferramenta consegue identificar estes padrões com o auxílio de assinaturas previamente inseridas em sua base de consulta. Muitos ataques conhecidos e maliciosos são identificados pela ferramenta e é por meio do conceito do funcionamento do SNORT que o a proposta de algoritmo desta pesquisa se baseia. A característica de funcionamento de um SDPI contribui na ação após a detecção de um elemento intruso na rede, essencial para a ação tomada sugerida nesta pesquisa.

## 2.2 O padrão 802.1X

De acordo com o desenvolvimento dos autores Congdon *et al*. (2003), a *Request for Comment* (RFC) 3580 trata exatamente do funcionamento do padrão 802.1X e seus protocolos.

O 802.1X é um padrão do IEEE para controle de acesso a redes que, em geral, permite ou nega o acesso baseado em regras e diretivas previamente criadas pelo administrador de rede. É um padrão comumente encontrado em redes sem fio e

também em redes cabeadas e trabalha em conjunto com o protocolo *Extensive Authentication Protocol* (EAP). É por meio do 802.1X que é possível permitir ou negar usuários no acesso a infraestrutura de rede, conforme proposto nessa pesquisa.

## 2.3 O servidor *proxy*

Segundo Marcelo (2005), o *proxy* é um serviço de rede que age como um intermediário entre as requisições de clientes a ele conectados a as suas solicitações. É por meio do *proxy* que informações como *sites* requisitados e tempo de navegação de cada cliente é coletado e utilizado para a análise na pesquisa. Seu papel é de vital importância para pesquisa e é por meio deste servidor que é possível a coleta dos dados necessários.

## 2.4 Aplicação de Estatísticas e suas distribuições

Segundo Morettin (2014), estatística é a ciência que faz uso das teorias de probabilidades, com o objetivo de demonstrar a frequência de ocorrência de determinadas situações ou eventos. A estatística é uma ciência que depende de coleta de dados e sua análise. Só dessa forma é possível compreender e definir as situações que podem se originar de determinada coleta.

## 2.5 Média Ponderada

Ainda de acordo com Morettin (2014), nos cálculos envolvendo média aritmética simples, todas as ocorrências possuem a mesma importância. É possível dizer que todos os valores presentes possuem o mesmo peso relativo.

Porém, existem casos onde as ocorrências possuem importâncias diferentes. Também considerados como pesos diferentes. O cálculo de média o qual leva em consideração diferente pesos e importância dos valores é conhecido por média ponderada.

O cálculo de média ponderada é dado na seguinte fórmula:

$$MP = \frac{p_1.x_1+p_2.x_2+p_3.x_3+\ldots.\ p_n.x_n}{p_1+p_2+p_3+p_n} \quad (1)$$

O cálculo da média ponderada auxilia na contagem dos *thresholds* dos usuários referente a seus respectivos acessos a endereços Web variando de acordo com cada dia. Neste trabalho, para caracterizar alguns valores de algumas características explicadas na seção 4, é considerada a divisão dos valores obtidos por 5, representando os últimos 5 dias de valores obtidos.

## 2.4 Análise Comportamental

De acordo com Kedma *et al*. (2013), o comportamento do ser humano é uma característica própria e pessoal. Cada indivíduo possui sua própria forma de lidar com as situações do dia a dia, variando entre atitudes, gestos e comportamento em si.

Na pesquisa de Pantic *et al.* (2007), os comportamentos humanos são dados por 3 elementos básicos:

- Face: Expressões faciais, movimento da cabeça e movimentação dos olhos;
- Corpo: Movimentação de pernas, braços e mãos, reconhecimento de postura, gestos derivantes de membros móveis do corpo e atividade em geral;
- Expressões vocais não linguísticas: Aumento no tom de voz, intensidade das palavras e expressões, tosse, entre outras.

Os 3 elementos tratados na pesquisa sobre a característica comportamental trazem informações que, além de apresentarem o humor do indivíduo em observação naquele momento, também trazem características próprias de identificação, similar a uma identidade do elemento em estudo.

Da mesma forma como as características citadas trazem uma identidade a cada ser humano, o comportamento na interação com sistemas informativos também apresenta características que trazem identidade ao elemento. Ainda no trabalho de Pantic *et al.* (2007), é proposta a relação e compreensão do comportamento humano para adaptação de sistemas mais próximos das características humanas.

O trabalho de Kedma *et al.* (2013) aproveita vários destes conceitos para identificar, além dos 3 elementos sugeridos, um quarto, relacionado à ações no computador, em especial em atividades *web*.

Esses comportamentos podem ainda ser previstos, conforme sugere o trabalho de Costa e Carmo (2007). Baseado na observação do comportamento humano em sistemas computacionais, inclusive navegação em sistemas *web*, é possível prever qual é a próxima ação ou tendência de realização de ação que um elemento humano terá no ambiente em estudo.

É pelo comportamento humano que se pode diferenciar as pessoas. Ocorre que o comportamento de cada ser humano também pode sofrer variações ao longo do tempo, dependendo das ações e experiências sofridas por este.

Apesar das várias formas de identificar seres humanos analisando seus respectivos comportamentos, ainda é um desafio trazer este conceito para os sistemas computacionais. Conforme é sugerido no trabalho de Kedma *et al.* (2013), é possível identificar seres humanos pelos seus respectivos comportamentos de navegação, trazendo inclusive as suas respectivas variações de humor e *stress*, entre outras características que sofrem alterações ao longo do tempo. Aliado a uma infraestrutura que consiga trazer estes dados sugeridos nas pesquisas citadas, o trabalho sugere a identificação dos seres humanos em um ambiente controlado, em uma empresa, por meio da análise comportamental, verificando se os dados presentes nas pesquisas são suficientes para a identificação dos indivíduos.

# 3. TRABALHOS RELACIONADOS

Na seção as principais características dos trabalhos que servem de base para esta pesquisa são comentadas e comparadas.

## 3.1 Introdução

A preocupação com segurança sempre foi constante nos profissionais de TI. Entre as preocupações está o fornecimento de credenciais digitais válidas em sistemas de informação.

Como citado no trabalho de Emight e Aron (2017), o furto de credenciais é uma vulnerabilidade difícil de ser mitigada. O furto de credencias normalmente ocorre por técnicas de engenharia social ou, em casos específicos, por instalação de *malwares* e programas maliciosos programados para tal ação.

A maioria dos sistemas de segurança confia nas credenciais usadas inicialmente para a autenticação, permissão e autorização dos usuários. O principal problema é que em caso de furto dessas credenciais, muitos sistemas não possuem meios de identificar um uso não autorizado destas credenciais.

Segundo os trabalhos de Alipour *et al*. (2015) e Loh *et al*. (2008), uma das formas de identificação de anormalidades em redes sem fio é a análise comportamental do meio físico, identificando padrões e alertando quaisquer mudanças que possam caracterizar comportamento anômalo. O mesmo acontece no furto de credenciais. Uma das formas possíveis de um sistema identificar legitimidade de usuário é por meio da identificação de padrões de comportamento.

## 3.2 Dados para compreensão e definição de perfil de usuário

Ainda em seus trabalhos, Alipour *et al*. (2015) e Loh *et al*. (2008), utilizam o comportamento do meio físico das redes sem fio para determinar se alguma anomalia

poderia representar alguma invasão ou tentativa de corrupção. As pesquisas se mostraram eficientes, porém, com altas taxas de alertas FP ou padrões de comportamento errôneos. Isto acontece porque os dados utilizados para definição dos padrões de comportamento são muito voláteis e quaisquer interferências do meio físico, sejam elas de caráter invasivo ou oriundas de alguma radiofrequência não relacionada a nenhum ataque, podem gerar alertas FP. Equipamentos também podem mudar seu comportamento físico em caso de atualizações de *firmware* ou sistema operacional, ocasionando novamente problemas na detecção destes dados.

Já no trabalho de Kedma *et al*. (2013), é proposto um método de identificação de comportamento para detectar traços de tendências criminosas e terroristas, baseado em dados de navegação *web*. Uma vez que padrões são estabelecidos, os autores conseguem determinar se o indivíduo possui ou não tendências para realizar tais atos. Os padrões estudados no trabalho estão listados na Tabela 1.

**Tabela 1 - Padrões de tráfego *web* coletados do trabalho de Kedma *et al*. (2013)**

| **Padrões** | **Descrição** |
|---|---|
| **Intensidade de navegação** | Identifica o quanto o elemento normalmente navega na Internet, considerando também quais endereços e a constância de seus acessos. |
| **Frequência de revisita/ frequência de *Refresh*** | Determina a frequência de acesso à endereços habitualmente acessados. |
| **Horário de Atividade** | Determina o horário de atividade do elemento e sua regularidade ou irregularidade. |
| **Nível de Interação** | Verifica a forma como o elemento utiliza os endereços. Há dois padrões de comportamento.<br>• Ativo: Utiliza formulários, envia fotos, textos, entre outras interações;<br>• Passivo: Só observa e apesar do acesso ser constante, não há interações com o endereço. |
| **Diversidade de tópicos de Interesse** | Determina os tópicos de interesse normalmente acessados pelo elemento. |
| **Correlação com tempos de Eventos de Interesse** | Determina os tempos de acesso dos tópicos de interesse. |

Fonte: Elaborado pelo Autor

A primeira coluna da Tabela 1 mostra os padrões analisados no tráfego *web* dos usuários, seguidos da descrição dos mesmos.

Outro método utilizado para identificação de comportamentos anômalos em redes é a utilização de sistemas SDPI. Bastante conhecido em ambientes de segurança, o SNORT, por meio de assinaturas, detecta ataques e ações maliciosas

que possam estar sendo aplicadas na estrutura de rede. Kim *et al*. (2015) trata exatamente da identificação de anomalias em redes sem fio por meio da aplicação do SNORT.

Uma situação similar é notada nos trabalhos de Trabelsi e Alketbi (2013), e Kumar e Sangwan (2012), onde são tratadas formas de implantação do SNORT para detecção de ataques de negação de serviço por meio de assinaturas previamente inseridas no SDPI.

Balachandran *et al*. (2002) demonstram em seu trabalho a detecção de comportamento de usuários em redes públicas de acesso sem fio em eventos por meio da análise do meio. A pesquisa leva em consideração vários dados como *throughput*, relação sinal ruído e associação de usuários nos Access Points (AP) presentes nas salas de evento, detectando os padrões de comportamento dos visitantes. Como cada usuário se comporta de forma diferente, nos diferentes dias de evento, os autores concluem a distribuição de acesso nos AP não é realizada de forma equilibrada.

Estes trabalhos apesar de conseguirem detectar e tomar ações diante das anomalias detectadas e atenderem a expectativa dos autores com sucesso, são técnicas estáticas, não adaptáveis a seus ambientes. Em suas pesquisas, os autores necessitam da inserção manual de assinaturas que demonstrem à ferramenta que seu comportamento é anômalo e malicioso naquele ambiente. Uma forma inteligente e adaptável onde a criação e adaptação das assinaturas ao ambiente não é tratada em seus trabalhos e, sem a devida atenção, é possível contornar seus sistemas de SDPI.

## 3.3 Técnicas estatísticas para definição de padrões de comportamento

Como demonstrado em alguns artigos, a utilização de técnicas estatísticas para definição dos dados é necessária. Sua compreensão de aplicação e funcionamento torna-se necessária tanto para a compreensão dos trabalhos relacionados, como também para a atual pesquisa.

O artigo de Silva e Guelfi (2010) demonstra como técnicas estatísticas auxiliam e são fundamentais para tratamento de dados em geral. No trabalho, para a correlação de alarmes e alertas isolados, técnicas estatísticas são utilizadas, auxiliando e classificando os respectivos alertas Verdadeiro Positivos (VP) de alertas FP gerados por um SDI.

Costa e Carmo (2007) também fazem uso de técnicas estatísticas para prever a navegação e, respectivamente, o comportamento de cada usuário. Porém, com técnicas de previsão de comportamento, e não necessariamente com padrões previamente aprendidos pelo SDPI.

No trabalho de Liu *et al*. (2014), também são observados métodos estatísticos comportamentais para detecção de comportamento humano em *flash crowds* web, ou seja, acessos muito intensos a alguns endereços em determinado período do dia, e como determinar se este é oriundo de comportamento humano ou até mesmo um ataque do tipo *Distributed denial of Service* (DDoS). Seu trabalho também cita e trabalha com aprendizado e análises de comportamento humano para definição de ataques e acessos.

A Tabela 2 mostra os trabalhos relacionados com esta pesquisa e os principais pontos de atuação.

**Tabela 2 - Comparação entre trabalhos relacionados**

| Trabalhos | Metodologias | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | I | II | III | IV | V | VI | VII | VIII |
| Emight e Aron (2017) | - | - | - | - | - | - | - | - |
| Alipour *et al*. (2015) | X | X | Camada física | X | X | - | - | - |
| Loh *et al*. (2008) | - | X | Camada física | - | X | - | - | - |
| Kedma *et al*. (2013) | X | X | - | X | X | X | - | - |
| Kim *et al*. (2015) | X | X | - | - | - | - | - | - |
| Trabelsi e Alketbi (2013) | X | X | - | X | X | - | - | - |
| Kumar e Sangwan (2012) | X | X | - | - | - | - | - | - |
| Silva e Guelfi (2010) | X | X | Camada de enlace | - | X | - | - | - |
| Costa e Carmo (2007) | - | - | - | X | X | - | - | X |
| Liu *et al*. (2014) | X | X | - | X | X | X | - | - |
| Balachandran *et al*. (2002) | - | X | Camadas física e enlace | X | X | - | Coleta por meio de associações | X |
| Pesquisa proposta | X | X | *Proxy* | X | X | X | Coleta por meio de associações | - |

X – Possui
– Não Possui

I. Detecção baseada em assinaturas pré-definidas;
II. Verificação constante do meio de tráfego;
III. Verificação de meio físico;
IV. Capacidade de aprendizado de padrões de comportamento;
V. Utilização de métodos estatísticos para obtenção de resultados;
VI. Não sofre problemas e falsos positivos com ações maliciosas de ataques em meio físico de transmissão dos dados;
VII. Utiliza o deslocamento físico do usuário em complemento como definição de padrão de comportamento;
VIII. Previsão do comportamento do usuário.

Como pode ser visto na Tabela 2, o item I baseia-se na detecção de anormalidades em assinaturas pré-definidas. A exemplo de ferramentas como o SNORT, a utilização de assinaturas pré-definidas depende de um aprendizado já efetuado, que é fornecido para o sistema realizar a detecção daquele padrão de comportamento. O sistema proposto neste trabalho utiliza um sistema de pesos que se assemelha bastante com assinaturas pré-definidas, justificando essa comparação.

O item II destaca-se por trabalhos onde a verificação constante do meio de tráfego é realizada. A verificação constante do meio de tráfego é necessária para analisar as possíveis anomalias e comportamentos do meio físico para detecção de invasores.

Já o Item III também destaca como verificação do meio de tráfego, o meio físico que os dados são trafegados. Essa verificação é realizada em camada física e analisa as variações presentes na transmissão do meio físico.

No item IV é possível analisar a capacidade de aprendizado de padrões de comportamento, ou seja, os trabalhos neste item conseguem se adequar e aprender novos padrões de comportamento, variando de acordo com o momento da análise.

O item V caracteriza os trabalhos que utilizam métodos estatísticos para obtenção de resultados de suas pesquisas. Esses trabalhos, além de contar com as características técnicas, também possuem métodos estatísticos para demonstração de seus resultados.

No item VI destacam-se trabalhos que não possuem problemas com alertas FP diante das características analisadas. Trabalhos que utilizam o meio físico de comunicação como somente a única forma de análise de comportamento estão sujeitos a problemas de falso positivo, uma vez que o meio físico está sujeito a interferências externas que não necessariamente caracterizam um ataque ou comportamento invasor. Esta característica de comparação é importante já que o presente trabalho analisa camadas superiores ao meio físico e não está sujeito a este tipo de interferência.

O item VII envolve trabalhos que realizam como parâmetros de comportamento o deslocamento físico de usuários por meio de associações e desassociações de AP,

trazendo os padrões de deslocamento de usuários em um determinado espaço físico. É, portanto, uma característica vital presente no trabalho atual.

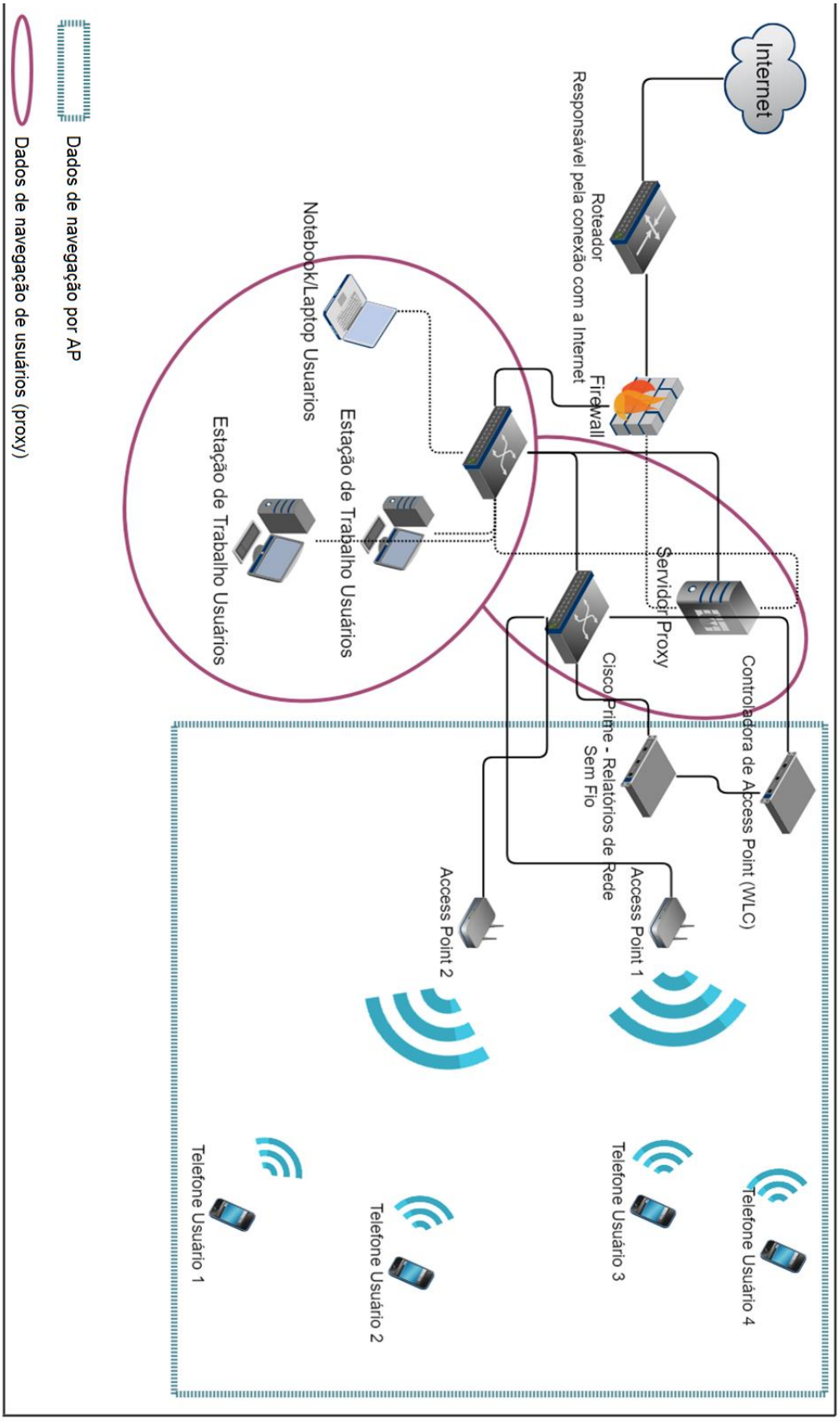
Por fim, o item VIII caracteriza trabalhos que junto ao seu aprendizado, fazem uma previsão do comportamento de determinado usuário, e conseguem, desta forma, identifica-los como normais ou anômalos. Este é um trabalho futuro a ser desenvolvido a partir do resultado desta pesquisa, por isso trabalhos relacionados com este item são citados, pois trazem características importantes para o desenvolvimento da atual pesquisa e de futuras a serem realizadas a partir dos resultados desta.

Todos estes itens são considerados como importantes para esta pesquisa uma vez que envolvem diferentes formas de identificação de padrões de comportamento e formas diferentes de detecção de intrusos. Desta forma, são itens considerados e comparados na pesquisa proposta.

## 4. PROPOSTA DE PESQUISA

Por meio da coleta de dados da rede, análises comportamentais são realizadas e o padrão de cada usuário é definido. Para isso, na Figura 1, é considerada a seguinte topologia:

**Figura 1 - Demonstração de Topologia disponível para o experimento**




Fonte: Elaborado pelo Autor

A Figura 1 mostra a captura de informações da rede que possibilita a análise e definição dos padrões de comportamento dos usuários. Para isso, são considerados (1) Dados de Navegação de usuários (DN-*proxy*); e (2) Dados de Navegação por AP (DN-AP).

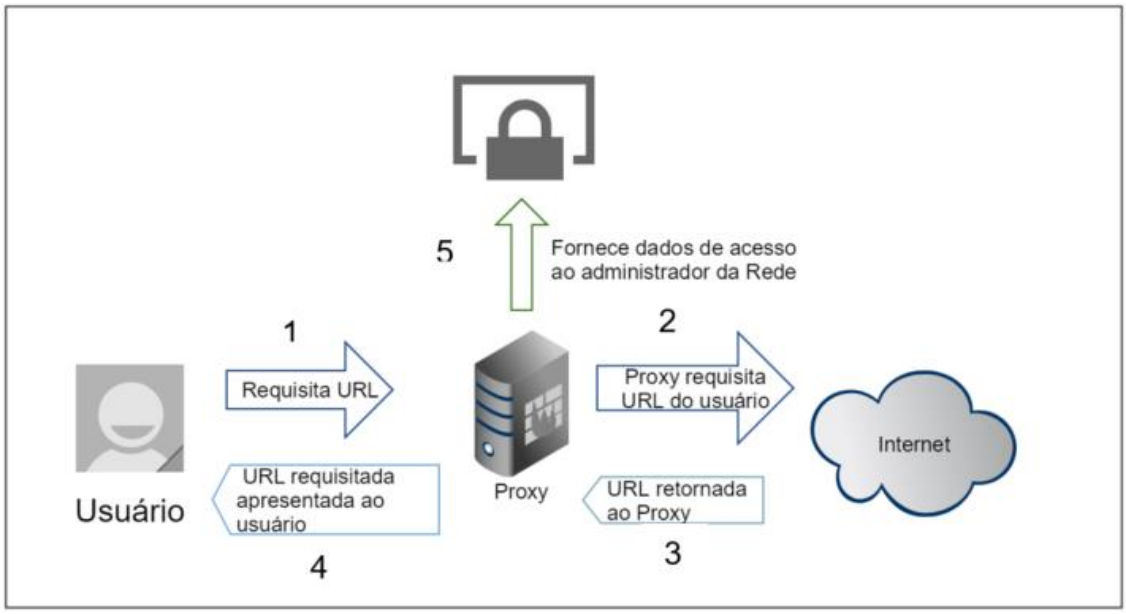
## 4.1 Dados de navegação de usuários (DN-*proxy*)

Por meio da análise do DN-*proxy,* tráfego da navegação dos usuários gerada por *proxy*, são coletados os dados dos funcionários ativos da companhia. Com esses dados, é possível determinar padrões e tendências de acesso do funcionário, demonstrando sua autenticidade. São coletadas as seguintes informações (Kedma et al. 2013):

- Os *sites* acessados, com as respectivas datas e horários;
- Velocidade de acesso e requisição de navegação de um *site* para outro;
- Períodos de maior e menor intensidade de uso da Internet e acesso a *sites*;
- *Sites* mais comuns acessados.

A Figura 2 demonstra o processo de captura dos dados.

**Figura 2 – DN-*proxy:* diagrama de coleta de informações**




Fonte: Elaborado pelo autor

Por meio do diagrama apresentado na Figura 2, é possível observar a forma de como os dados são coletados, com a requisição de *sites* pelo usuário, a intercepção do *proxy* e o fornecimento de informações para o administrador.

Para a criação do perfil comportamental do usuário, duas características são analisadas a partir do DN-*proxy:* (1) Quantidade de Acessos à uma URL (QA); e (2) Intervalo de Acesso entre URL (IA).

## 4.2 Dados de navegação por AP (DN-AP)

As soluções corporativas de acesso à rede sem fio fornecem relatórios nos quais são possíveis verificações de padrões de comportamento. Este é um dado extremamente valioso uma vez que define também padrões de localização dos usuários.

Como pode ser observado no trabalho de Balachandran *et al*. (2002), a movimentação de usuários analisados em um evento é fundamental para definir o comportamento deles. O mesmo se aplica em ambientes empresariais, que podem caracterizar o comportamento dos usuários de acordo com os AP associados em vários períodos do dia.

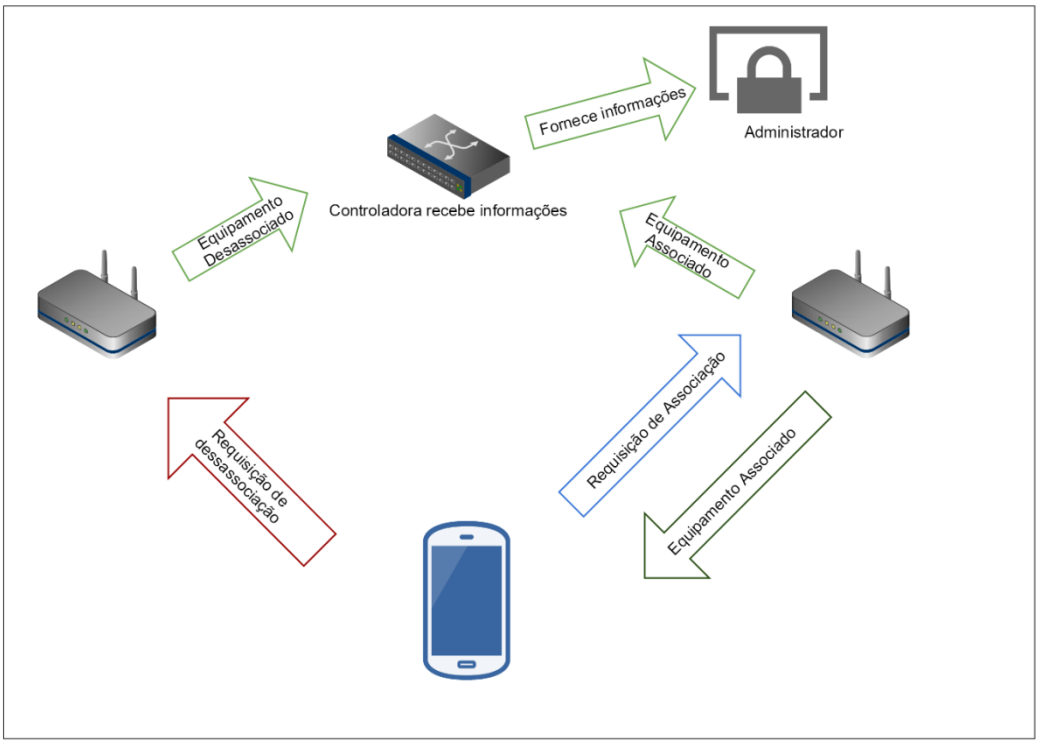
Como pode-se observar na Figura 3 os dispositivos móveis dos funcionários conectam-se a rede corporativa por meio dos AP disponíveis na companhia. Cada funcionário pertence a um setor. Cada setor encontra-se fisicamente em localidades específicas, sendo elas prédios ou andares, com um AP fixo responsável por fornecer o acesso à rede em cada uma delas. Os funcionários de um setor, em geral, se conectam aos AP locais, mas por necessidade de locomoção podem se conectar a AP diferentes.

O DN-AP, dados da associação de usuários aos AP, permite verificar o deslocamento do funcionário pela companhia em determinados horários e dias. Estes também são dados considerados nos padrões comportamentais para determinar legitimidade ou não do funcionário ingressante daquela rede.

Dados recolhidos:

- Horário de associação;
- Horário de desassociação;
- AP que o equipamento móvel do funcionário se associa durante o dia.

**Figura 3 – DN-AP: diagrama de coleta de informações**




Fonte: Elaborado pelo autor

É possível observar por meio da representação da Figura 3 como os dispositivos da rede conseguem coletar as informações necessárias de captura de rede sem fio. Informações de associação e desassociação são diretamente enviadas para a controladora que, por sua vez, envia para o administrador.

Para a formação do perfil comportamental dos usuários, a característica analisada a partir da análise DN-AP é o Horário Comum de Acesso (HCA).

## 4.3 Extração dos Dados

Os dados de navegação de usuários coletados por meio da DN-*proxy* são analisados e formatados pelo sistema, gerando tabelas com os dados de navegação de determinados períodos, como pode ser visto na Tabela 3.

**Tabela 3 - Exemplo de Tabela gerada pelo *proxy* e dados extraídos.**

| Usuário | Acesso | Horário | IA |
|---|---|---|---|
| **Usuário 1** | www.uol.com.br | 13:52:54 – 20/05/2018 | 10:01 |
| **Usuário 1** | www.itau.com.br | 14:02:53 – 20/05/2018 | |
| **Usuário 1** | www.estadao.com.br | 14:02:55 – 20/05/2018 | 00:07 |
| **Usuário 1** | www.youtube.com.br | 14:03:03 – 20/05/2018 | |
| **Usuário 1** | www.facebook.com | 14:07:02 – 19/05/2018 | 01:01 |
| **Usuário 1** | www.linkedin.com | 14:08:03 – 19/05/2018 | |

Fonte: Elaborado pelo autor.

A

Tabela 3 é um exemplo da extração DN-*proxy* para um determinado usuário (1) que permite cálculos de IA. Na Tabela 4 verifica-se um exemplo da quantidade de *sites* acessados em um determinado período, que permite o cálculo de QA.

**Tabela 4- Exemplo de *sites* acessados e quantidade por período.**

| Usuário | *Sites* acessados | QA |
|---|---|---|
| **Usuário 1** | www.youtube.com | 230 |
| **Usuário 2** | www.estadao.com.br | 193 |
| **Usuário 3** | www.uol.com.br | 52 |

Fonte: Elaborado pelo autor

A Tabela 4 exemplifica o DN-*proxy*, demonstrando os *sites* acessados e as respectivas QA (quantas vezes cada um deles são acessados), auxiliando na definição dos padrões de comportamento dos usuários.

Os DN-AP são atrelados principalmente ao percurso que normalmente o usuário realiza em seu dia a dia. Por exemplo, qualquer aparelho celular, *tablet* ou *notebook* que se associar aos AP tem seu movimento registrado. Ao final, também é

gerada uma tabela, demonstrando os locais onde os usuários costumam trafegar e o horário médio de início e fim de trabalho (exemplo na Tabela 5).

**Tabela 5- Exemplos de coleta de DN-AP**

| Usuário | AP atrelado e horários de acesso |
|---|---|
| **Usuário 1** | AP 1 – Associação 13:52:31/Desassociação 14:02:41 – 20/05/2018<br>AP 2 – Associação 14:02:58/Desassociação 15:02:35 – 20/05/2018 |
| **Usuário 1** | AP 1 – Associação 13:55:31/Desassociação 14:45:41 – 19/05/2018<br>AP 2 – Associação 14:02:58/Desassociação 15:02:35 – 19/05/2018 |
| **Usuário 1** | AP 1 – Associação 10:52:31/Desassociação 14:02:41 – 20/04/2018<br>AP 2 – Associação 14:02:58/Desassociação 15:02:35 – 20/04/2018 |
| **Usuário 1** | AP 1 – Associação 08:52:31/Desassociação 14:02:41 – 15/05/2018<br>AP 2 – Associação 20:02:58/Desassociação 15:02:35 – 16/05/2018 |

Fonte: Elaborado pelo autor.

A Tabela 5 demonstra um exemplo de DN-AP, com dados extraídos da solução de rede sem fio com os relatórios de associação/desassociação por usuário na rede. É possível observar que informações como presença física do usuário na companhia, assim como seus hábitos de deslocamento, entrada e saída, são extraídos e usados na análise HCA.

## 4.4 – Considerações de Pesos para a Geração do Ranking

No trabalho de Kedma *et al*. (2013), são propostos cenários que remetem à comportamentos possíveis para identificação de comportamentos terroristas. No trabalho há dois cenários propostos nos quais os autores demonstram os principais casos e identificação de usuários. O primeiro cenário classificado como *Hit and Run Accident* classifica indivíduos que acidentalmente cometeram algo que possa lhes comprometer e, diante da ação, a busca referente ao assunto torna-se muito constante e comum, sobrepondo as ações naturais normalmente efetuadas pelo usuário. Neste cenário, o usuário é classificado como convencional, visto que o mesmo não necessariamente possui características para um ataque terrorista. Porém,

a ação fora do padrão é analisada e levada em conta na definição do padrão de comportamento daquele usuário. Outro cenário característico é o *Terrorist Attack*. Basicamente, o usuário possui intenções de efetuar um ataque terrorista e, para monitorar a situação de ação de tal ataque, o usuário pesquisa informações sobre acidentes ou atentados por meio de *strings* específicas de notícias para verificar se o ataque ocorreu de forma bem-sucedida ou não. Caracteriza-se também como um comportamento recorrente daquele usuário, podendo ser classificado também como forma de padronização de comportamento.

Tendo por base os trabalhos de Kedma *et al*. (2013) e Carmo e Costa (2007), além de observações de comportamento característico dos usuários da empresa, nesta dissertação o ranking é classificado de acordo com pesos atribuídos às três características previamente citadas:

1. QA: quantidade de acesso à uma URL, obtida a partir do DN-*proxy*;
2. IA: intervalo de acesso de um domínio a outro, obtida a partir do DN-*proxy*;
3. HCA: horário comum de acesso, obtida a partir do DN-AP.

### 4.4.1 Quantidades de Acessos à uma URL (QA)

Segundo Kedma *et al*. (2013), em todos os cenários de possível detecção de terroristas e padrões comportamentais é essencial entender quais são os endereços acessados pelos usuários de uma determinada rede. Segundo o autor trata-se da detecção de padrões de interesse e tendências dos usuários. Além disso, endereços de acesso também são essenciais em ações de marketing e vendas. Portanto, a identificação dos endereços de interesse do usuário é classificada como de alta importância para autenticação de um dado.

Em relação à quantidade de acessos efetuados a esses endereços, um valor elevado acima de um *threshold* pode indicar um interesse grande em um assunto e caracterizar um possível comportamento anômalo.

Sendo assim, o peso que a ocorrência desta característica atribui ao ranking é calculado por meio da comparação entre o valor médio diário da quantidade de visitas aos dez *sites* mais acessados e a média ponderada dos acessos dos últimos cinco dias (*threshold*):

1. Soma-se a quantidade diária de acessos dos dez *sites* mais visitados. O valor é armazenado em QA;
2. Calcula-se a média ponderada dos acessos aos dez *sites* mais visitados dos últimos cinco dias de acordo com a fórmula (2). O valor é armazenado como *threshold*;
3. Comparam-se os valores de QA e do *threshold.* Caso QA > *threshold,* considera-se que a característica analisada corresponde a um *match*. Caso contrário, considera-se que a característica está dentro do perfil*;*
4. Quando ocorre um *match,* o peso atribuído no ranking à característica QA é 4. Caso o *match* não ocorra, não é atribuído valor algum à QA.

A média ponderada adaptada considerada para a definição de legitimidade do usuário na geração de ranking é obtida por meio da seguinte fórmula:

$$DC = \left(\frac{(DC-1)5+(DC-2)4+(DC-3)3+(DC-4)2+(DC-5)1}{5}\right) \qquad (2)$$

Onde: DC= Dia Correspondente

Para a definição dos perfis de QA, é observado o período de uma semana de trabalho (5 dias úteis). Inicialmente a ideia era definir uma média aritmética simples dos horários comuns do funcionário para se estabelecer o perfil comportamental individual para a característica QA. Contudo, frente ao fato de que atualizações diárias na rotina dos funcionários poderem influir diretamente nos perfis comportamentais especificados em uma semana, a opção encontrada é a utilização de uma média ponderada diária para os cálculos das características.

A média ponderada adaptada aplicada leva em consideração o padrão de comportamento de cinco dias atrás, considerando o dia anterior com um peso maior (Peso 5) até o quinto dia com um peso menor (Peso 1), dividido pelo número de dias considerados.

A adoção desta métrica adiciona relevância aos mais recentes comportamentos do usuário em detrimento aos mais anteriores. A adoção da média ponderada traz mais fidelidade ao comportamento do usuário.

O valor máximo que pode ser atribuído como peso ao ranking, para cada característica, é 4. Justamente pela importância da característica QA, o valor máximo é adotado em caso de *match.*

### 4.4.2 Intervalo de Acesso de um domínio a outro (IA)

Da mesma forma que o trabalho de Kedma *et al.* (2013), na pesquisa de Costa e Carmo *et al.* (2007) as URL acessadas e a quantidade de acessos são definidos como assinaturas de comportamento. Outro dado importante levado em conta pelos autores para definição de uma assinatura é o intervalo de acesso, ou seja, o tempo médio que um usuário leva para acessar um endereço e, posteriormente, trocar para outro em um domínio diferente.

No trabalho de Pan, Hu e Liu (2014) é tratado o comportamento humano em situações onde *sites* sofrem com eventos de *Flash Crowd* - um conjunto muito grande de acessos legítimos a um determinado endereço em um curto período de tempo. O tráfego Flash Crowd costuma ser tão intenso que muitas vezes tem características parecidas a um ataque DDoS. Segundo Yu *et al*. (2012), ataques de DDoS são comumente efetuados por máquinas e o valor do intervalo de acesso tende a ser diferente de uma ação originada pelo comportamento humano.

Em relação ao tempo de permanência em cada domínio acessado alguns cuidados devem ser levados em conta. Um usuário pode acessar um *site* e sair para realizar outra atividade, aumentando demasiadamente o tempo que o domínio permanece com acesso. Como isso enviesa o cálculo da média dos intervalos de acesso, 20 minutos é considerado como o tempo máximo de permanência em um domínio. A razão é que valores de tempo mais altos caracterizam a saída do funcionário de estação de trabalho – após 20 minutos, as estações de trabalho estão programadas para bloqueio de sessão. Da mesma forma, tempos de acesso muito curtos, abaixo de 3 segundos, também são desconsiderados, uma vez que

caracterizam *sites* acessados por engano ou requisições de agentes de software requisitando informações por meio da rede.

Sendo assim, tendo em vista que cada usuário possui uma média diferente no tempo de acesso aos domínios acessados, esta característica tem grande importância para o processo de identificação comportamental e é analisada de acordo com os seguintes passos:

1. Os tempos de permanência em cada *site* visitado são coletados diariamente;
2. Após a eliminação dos *outliers* (tempos muito curtos, abaixo de 3 segundos ou maiores que 20 minutos), calcula-se a média diária destes tempos. Este cálculo é uma média de IA e não precisa ser realizado necessariamente no fim do dia, pode ser feito a qualquer momento;
3. Comparam-se o valor da média de IA do dia atual com o menor valor de média de IA do mês anterior (*threshold)*;
4. Caso o valor da média de IA analisada seja menor que o valor do *threshold,* o *match* acontece. Neste caso um peso de valor 4 é atribuído ao ranking. Caso contrário valor algum é adicionado.

### 4.4.3 Horário Comum de Acesso (HCA)

De acordo com Kedma *et al*. (2013), um dos pontos importantes para detecção de padrões terroristas ou tendências a atos deste tipo são os acessos a conteúdos em horários não comuns. O autor explica que este é um indício de algo errado no comportamento do usuário. No ambiente corporativo, a situação é similar, uma vez que os usuários possuem horários regulares que podem acessar a rede interna da companhia.

Horários não coincidentes com o período de trabalho, podem caracterizar uma situação anômala passível de investigação. Esses dados são coletados por meio das associações e desassociações dos AP presentes na companhia, uma vez que os funcionários submetidos à coleta de dados possuem acesso a estes dispositivos por meio da conexão de seus celulares à rede *wireless* corporativa.

Os funcionários analisados possuem horário de trabalho convencional com entrada as 08:00h e saída as 17:00h. Apesar dos horários pré-definidos, não necessariamente os funcionários cumprem à risca dos horários. A companhia disponibiliza horários móveis, dessa forma os funcionários podem entrar a partir das 07:00h e sair até as 18h, sempre cumprindo 8 horas de trabalho diárias. Cada funcionário é avaliado de acordo com seus costumes de horário de entrada e saída entre os horários permitidos de trabalho, considerando também os horários de saída e entrada de almoço. Esta característica tem como *threshold* o HCA do mês anterior: registra-se o horário mais cedo que um usuário chegou à empresa e o horário mais tarde que ele saiu. Esses registros indicam, de acordo com Kedma *et al*. (2013), comportamentos que podem ser considerados suspeitos. Usuários que chegam antes (do menor) ou saem depois (do maior) horário histórico registrado no mês anterior, estão descumprindo o **horário comum de acesso**. Por exemplo, quando um determinado usuário, cujo perfil comportamental está em avaliação, chega à empresa em um horário **anterior** ao horário histórico registrado, considera-se que o **HCA** foi descumprido. A mesma lógica se aplica quando um usuário sai depois do maior horário registrado no histórico. Em ambos os casos o peso atribuído à esta característica tem valor 2.

É importante notar que para a obtenção desta métrica é preciso que haja uma conexão ao AP, feita pelo celular do usuário. Como o celular não é um equipamento obrigatório de uso no trabalho, esta característica tem menos relevância que as anteriores. Ao final do dia, é coletada a observação com os dados de associação/desassociação dos equipamentos, sendo verificados seu primeiro horário de associação como entrada do funcionário na companhia e seu último horário de desassociação como seu horário de saída.

A Tabela 6 resume as características analisadas para a geração do ranking e os pesos máximos usados de acordo com a relevância da característica.

**Tabela 6 – Características e pesos considerados para formar o ranking.**

| Características Consideradas | Pesos |
|---|---|
| **QA** | 4 |
| **Média de IA** | 4 |

| | |
|---|---|
| **HCA** | 2 |

Fonte: Elaborado pelo autor.

A Tabela 6 demonstra o peso considerado para cada característica – estes pesos são adicionados ao valor do ranking quando uma das características listadas gera um *match.* Quando uma característica não dá *match* o valor do peso para essa característica específica é 0.

Para a definição dos perfis, é observado o período de uma semana de trabalho (5 dias úteis). Inicialmente a ideia era definir uma média aritmética simples dos horários comuns do funcionário para se estabelecer o perfil comportamental individual para cada característica listada na Tabela 6. Contudo, frente ao fato de que atualizações diárias na rotina dos funcionários poderem influir diretamente nos perfis comportamentais especificados em uma semana, a opção encontrada é a utilização de uma média ponderada diária para os cálculos das características.

Um resumo sobre os valores de pesos e os *thresholds* adotados em cada característica analisada pode ser visto na Tabela 7

**Tabela 7 – *Thresholds* usados para a comparação**

| Características | *Threshold* | *Match* | Peso |
|---|---|---|---|
| **soma do QA diário** | Média ponderada dos últimos 5 dias | QA > *threshold* | 4 |
| **Média do IA diário** | Menor média de IA do mês anterior | Média IA < *threshold* | 4 |
| **HCA na entrada e na saída** | horário de entrada mês anterior<br>horário de saída mês anterior | HCA < horário de entrada mês anterior<br>HCA > horário de saída mês anterior | 2 |

Fonte: Elaborado pelo usuário.

É possível verificar as diferentes métricas de comparação entre as características consideradas e seus respectivos pesos. Como pode ser visto na Tabela 7, os pesos de cada característica só são atribuídos ao ranking de um usuário que está sendo avaliado quando os valores observados de cada característica apresentam *match.* O valor máximo que o ranking pode receber durante uma análise é 10 (soma dos pesos). É importante perceber que a característica HCA tem um peso

com valor máximo de 2, mesmo quando são observados, ao mesmo tempo, o menor horário de entrada e o maior horário de saída.

## 4.5 Ranking do usuário

O ranking de cada usuário é calculado diariamente - seu valor inicial é nulo (0). Quando o *match* de alguma das características analisadas ocorre, o peso da mesma é somado ao cálculo do ranking. É justamente com base nos valores de ranking que um usuário é considerado autêntico ou não:

- Usuário autêntico: quando o ranking tem o valor 0 (nenhuma característica apresenta *match*). Para os usuários autênticos também são válidos os valores 2 ou 4 (uma única característica apresenta *match*);
- Usuário não autêntico: neste caso, pelo menos duas características avaliadas apresentam *match, o*u seja, o ranking tem valores ≥ 6.

O valor de *threshold* máximo considerado para cada usuário até seu reconhecimento de não legítimo é 6, justamente a soma dos pesos de duas ou mais características listadas na Tabela 6. De acordo com o que é detalhado nos trabalhos de Kedma *et al.* (2013) e Costa e Carmo (2012), por serem dependentes uns dos outros, estes dados são considerados determinantes para detecção de padrões.

## 4.6 Detecção de usuário intruso e revogação de acesso

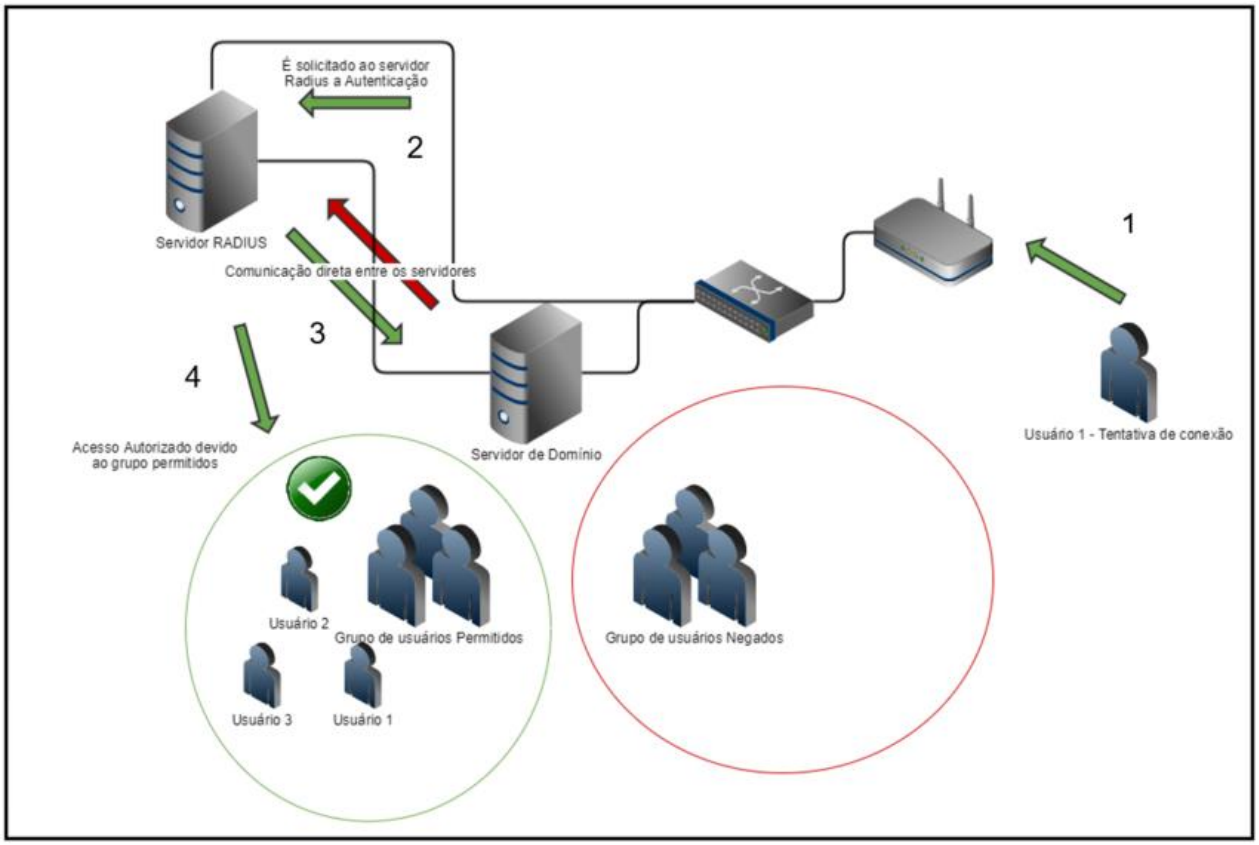
Ao detectar um comportamento anômalo que fuja dos padrões do ranking gerado, o sistema precisa tomar uma ação para eliminar o usuário intruso da rede.

Uma comunicação direta entre o algoritmo gerador do ranking e o servidor Radius responsável pela autorização/negação dos acessos, permite que o sistema revogue o acesso do usuário em questão. Em comunicação direta, o usuário é: (1) classificado como anormal; (2) movido de um grupo considerado confiável para um não-confiável; e (3) seu acesso é revogado temporariamente, até a intervenção de um administrador de rede para análise da anormalidade.

Esta integração é possível devido à comunicação direta do sistema com o 802.1X provido pelo servidor Radius e o servidor de domínio, que identifica os usuários

em grupos de acesso. Na Figura 3, é possível verificar o funcionamento na tentativa de conexão de um usuário à rede proposta.

**Figura 3 - Tentativa de Conexão com 802.1X**

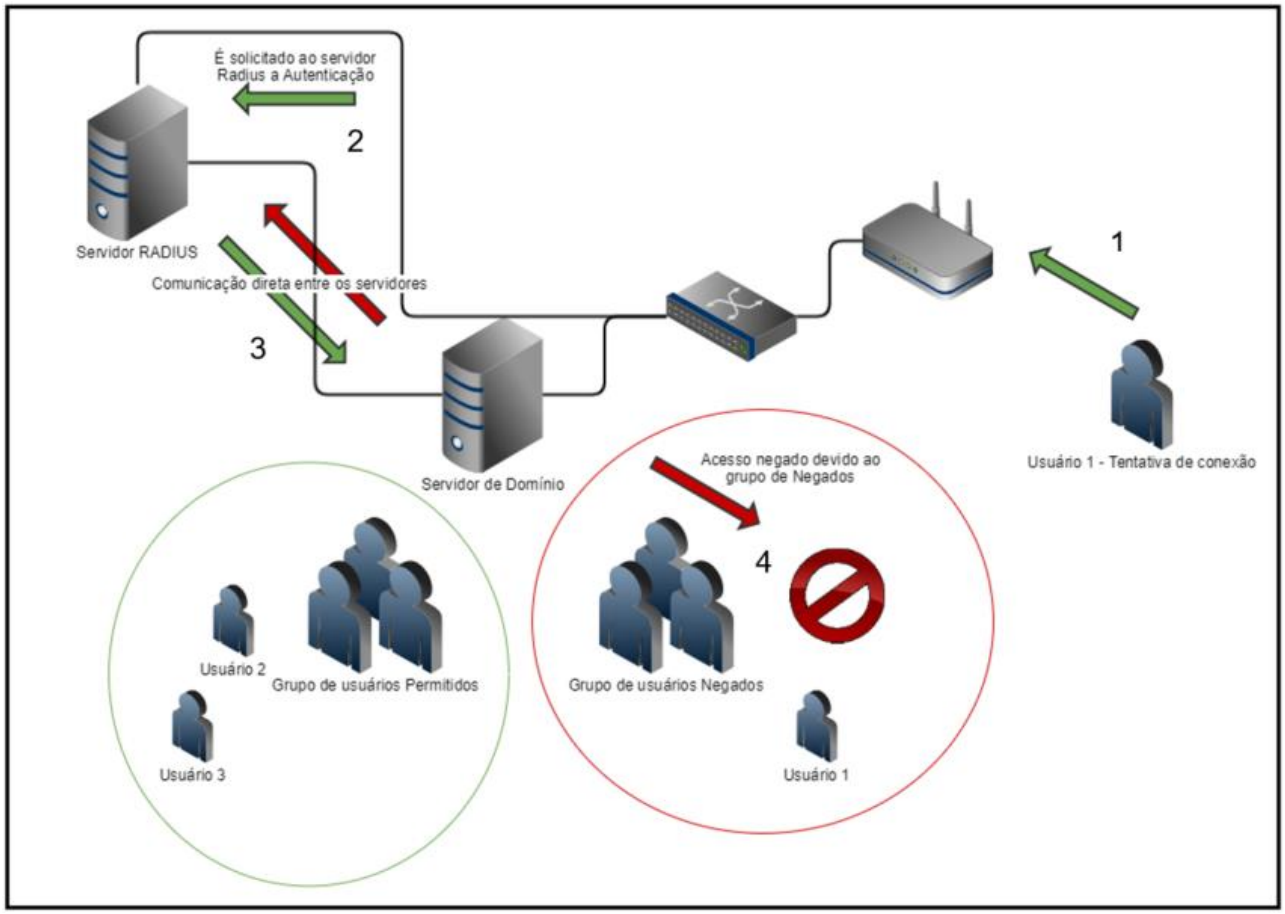



Fonte: Elaborado pelo Autor

Como mostra a Figura 3, ao efetuar a tentativa de ingresso na rede (1), o Usuário 1 envia suas credenciais para os servidores de autenticação (2). Como o servidor Radius possui uma integração direta com o servidor de domínio, o grupo ao qual o Usuário 1 pertence é verificado (3).

Por funcionamento padrão do sistema de domínio, todos os usuários encontram-se no grupo Permitidos, e conseguem efetuar o acesso à rede normalmente (4). Porém, com a constante verificação dos padrões de comportamento, um usuário poderá ser considerado não autêntico, e ter sua classificação alterada. Nestes casos, o algoritmo gerador de ranking responsável pela verificação notifica o servidor de domínio, alterando o grupo do usuário de Permitidos para Negados, conforme pode ser verificado na Figura 4 - Tentativa de conexão negada com 802.1X.

**Figura 4 - Tentativa de conexão negada com 802.1X**




Fonte: Elaborado pelo autor.

Após a detecção de não autenticidade do usuário, suas credenciais são movidas do grupo de Permitidos para Negados. É possível verificar na Figura 4 que o usuário tenta efetuar seu ingresso na rede logo em seguida. Ocorre que como seu grupo de acesso agora não permite seu ingresso, o usuário considerado não autentico tem seu acesso negado. Dessa forma, o usuário suspeito não ingressa na rede e evita problemas que invasores possam causar no perímetro.

## 5. DETECÇÃO POR MEIO DE ANÁLISE COMPORTAMENTAL

Nesta seção são descritos os resultados de um experimento envolvendo a geração de ranking e a análise do perfil de três funcionários pelo período de um ano, analisando seus dados de QA, HCA e IA. Posteriormente, estes dados são comparados entre si, formando os *thresholds*.

## 5.1 Funcionário 1 – Dados obtidos

A amostra obtida para o funcionário 1 corresponde a 3 meses de observação. Os dados relativos ao *threshold* de (QA) dos 10 domínios por período para o mês de janeiro estão presentes na Tabela 7, aplicando os dados obtidos na fórmula 2.

**Tabela 7 – Dados obtidos do funcionário 1 Mês de Janeiro – Dados para consideração de *Threshold***

| Período (dias) | Média Aritmética de QA | QA - *Threshold* |
|---|---|---|
| **2 a 8** | 123,50 | |
| **3 a 9** | 107,33 | |
| **4 a 10** | 105,00 | |
| **5 a 11** | 105,00 | |
| **6 a 12** | 105,00 | 319,63 |
| **7 a 13** | 102,50 | 312,96 |
| **8 a 14** | 79,00 | 287,00 |
| **9 a 15** | 58,00 | 245,70 |
| **10 a 16** | 87,75 | 243,55 |
| **11 a 17** | 92,20 | 249,30 |
| **12 a 18** | 97,17 | 262,57 |
| **13 a 19** | 114,71 | 294,46 |
| **14 a 20** | 116,50 | 321,00 |
| **15 a 21** | 149,33 | 368,66 |
| **16 a 22** | 171,60 | 426,28 |
| **17 a 23** | 170,25 | 466,67 |
| **18 a 24** | 152,25 | 474,44 |
| **19 a 25** | 131,75 | 454,20 |
| **20 a 26** | 88,50 | 387,67 |
| **21 a 27** | 43,25 | 288,05 |
| **22 a 28** | 43,25 | 214,10 |
| **23 a 29** | 43,25 | 165,55 |
| **24 a 30** | 63,80 | 159,35 |
| **25 a 31** | 71,80 | 174,74 |

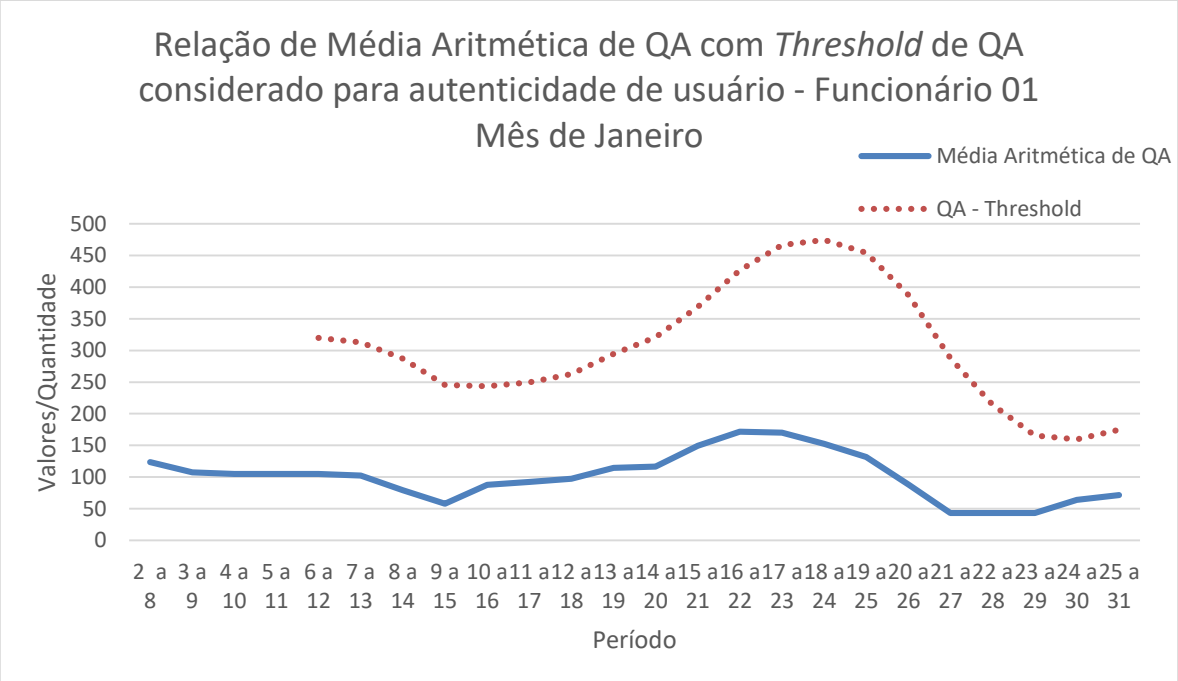
Fonte: Elaborado pelo autor.

A Tabela 7 contém os dados calculados (QA) a partir dos dados presentes na observação da quantidade de *sites* acessados diariamente em um mês. Os dados calculados presentes são específicos do mês de janeiro de 2017. Pode-se perceber

que nos quatro primeiros dias a média ponderada não é calculada, já que depende de pelo menos quatro observações anteriores da média aritmética.

Também é possível observar o comportamento do usuário por meio de gráficos, conforme apresentado na Figura 5.

**Figura 5 - Representação gráfica dos valores de QA (Tabela 7) apresentados no mês de Janeiro do funcionário 1**

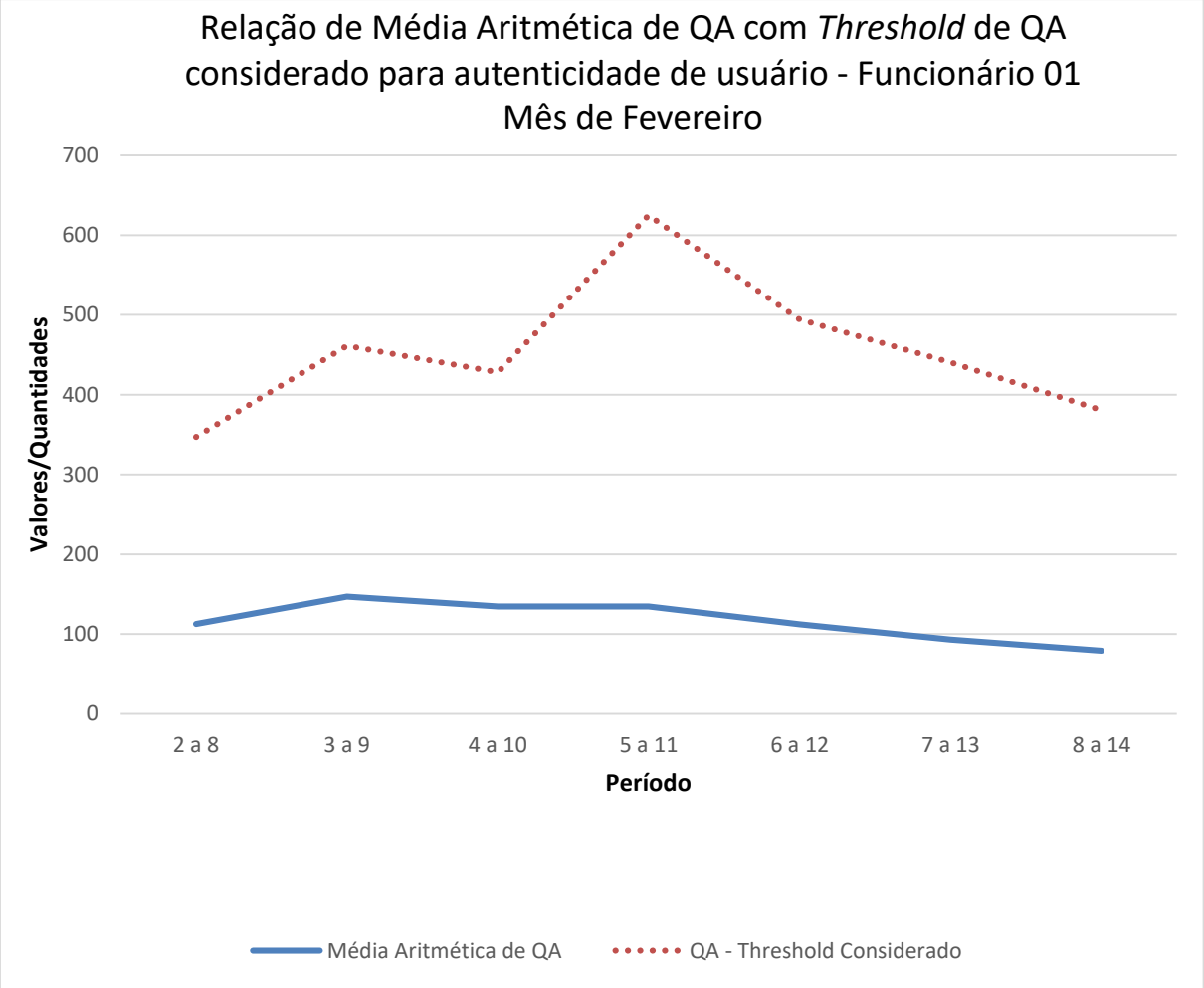



Fonte: Elaborado pelo autor

Na Figura 5 é possível observar que as curvas sólidas e pontilhadas são bastante parecidas, já que apesar de estarem em patamares diferentes de valores, apresentam um comportamento bastante parecido. A linha pontilhada denominada QA é o *threshold* obtido por meio da média ponderada dos últimos cinco dias, enquanto a linha sólida indica a média aritmética dos dias que foram analisados. Como o mês de janeiro não possui dados anteriores para verificação das outras características observadas do usuário para definir sua autenticidade, não é possível gerar seu ranking. Porém é importante destacar que a linha pontilhada é exatamente o *threshold* que delimitará o comportamento do funcionário. Nos meses apresentados a seguir, é possível verificar seu papel na definição de autenticidade dos usuários.

Já no mês de fevereiro, é possível observar o comportamento do funcionário 1 na Figura 6, de forma similar ao verificado no mês anterior.

**Figura 6 - Representação gráfica dos valores de média aritmética de QA e *threshold* considerado apresentados no mês de Fevereiro do funcionário 1**

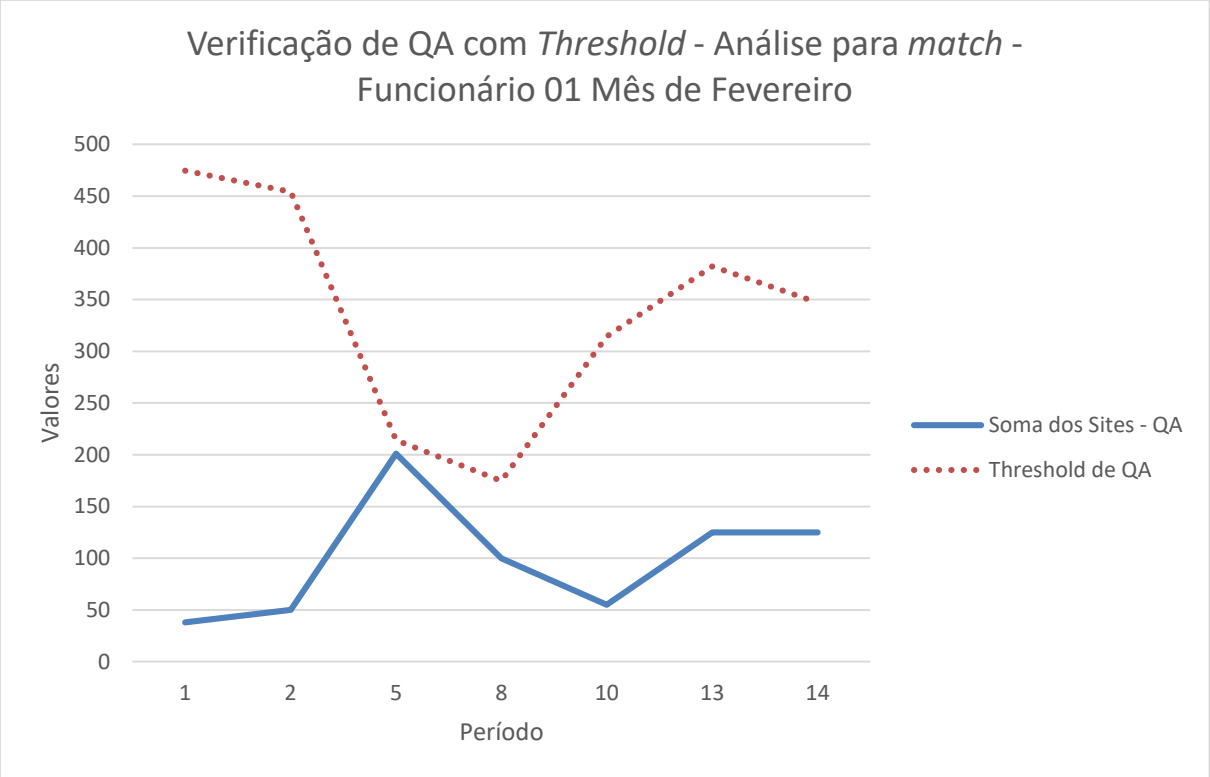



Fonte: Elaborado pelo autor.

É possível verificar comparando as Figuras 5 e 6 que os valores de *Threshold* são diferentes (alguns superiores aos que o mês de janeiro), uma vez que o comportamento do funcionário 1 passa por alterações conforme o passar do tempo. É importante ressaltar que o intervalo de observação segue até o período de 8 a 14 pois posteriormente o funcionário entrou em férias, retornando somente no mês de março.

Na Figura 7 é possível observar o comportamento diário do usuário em relação a seu *threshold*.

**Figura 7 - Relação de *Sites* acessados por *Threshold* obtidos de QA**



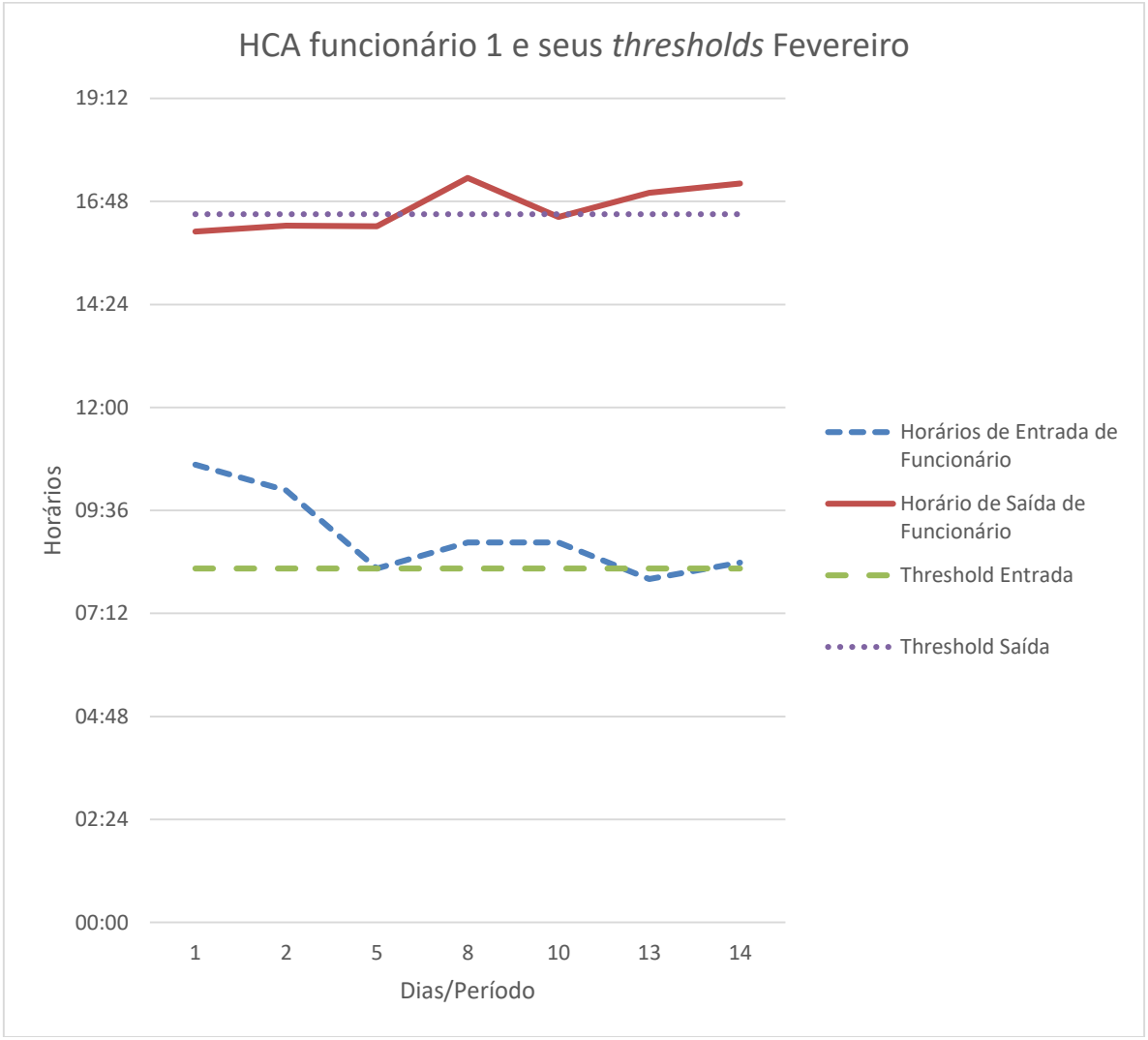

Fonte: Elaborado pelo autor.

Na Figura 7, os valores dos dias coletados referentes a soma dos sites acessados não ultrapassaram os limites de *threshold*. Dessa forma, a característica de **QA** não teve anormalidades, não somando valores ao ranking.

A partir do mês de fevereiro, é possível averiguar dois dados que dependiam da observação do mês de janeiro. Neste mês, já é possível contar com os dados de intervalo de acesso entre domínios, compondo a característica **IA,** e com o horário máximo de entrada e saída por meio das associações e desassociações dos equipamentos aos AP, compondo a característica **HCA**.

A partir da observação do mês anterior o valor de *threshold* de IA é de 11 segundos, assim como o valor de HCA fica entre 08:15 e 16:30, como pode-se observar na Figura 8 e Figura 9 respectivamente.

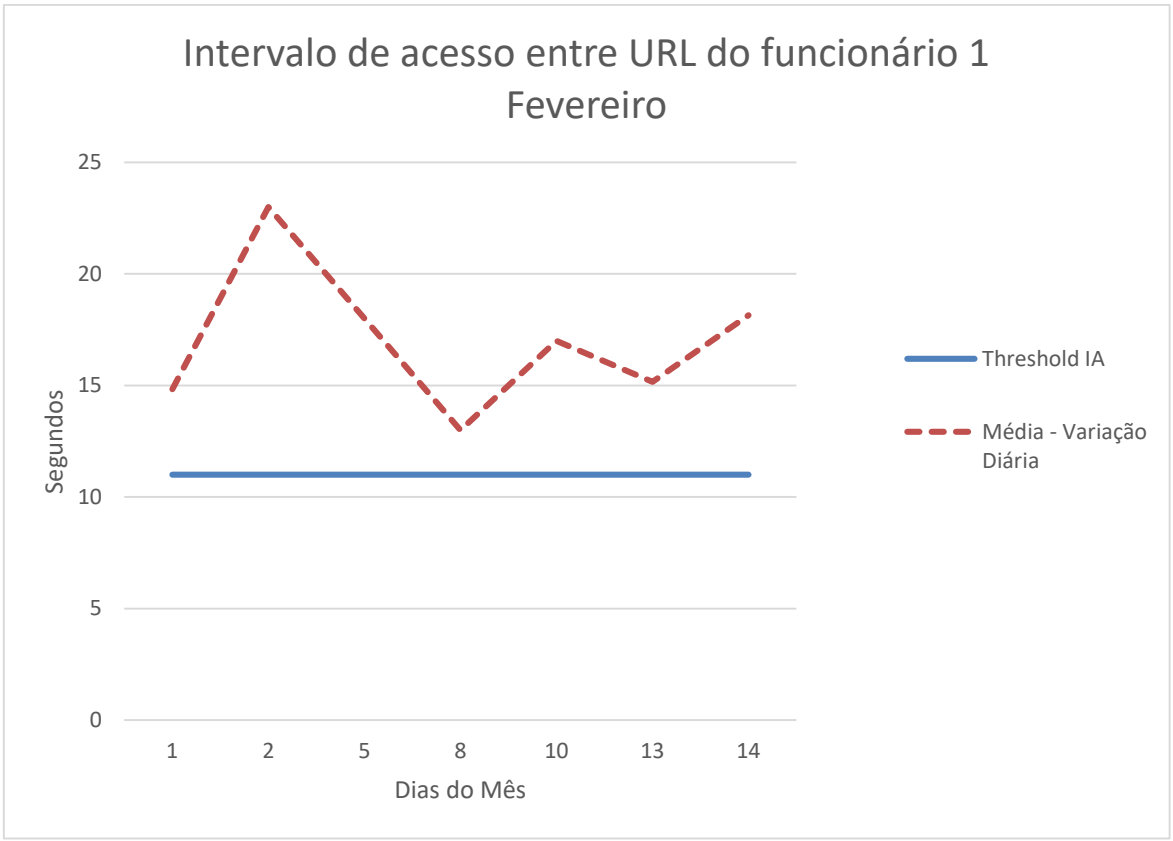
**Figura 8 - Horários Comum de Acesso (HCA) do funcionário 1 baseado em associações e desassociações de AP**




Fonte: Elaborado pelo autor.

Pode-se observar na Figura 8 que em alguns dias o *threshold* de horário de entrada e saída teve seu *match* alcançado. No ranking, esses dias que apresentam horário fora do *threshold* definido, ou seja, que apresentaram *match* recebem o valor 2.

**Figura 9 - Intervalo de acesso entre URL (IA)**



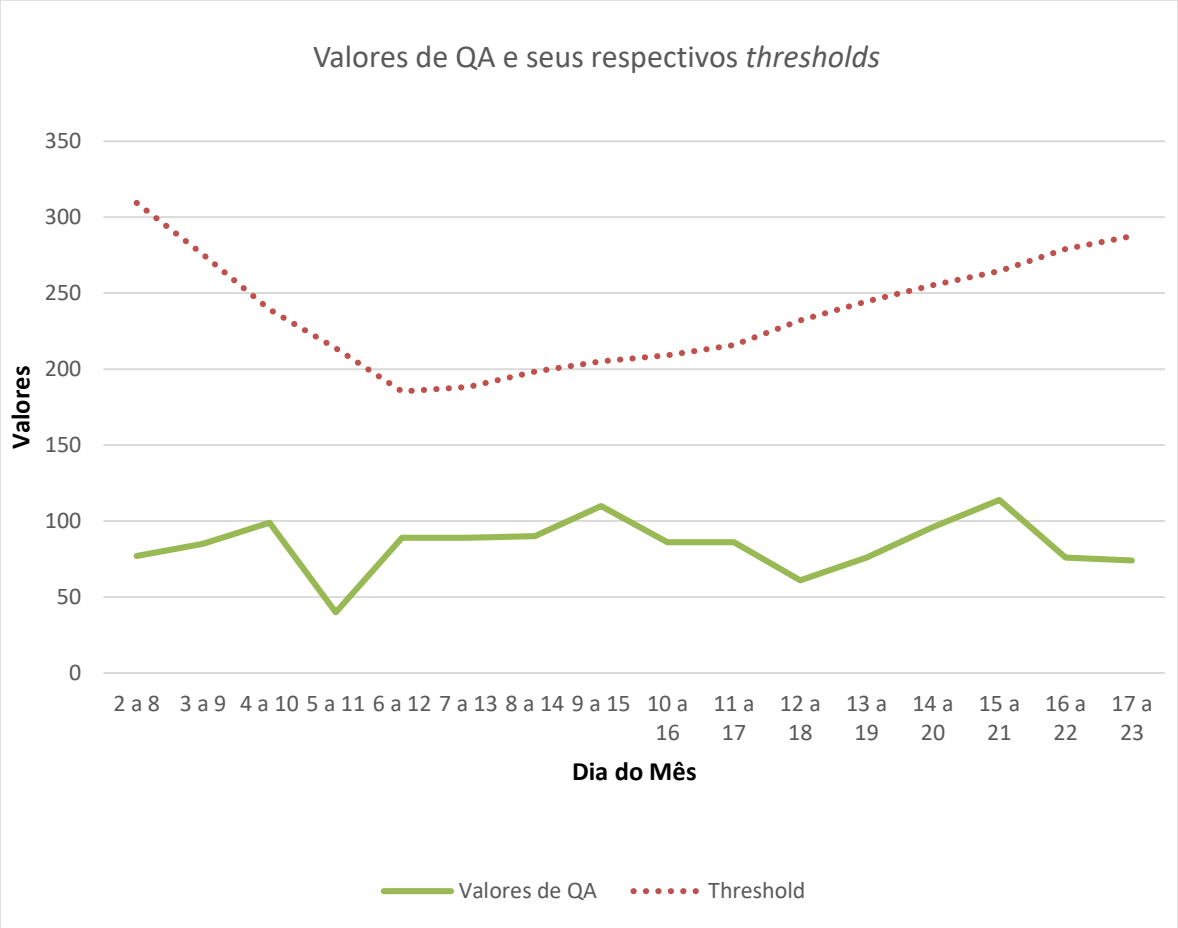

Fonte: Elaborado pelo autor.

Na Figura 9 pode-se observar o *threshold* calculado a partir do mês anterior (janeiro) em comparação com as médias de IA dos dias 1 a 14. É possível verificar que não houveram valores fora da linha de *threshold*, não apresentando *match* em nenhum dia observado. Portanto, essa característica para os dias observados no mês de fevereiro não apresentou divergências que caracterizem a não autenticidade do funcionário.

A partir dos dados obtidos, com exceção de alguns dias de HCA, não houveram valores que atingiram a linha de *threshold.* Dessa forma, para o mês de fevereiro o *match* de duas características não aconteceu e em momento algum o ranking ultrapassou o valor de máximo de 6. Assim, é possível concluir que o funcionário 1 no mês de fevereiro era autêntico, não sendo necessária nenhuma medida para sua remoção da rede.

No mês de março, é possível verificar valores de comportamento do mesmo funcionário variando de acordo com os dias observados na Figura 10.

**Figura 10 - Valores de QA e os respectivos *thresholds* do Mês de Março do funcionário 1**




Fonte: Elaborado pelo Autor.

Como demonstra a Figura 10, não houveram valores que apresentassem *match* nos dias observados. Dessa forma, não houve dia que o funcionário apresentou comportamento não autêntico.

Ainda no mês de março, é possível observar que os dados de **IA** e **HCA** ficaram entre 11 segundos, 07:05 e 17:15 respectivamente, constando somente um dia observado com valores fora do padrão na característica **HCA,** apresentando *match* com o valor de *threshold*.

Ao final do mês, observa-se na Tabela 8 o ranking do qual compõe os valores apresentados pelo funcionário 1, ao final demonstrando seu status de autêntico ou não.

**Tabela 8 - Ranking do Mês de Março do funcionário 1**

| Dia do Mês | *Match* IA | *Match* QA | *Match* HCA | Soma ranking | Status do usuário |
|---|---|---|---|---|---|
| **1** | 0 | 0 | 0 | 0 | Autêntico |
| **2** | 0 | 0 | 0 | 0 | Autêntico |
| **3** | 0 | 0 | 0 | 0 | Autêntico |
| **4** | 0 | 0 | 0 | 0 | Autêntico |
| **5** | 0 | 0 | 0 | 0 | Autêntico |
| **6** | Ausente | Ausente | Ausente | Ausente | Ausente |
| **7** | 0 | 0 | 0 | 0 | Autêntico |
| **8** | 0 | 0 | 0 | 0 | Autêntico |
| **9** | Ausente | 0 | 0 | Ausente | Ausente |
| **10** | Ausente | Ausente | Ausente | Ausente | Ausente |
| **11** | Ausente | Ausente | Ausente | Ausente | Ausente |
| **12** | Ausente | Ausente | Ausente | Ausente | Ausente |
| **13** | 0 | 0 | 0 | 0 | Autêntico |
| **14** | 0 | 0 | 0 | 0 | Autêntico |
| **15** | 0 | 0 | 0 | 0 | Autêntico |
| **16** | 0 | 0 | 0 | 0 | Autêntico |
| **17** | 0 | 0 | 0 | 0 | Autêntico |
| **18** | Ausente | Ausente | 0 | Ausente | Ausente |
| **19** | Ausente | Ausente | 0 | Ausente | Ausente |
| **20** | 0 | 0 | 0 | 0 | Autêntico |
| **21** | 0 | 0 | 0 | 0 | Autêntico |
| **22** | 0 | 0 | 0 | 0 | Autêntico |
| **23** | 0 | 0 | 0 | 0 | Autêntico |
| **24** | 0 | 0 | 0 | 0 | Autêntico |
| **25** | Ausente | Ausente | Ausente | Ausente | Ausente |
| **26** | Ausente | Ausente | Ausente | Ausente | Ausente |
| **27** | 0 | 0 | 0 | 0 | Autêntico |
| **28** | 0 | 0 | 0 | 0 | Autêntico |
| **29** | 0 | 0 | 0 | 0 | Autêntico |
| **30** | **0** | **0** | **2** | **2** | **Autêntico** |
| **31** | 0 | 0 | 0 | 0 | Autêntico |

Fonte: Elaborado pelo autor.

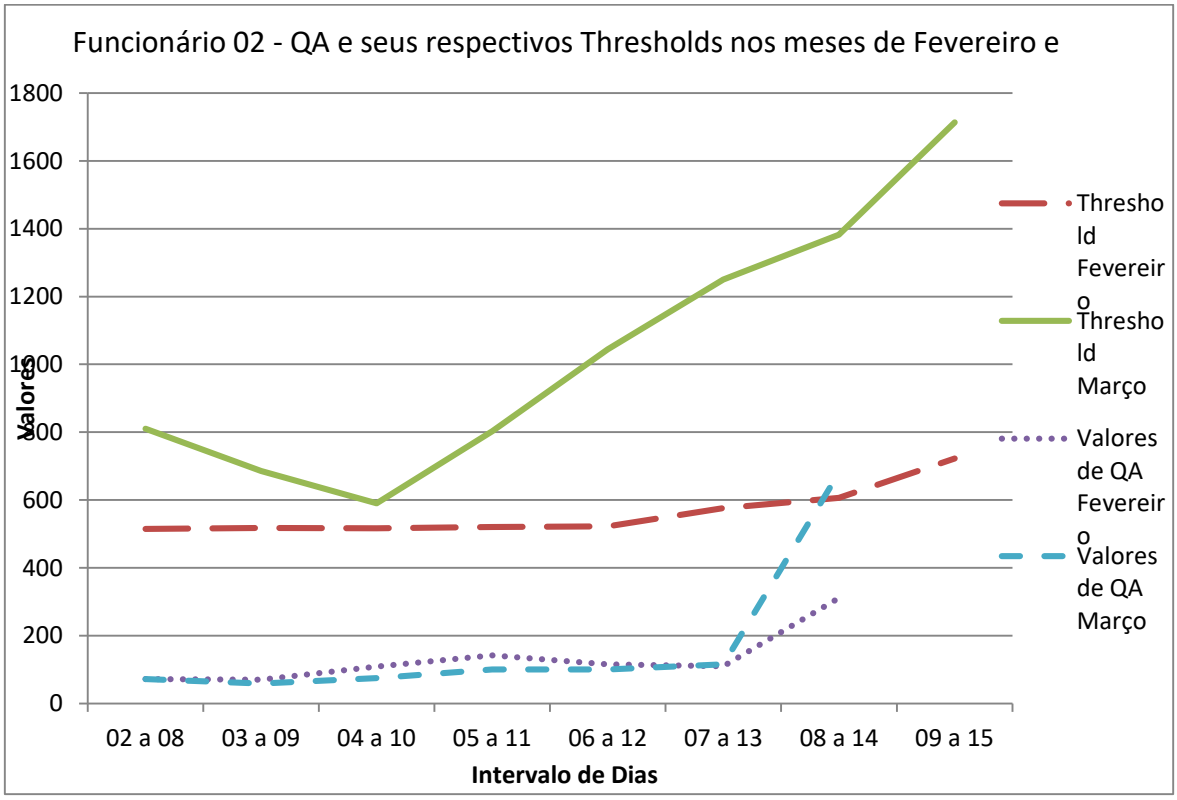
A Tabela 8 representa a geração de ranking do funcionário 1 no mês de março. Cada coluna representa uma característica observada. Caso o valor dos *matches* de pelo menos duas características ultrapasse o valor de *threshold* considerado para o dia por meio das métricas previamente apresentadas, o funcionário daquele período

é considerado não autêntico. Pode-se observar que no dia 30 a característica **HCA** apresenta um *match.* Dessa forma, é atribuído o valor 2 para a soma no ranking final. Como esse valor é menor que o limite de 6, o funcionário 1 ainda é considerado autêntico. Nos demais dias, não são constatados *matches*.

## 5.2 Funcionário 2 – Dados Amostrais

Seguindo a mesma lógica aplicada ao funcionário 1, o funcionário 2 também possui seus próprios valores de *threshold* nas características observadas. De acordo com a Figura 11, observa-se que não há *match* com o *threshold* de QA. Similar, na na Figura 12, verifica-se que também não aconteceram *matches*. Dessa forma, o usuário é considerado autêntico nos três meses de observação. É importante ressaltar que o funcionário 2 registrou ações no mês de março até a data do dia 16 e fevereiro até a data do dia 15. O funcionário em questão precisou se ausentar nos demais dias devido a problemas de saúde.

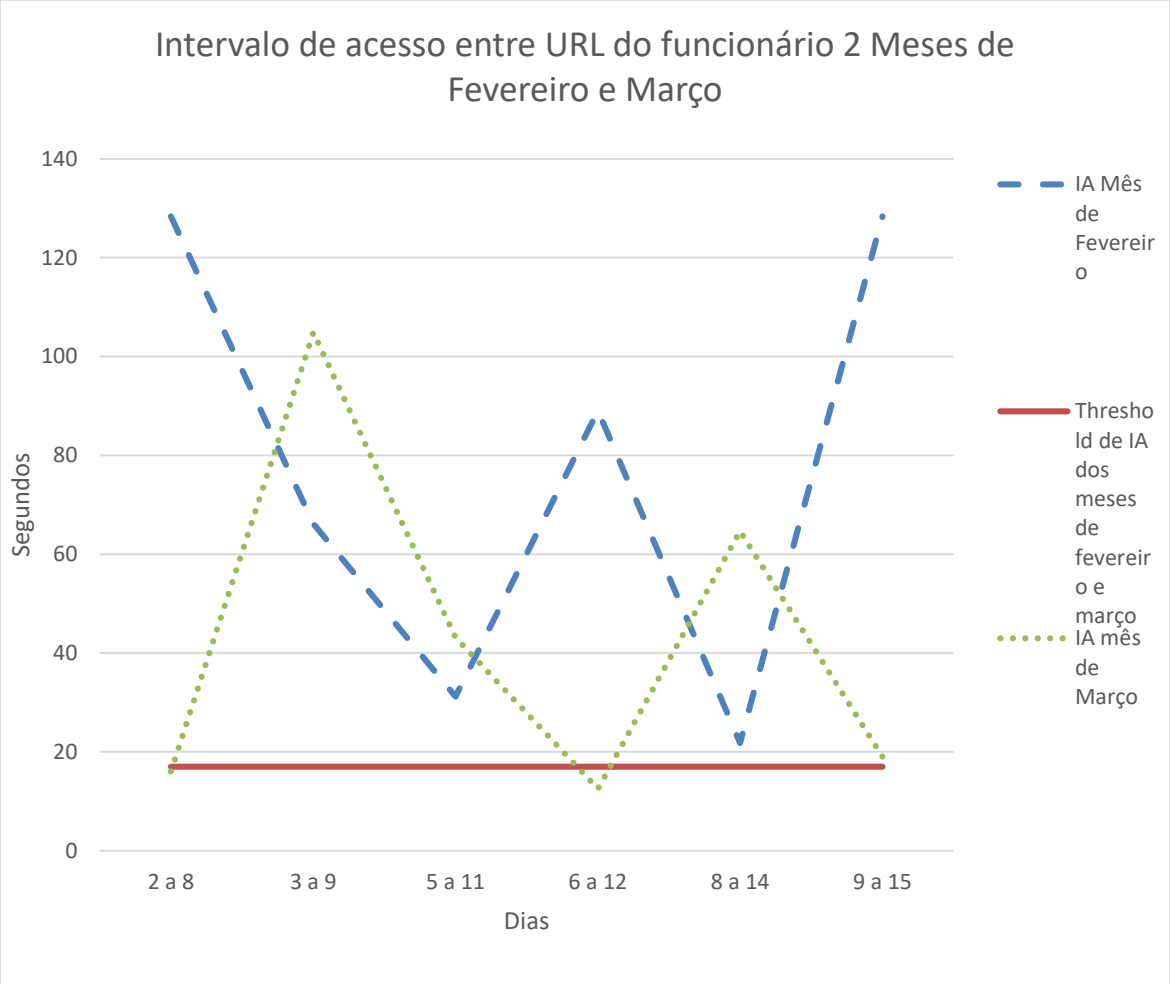
**Figura 11 - Valores de QA e os respectivos *thresholds* dos meses de fevereiro e março do funcionário 2**




Fonte: Elaborado pelo autor.

Observa-se na Figura 11 os valores de *threshold* de janeiro, fevereiro e março de QA, junto aos valores obtidos no mês de fevereiro e março. Não se verifica os valores de QA de janeiro ou seu *threshold* na figura, uma vez que não é possível realizar o cálculo de ranking neste mês, já que depende de observações anteriores, portanto é omitido nesta figura os valores de QA e *threshold*.

**Figura 12 - Intervalo de acesso entre URL (IA) do funcionário 2**

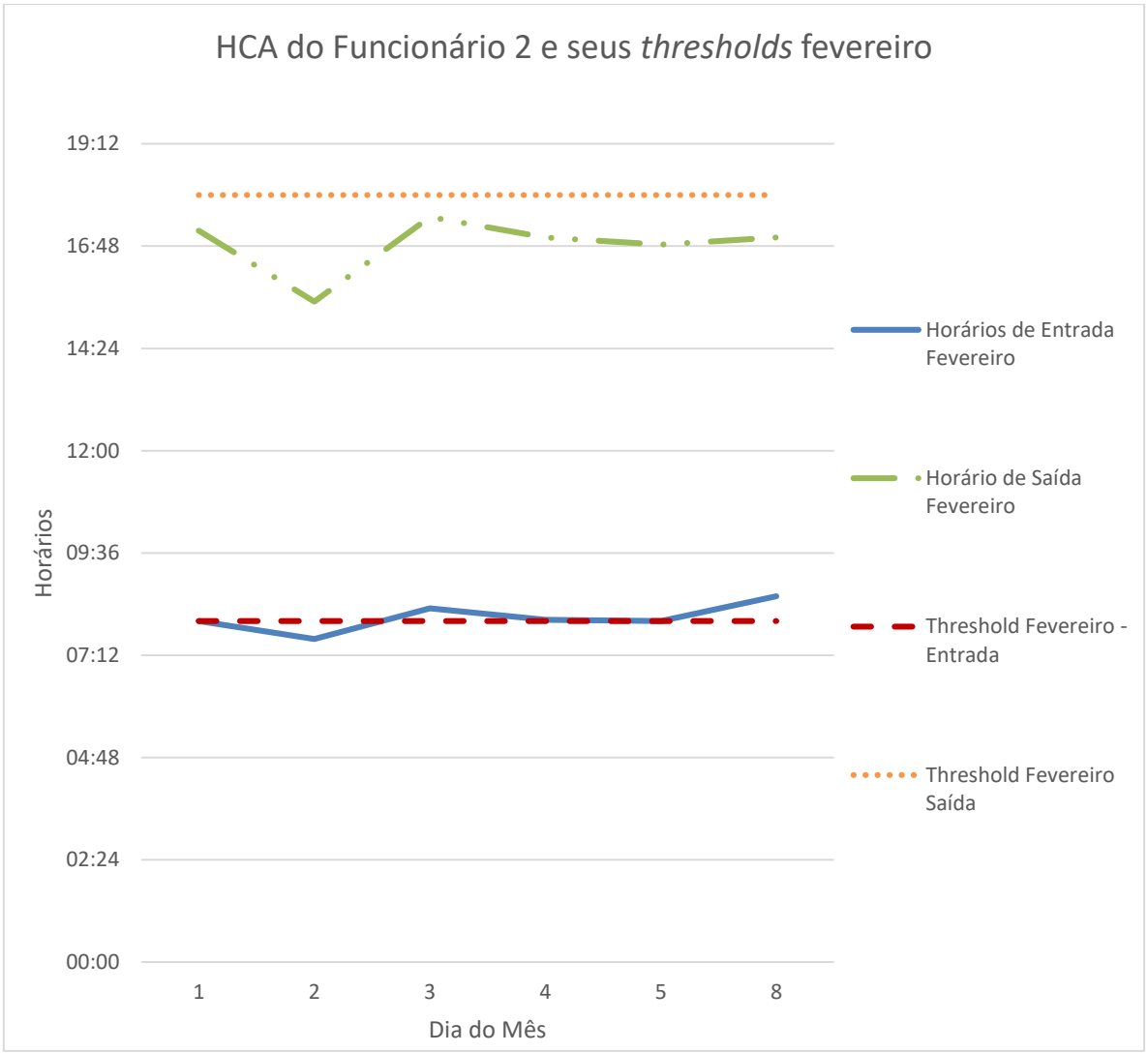



Fonte: Elaborado pelo autor.

É possível observar na Figura 12 que no intervalo de 06 a 12, especificamente no dia 13 de março, o valor de intervalo de navegação cruza o *threshold,* apresentando um *match*. Portanto neste dia, o ranking do funcionário 2 teve um acréscimo de 2 em seu valor. Porém, como apresentado na Figura 11, não houve outro valor fora do padrão. Portanto, apesar da apresentação do *match* nesta data, o usuário ainda é considerado autêntico pois a soma dos valores de seu ranking ainda é inferior a 6.

Já na Figura 13, é possível averiguar os valores de HCA e seus respectivos *thresholds*.

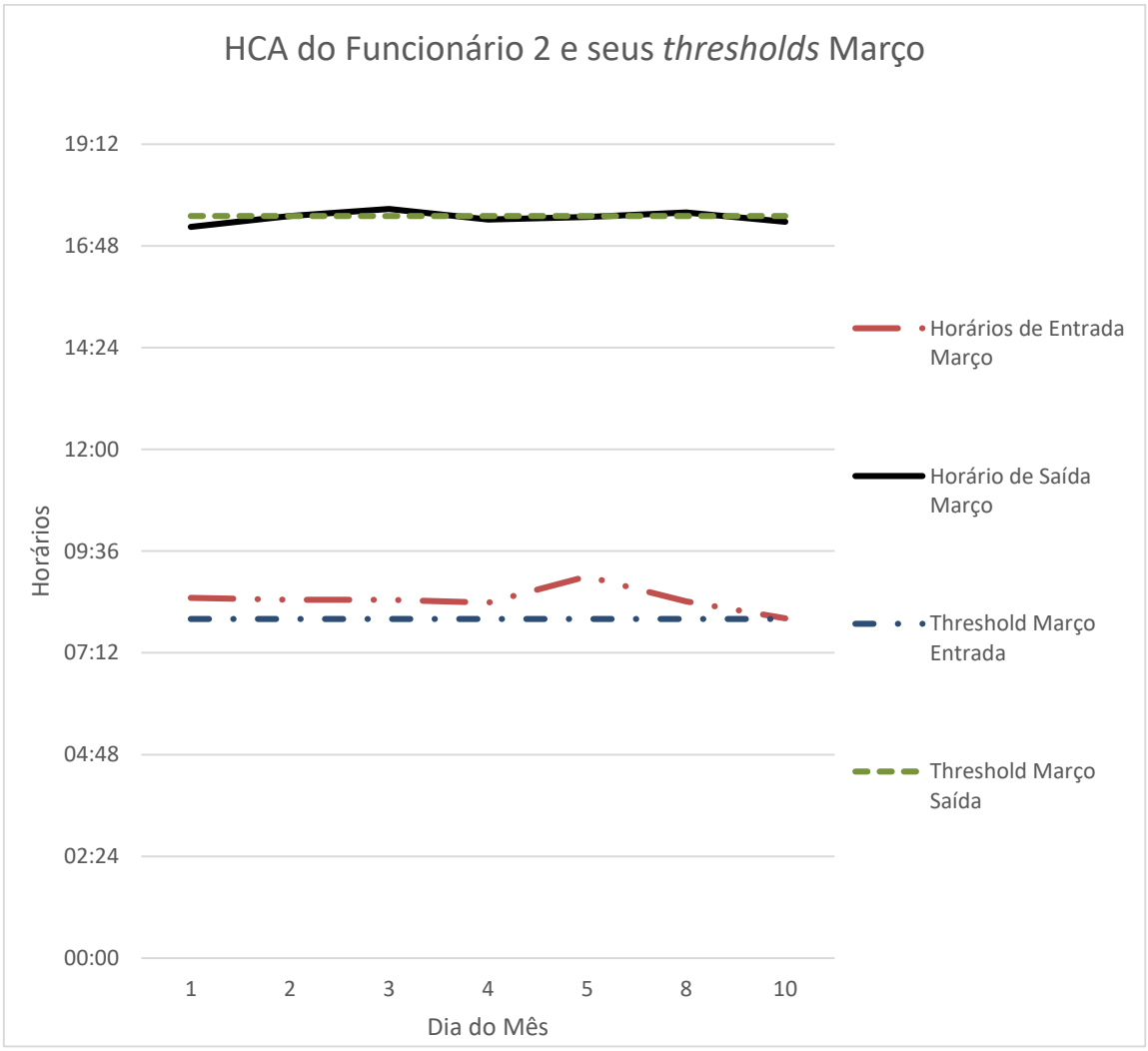
**Figura 13 - Horários Comum de Acesso (HCA) do funcionário 2 baseado em associações e desassociações de AP no mês de Fevereiro**




Fonte: Elaborado pelo autor.

Nas Figuras 13 e 14 observa-se que poucos horários apresentam divergência com os observados no mês anterior apresentando *match* com seus valores de *threshold*. Como a soma das demais características nos dias que o *match* é verificado apresenta valor inferior a 6, o usuário ainda é considerado autêntico. Tanto nas Figuras 13 e 14 não é possível verificar valores após o dia 8 e 10 de cada mês. Nos próximos dias que o usuário esteve presente na companhia não foi registrada a presença de conexão de seu aparelho celular.

**Figura 14 - Horários Comum de Acesso (HCA) do funcionário 2 baseado em associações e desassociações de AP no mês de Março**



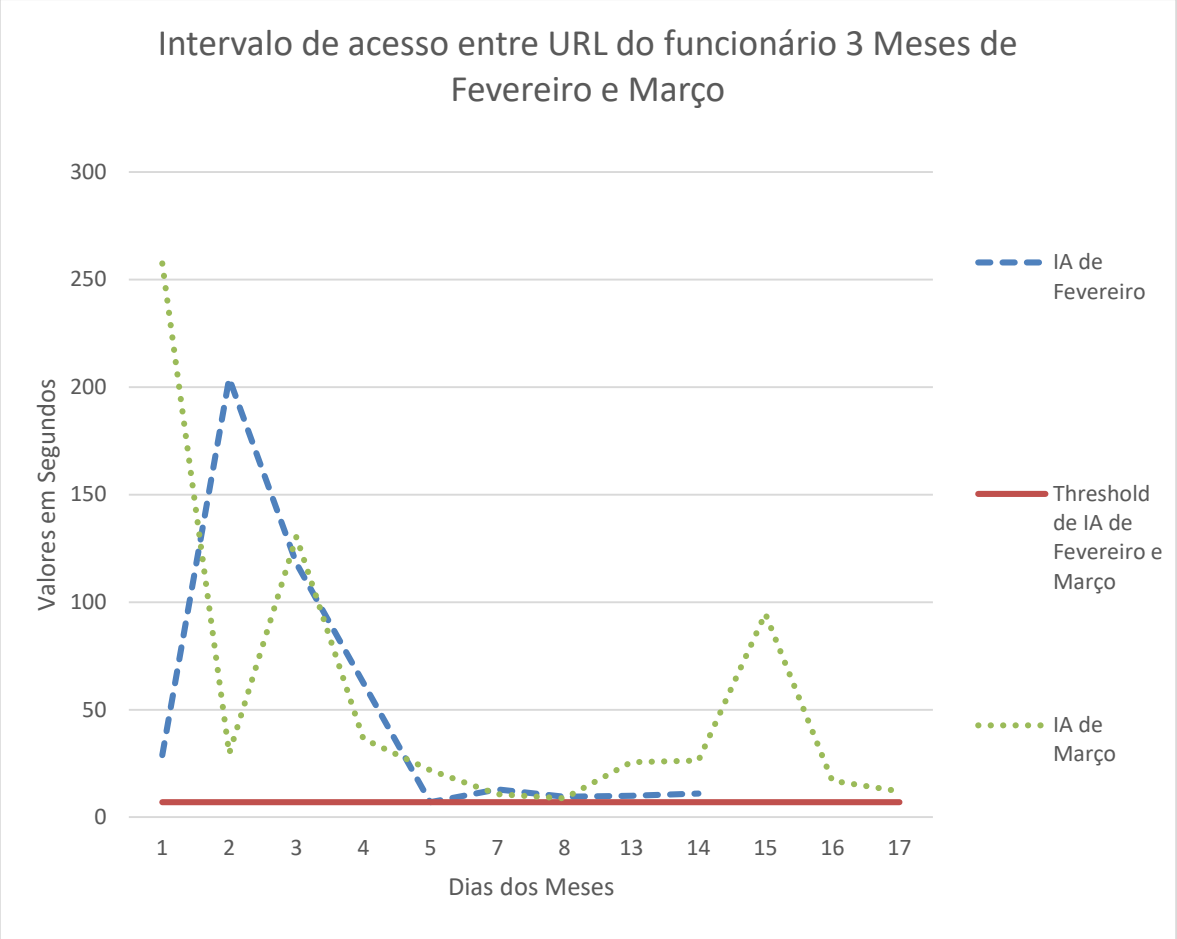

Fonte: Elaborado pelo autor.

## 5.3 Funcionário 3 – Dados Amostrais

Da mesma forma como observado nos funcionários 1 e 2, o funcionário 3 apesar de possuir valores de características e *threshold* diferentes, também não obteve valores que apresentem uma soma igual ou superior a 6, conforme observa-se nas Figuras 15, 16 e 17. O funcionário 3 teve sua observação até os dias 14 e 17 dos meses de fevereiro e março respectivamente pois sua característica de serviço

demanda saídas da companhia por longos períodos. Dessa forma, as datas ausentes nas figuras representam os dias de trabalho em campo do funcionário.

**Figura 15 - Intervalo de acesso entre URL (IA) do funcionário 3 nos meses de fevereiro e março**

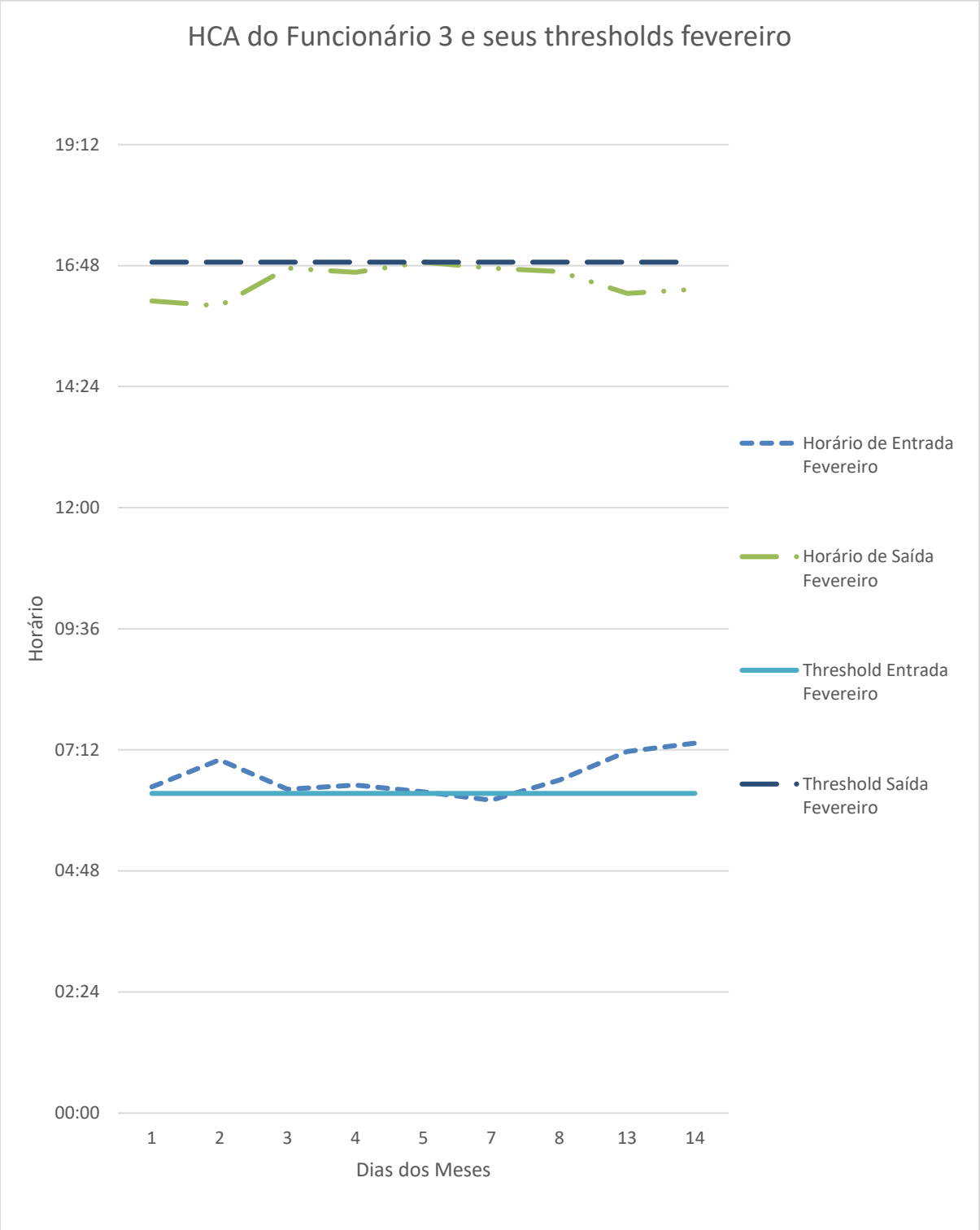



Fonte: Elaborado pelo autor.

A Figura 15 representa os valores de IA do funcionário 3. Pode-se observar que alguns dias a linha de *Threshold* é atingida pelo IA, tanto em fevereiro como em março, nos dias 05 e 08 respectivamente. Como os valores apresentados são inferiores ao *threshold* estabelecido, o funcionário ainda apresentou comportamento autêntico e não teve o valor 2 somado a seu ranking.

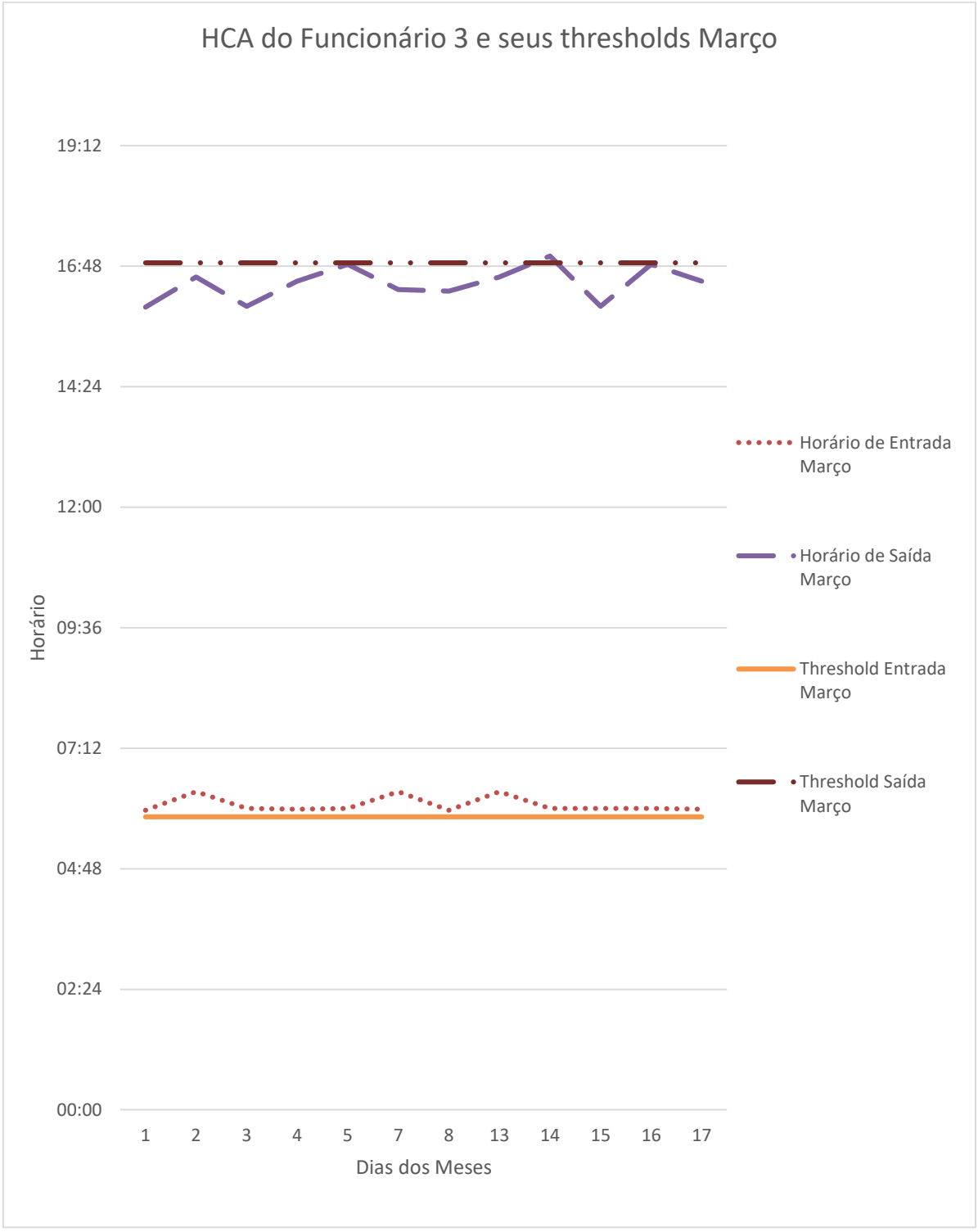
Já nas Figuras 16, 17 e 18, é possível observar que, similar as amostras dos funcionários 1 e 2, os dados de HCA e QA raramente ultrapassam os valores de *threshold* definidos pelas métricas propostas.

**Figura 16 - Horários Comum de Acesso (HCA) do funcionário 3 baseado em associações e desassociações de AP no mês de Fevereiro**



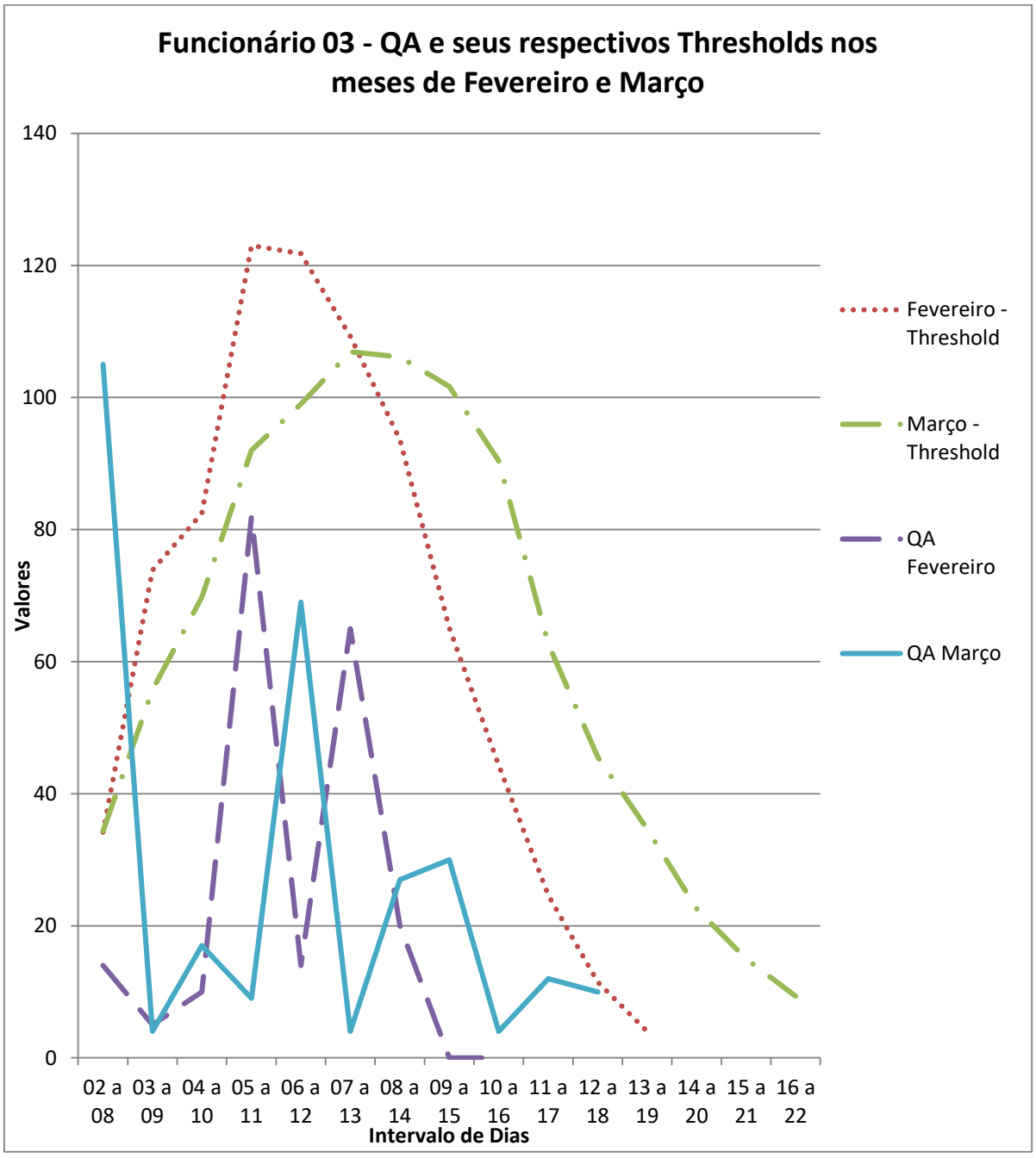

Fonte: Elaborado pelo autor.

**Figura 17 - Horários Comum de Acesso (HCA) do funcionário 3 baseado em associações e desassociações de AP no mês de Março**




Fonte: Elaborado pelo autor.

**Figura 18 - Valores de QA e os respectivos *thresholds* dos meses de fevereiro e março do funcionário 3**



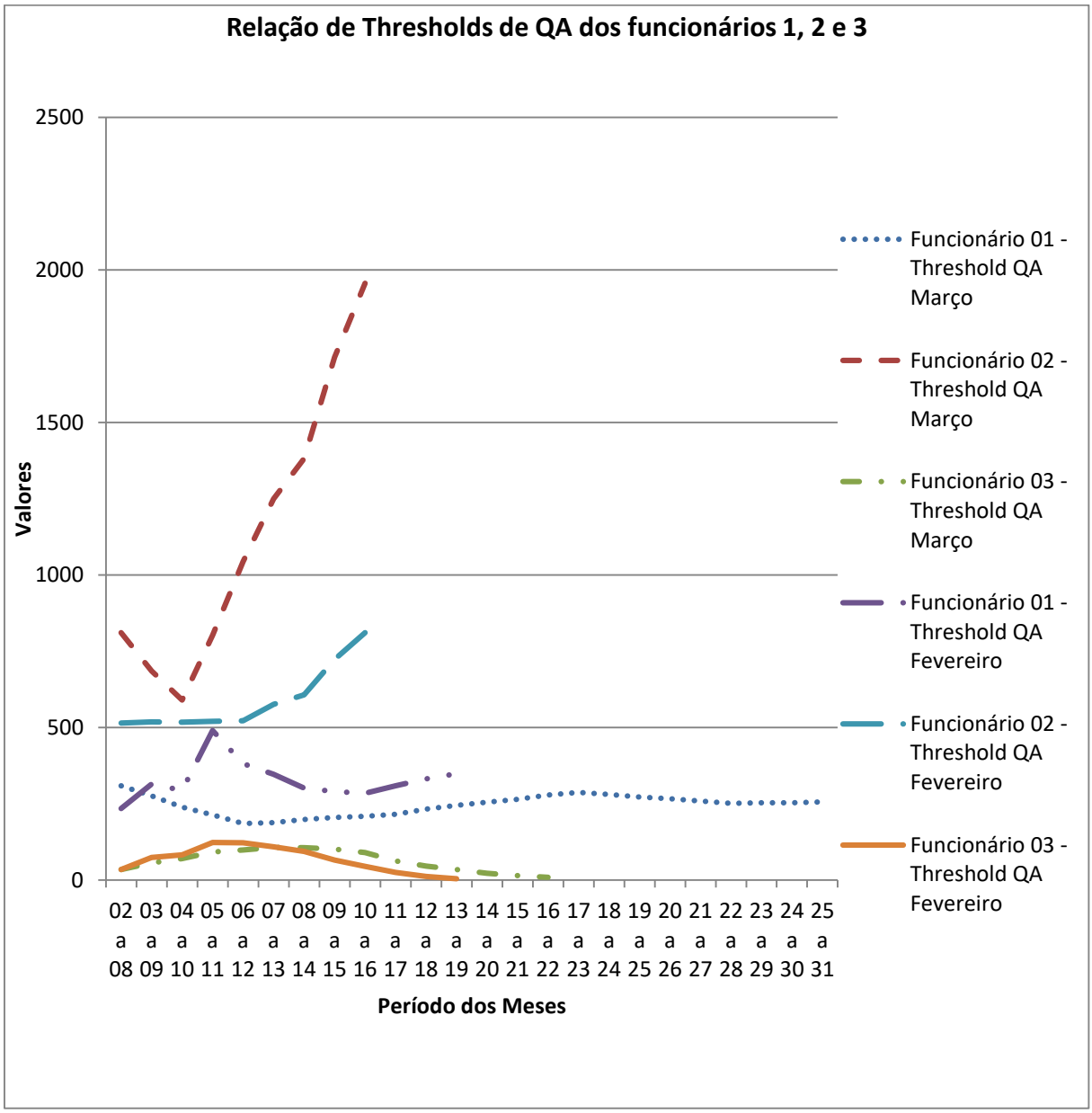

Fonte: Elaborado pelo autor.

Conforme observado nas Figuras 15, 16, 17 e 18, o funcionário 3 não teve nenhum período onde seu ranking seja igual ou superior o valor de 6. Dessa forma, mesmo com alguns *matches*, o usuário é considerado autêntico em todos os dias observados.

## 5.4 Detecção de Usuário não autêntico

Conforme apresentado nas seções anteriores, os funcionários observados possuem valores de *threshold* diferentes para as características observadas. Ao cruzar estas informações, é possível verificar a grande diferença dos *thresholds* destes funcionários, especialmente nas características de **HCA** e **QA**. Isso deve-se principalmente a atividade fim de cada funcionário, uma vez que cada um exerce atividades diferentes, com demandas diferentes. Apesar do horário de entrada e saída ser comum a todos os funcionários, cada setor tem a liberdade de flexibilizar os horários de seus respectivos funcionários. Dessa forma, é possível que os funcionários observados possuam diferentes horários de entrada e saída. É importante ainda ressaltar que a faixa etária dos funcionários observados é bastante heterogênea. Os dois primeiros funcionários são mais velhos, em torno de 60 anos de idade, enquanto o funcionário 3 é um funcionário mais novo, por volta de 30 anos. A diferença de idade pode impactar bastante nas características observadas, principalmente na facilidade de lidar com as novas tecnologias, principalmente relacionadas a acesso a sites e sistemas online. Este trabalho não realiza um estudo baseado na idade, porém é importante ressaltar essa característica, uma vez que pode contribuir para a diferenciação de comportamento entre os perfis. As Figuras 18 e 19 demonstram a diferença dos *thresholds* das características citadas.

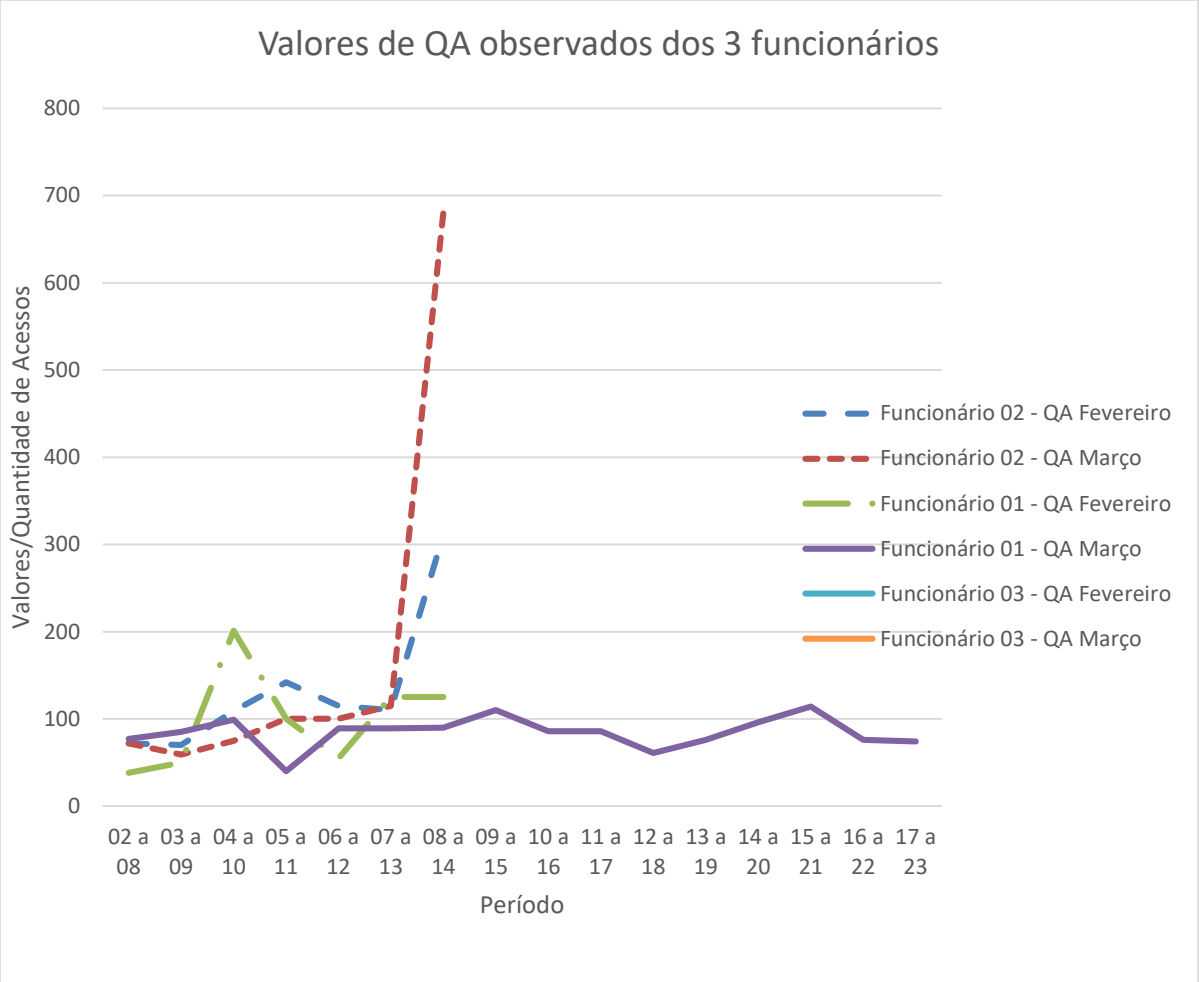
**Figura 19 - Threshold de QA dos 3 funcionários observados em relação aos meses de fevereiro e março.**




Fonte: Elaborado pelo autor.

Nas Figuras 19 e 20 é possível verificar a grande diferença dos *thresholds* de cada funcionário. Os valores de *threshold* observados são diretamente verificados das ações de cada funcionário no período observado. É possível notar que estes são bastante diferentes entre si, e cada funcionário com seu respectivo *threshold* possui seu próprio padrão de comportamento, sua própria identidade comportamental na rede.

**Figura 20 - Valores de QA dos funcionários obtidos por período**

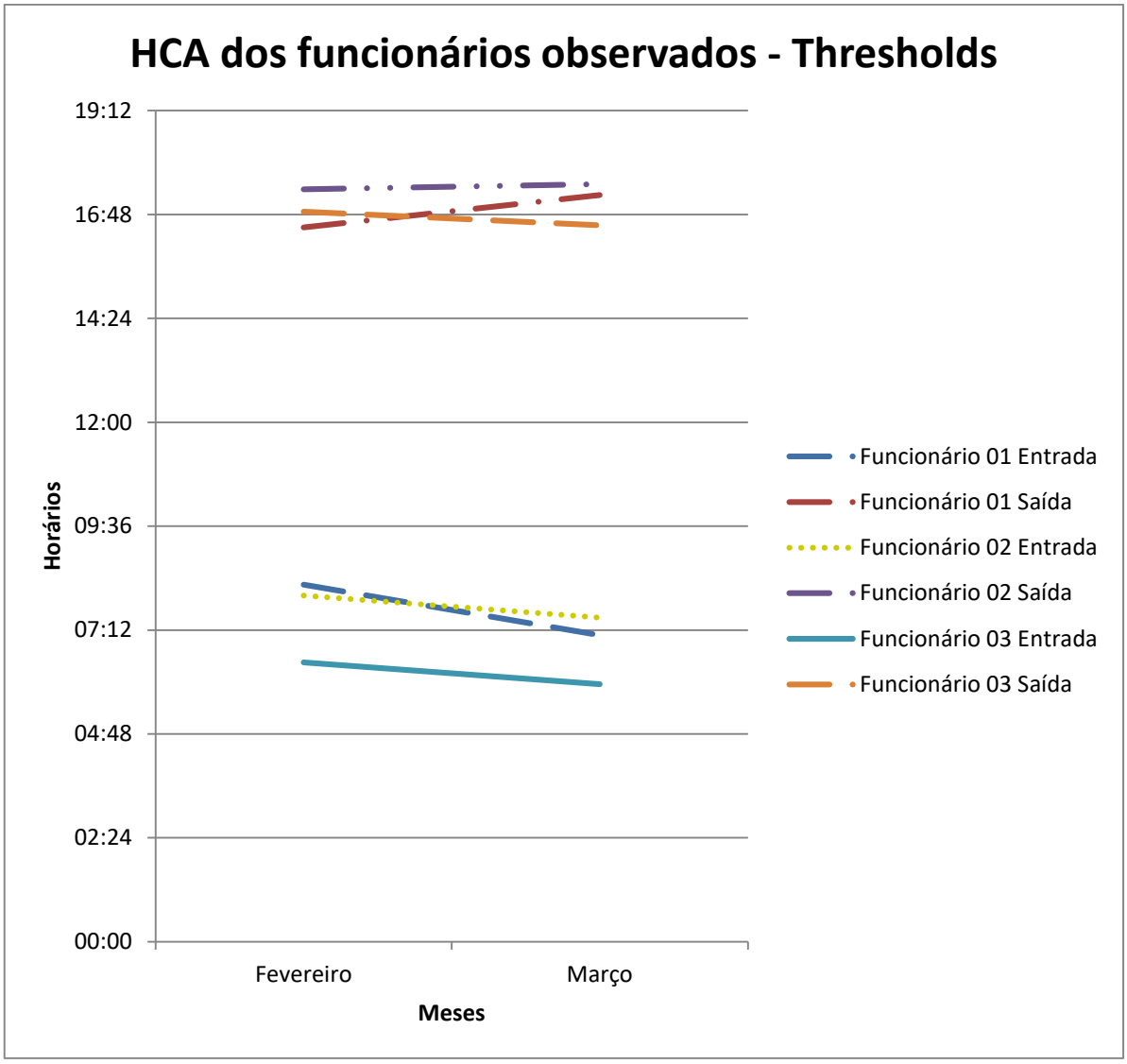



Fonte: Elaborado pelo autor.

É possível observar na Figura 20 os diferentes valores de *threshold* dos funcionários observados no período de fevereiro e março. Ainda é possível verificar que um funcionário apresenta mudanças de comportamento no decorrer das observações. Essas mudanças podem ocorrer diante de mudanças de fluxo de trabalho, demandas emergenciais ou até mesmo problemas pessoais. Mesmo diante dessas mudanças, cada comportamento é único e é possível observar suas diferenças ao cruzar com os outros funcionários observados. Caso os valores de QA de um funcionário fossem usados com os *thresholds* de outro, a ocorrência de *matches* aumentaria significativamente.

Na Figura 21, observa-se também os diferentes *thresholds* de HCA.

**Figura 21 - *Thresholds* de HCA dos 3 funcionários observados**

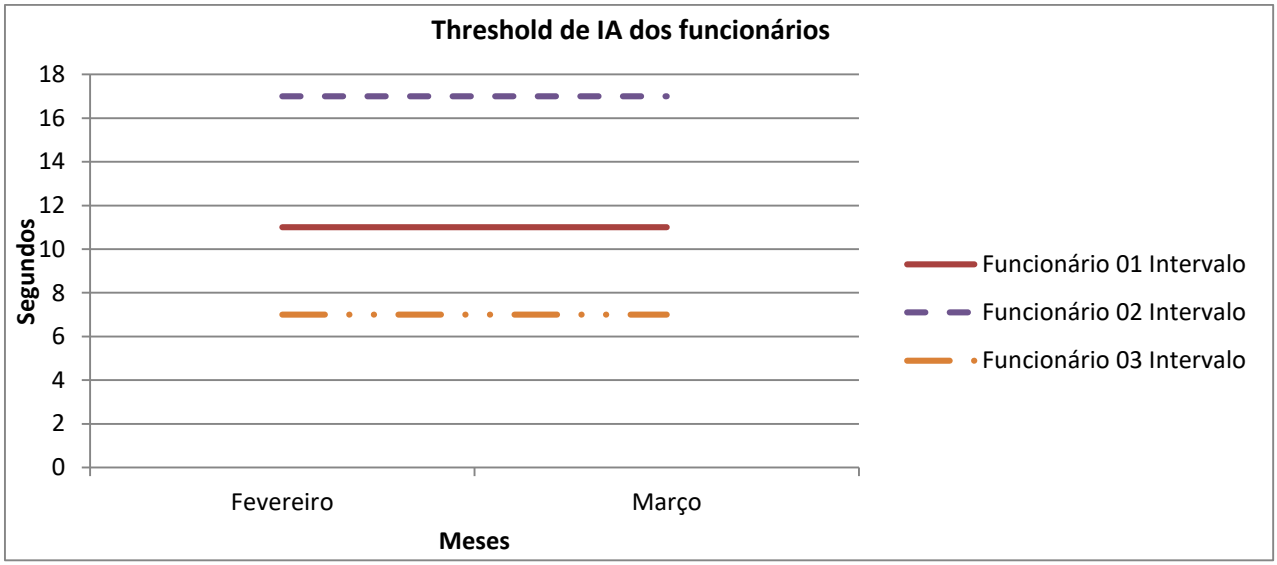



Fonte: Elaborado pelo autor.

Verifica-se o mesmo comportamento das Figuras 19 e 20, na Figura 21: A diferença de horário entre os funcionários. Ao cruzarmos os *thresholds* dos funcionários com os diferentes HCA que estes possuem, é possível observar que, de forma similar aos dados de *threshold* de QA, que a quantidade de *matches* também aumentaria. Isso pode ser observado comparando-se as Figura 8, Figura 13 e 15 com a Figura 21.

Ainda na Figura 22, também verifica-se um comportamento similar na IA, que apresenta diferentes *thresholds* de intervalo de acesso entre os funcionários.

**Figura 22 - *Threshold* de IA dos funcionários 01, 02 e 03**




Fonte: Elaborado pelo autor.

Ao verificar os *thresholds* de cada funcionário na Figura 22, é possível concluir que os valores em segundos de IA são diferentes, e trazem o comportamento específico de cada funcionário, especialmente quando se comparam a Figura 9, Figura 12 e Figura 15. Novamente, é importante destacar a diferença de perfis dos funcionários observados, suas diferentes atividades na companhia e a faixa etária de cada um. Essas são características que podem interferir no comportamento de cada funcionário, tendo intervalos de acesso mais rápido por terem uma facilidade maior com sistemas computacionais, ou até mesmo uma facilidade com a leitura da tela do computador.

Ao final, é possível concluir que ao menos duas das 3 características dos funcionários, ao cruzar seus dados com seus respectivos *thresholds* observados, os valores apresentam *match* e somam a quantidade suficiente no ranking para sua consideração de não autênticas. Desta forma, o funcionário categorizado como não autêntico é automaticamente excluído de acesso a rede utilizando as propriedades do protocolo 802.1X. Dessa forma, verifica-se que os dados apresentados das características propostas, em uma situação onde um intruso assuma a credencial de um usuário pertencente na rede, logo poderá ter seu acesso bloqueado e excluído da infraestrutura até uma ação da equipe de TI responsável.

A Tabela 9 representa as quantidades de rankings que apresentam alertas “não autêntico” caso os *thresholds* dos funcionários fossem cruzados entre si.

**Tabela 9 - Relação de não autênticos ao cruzar thresholds de funcionários**

| Funcionários | *Thresholds* | Valores de ranking (fev-mar) | | |
|---|---|---|---|---|
| | | 6 | 8 | 10 |
| Funcionário 1 | Funcionário 2 | 8 | 0 | 0 |
| | Funcionário 3 | 17 | 0 | 0 |
| Funcionário 2 | Funcionário 1 | 2 | 0 | 0 |
| | Funcionário 3 | 4 | 0 | 0 |
| Funcionário 3 | Funcionário 1 | 3 | 0 | 0 |
| | Funcionário 2 | 5 | 0 | 0 |

Fonte: Elaborado pelo autor

É possível ainda observar que a Tabela 9 não apresentou valores de ranking maiores que 6 pois não foram constatadas características que somassem os demais valores. Porém, mesmo com o valor de 6 atingido, houveram várias incidências de usuários não autênticos, demonstrando o funcionamento da identificação de usuários não autênticos.

No total, a Tabela 9 representa uma análise de 112 comparações de dados obtidos com os respectivos *thresholds* cruzados entre os usuários. Houve no total um acerto de 39 verificações não autênticas. Esse valor representa um índice de 43,68% de identificação de perfis não autênticos de funcionários ao cruzar os dados de um funcionário com o *threshold* de outro funcionário. Apesar de ser um número relativamente mediano para acertos, é importante ressaltar que no momento da detecção o usuário é automaticamente removido da rede. Portanto, na situação hipotética onde funcionários fossem identificados como não autênticos, não haveriam a continuidade de seu acesso e, consequentemente, não haveriam outras checagens posteriores, uma vez que os funcionários seriam removidos da rede e não gerariam mais tráfego para análise. Ainda também, é importante ressaltar que para o ambiente ao qual a lógica proposta deste trabalho é apresentada foca exatamente na inibição de utilização de credenciais de um funcionário por outro. Mesmo apresentando um valor mediano de detecção, seria uma aplicação viável, uma vez que na ocorrência

de utilização, seu acesso seria bloqueado e logo percebida este bloqueio pelo funcionário.

Nas amostras obtidas nos meses em observação dos funcionários, nenhum apresentou valores em seus respectivos *thresholds* que indicassem falso positivo, ou seja, que somassem 6. Portanto, o índice de assertividade para definição de usuários autênticos gira próximo a 100% nas amostras e no período observado.

## 5.5 O 802.1X na detecção de comportamento anômalo

A partir do momento da detecção de não autenticidade do funcionário, uma ação é tomada utilizando o padrão 802.1X. Verificada sua não autenticidade, o funcionário é removido do grupo Autorizados, não sendo mais liberado para acesso à rede, ou seja, ele é movido para o grupo Negados.

Nos dados coletados tanto pelo *proxy* como também pelo AP, é possível verificar o *Media Access Control* (MAC) do dispositivo que está efetuando o acesso à rede. A partir do momento de sua identificação, um comando para desassociação do respectivo AP associado é enviado. A infraestrutura aplicada neste trabalho é composta por equipamentos do fabricante Cisco, que possuem seus próprios comandos para esta ação. Porém, é importante ressaltar que outros fabricantes de linha corporativa possuem também este tipo de função e são capazes de realizar a mesma operação, cada um com sua respectiva sintaxe de comando. Ainda também é importante ressaltar que *switches* com suporte ao padrão 802.1X também realizam esta ação de forma similar ao ambiente *wireless*, com a diferença vinculada ao meio físico de transmissão. O AP Cisco recebe o seguinte comando para sua remoção de intruso: “*wireless cliente mac-address <mac-do-usuario> deauthenticate”*

Assim que realizada a execução do comando, o funcionário identificado como não autentico é desconectado da rede e, como padrão da maioria dos dispositivos com conexão *wireless* ou até mesmo cabeados, o dispositivo tenta efetuar sua conexão novamente de forma automática. Ocorre que para efetuar a conexão à rede, este fornecerá novamente suas credenciais. O 802.1X realiza seu papel, verificando as credenciais fornecidas pelo equipamento do cliente não autêntico e, verifica a qual

grupo o funcionário solicitante pertence. Como o funcionário já não se encontra mais no grupo Autorizados, mas sim agora no grupo de Negados, seu acesso não é permitido e ele é automaticamente desconectado da rede em questão. O funcionário somente poderá efetuar sua conexão novamente a partir da ação de transferência de seu usuário do grupo Negados para o grupo Autorizados, ou seja, somente via ação conjunta com o departamento de TI.

É possível verificar essa situação no exemplo demonstrado na seção 5.4, da Tabela 9, que utiliza padrões de funcionários com as métricas obtidas de outros. É possível notar que em vários casos houveram incidentes de identificação de usuário não autêntico. Nestes exemplos, o 802.1X verificaria que o funcionário não pertence mais ao grupo Autorizados, removendo-os em conjunto com a ação de desassociação do AP e negando sua entrada na rede.

# 6. CONCLUSÃO

A identificação e reconhecimento de autenticidade de um usuário é um desafio que a área de segurança enfrenta constantemente. No ambiente em estudo deste trabalho, há um grande problema de uso indevido de credenciais de terceiros, ou seja, funcionários utilizam as credenciais de outros, sejam estes seus colegas, companheiros de trabalho e até gestores, mesmo com toda a orientação e recomendação da equipe de TI sobre os problemas de segurança que tal prática possa trazer. A partir de dados de serviços relativamente comuns disponíveis em uma estrutura de rede, verifica-se a possibilidade de identificação de alguns padrões de comportamento.

Dados como navegação de sites, intervalo de acesso entre domínios, associação e desassociação de equipamentos *wireless* são informações que alguns serviços de rede relativamente comuns nas empresas podem nos oferecer. Diante destes, é possível relacionar determinados dados e verificar padrões de comportamento de usuários, sendo possível diferenciá-los em situações de observação.

O serviço de *proxy* bem como a infraestrutura de rede *wireless* disponível no cenário deste trabalho, possibilita a captura destes dados. Com seus respectivos cruzamentos, é possível determinar padrões de comportamento para os usuários pertencentes a rede.

Neste trabalho ainda, com o objetivo de levar em consideração os padrões dos últimos 5 dias de comportamento do usuário, especificamente na quantidade de *sites* acessados, é adotada uma média ponderada, a qual possibilita que esta característica considerem o nível de trabalho do funcionário, pois este, como observado neste trabalho, sofre alterações no decorrer do tempo.

Dados como intervalo de acesso entre domínios e horários de associação e desassociação também trazem valores importantes para a definição de autenticidade do funcionário, uma vez que cada um deles apresenta um comportamento diferente.

A utilização da lógica proposta para definição de autenticidade de usuário na infraestrutura correspondente traz um índice de acerto considerável. É possível por

meio da análise dos dados capturados, determinar informações necessárias para autenticidade dos usuários.

Tanto os dados de DN-*Proxy*, como os dados de DN-AP na infraestrutura em estudo foram primordiais para demonstrar os padrões e comportamentos de cada usuário na rede.

. A situação de bloqueio só é verificada a partir do momento que são aplicadas métricas de um funcionário com dados de outro, exatamente para a validação de uso de um usuário invasor com posse das credenciais não pertencentes a ele. Neste cenário, é possível verificar que muitos dos valores apresentados nos dias são divergentes e o ranking logo atinge o valor máximo para exclusão do usuário do grupo de permitidos para acesso a infraestrutura. Isso ocorre pois fica constatado que o padrão de comportamento de um usuário para outro varia em relação as características observadas. Essa variação deve-se não somente ao perfil de cada um, mas sim também pela função de trabalho que cada usuário exerce na companhia. Ainda também o fator de idade e maior familiaridade com sistemas eletrônicos afetam diretamente os valores das características observadas.

Dessa forma, conclui-se que a lógica proposta pode ser utilizada na infraestrutura da companhia, trazendo uma maior segurança e principalmente, eliminando o problema de uso de credenciais de outros usuários com informações sensíveis, forçando os funcionários da companhia a não fornecerem seus acessos para outros colegas, sob risco de terem seus *logins* bloqueados devido a alterações de comportamento no dia a dia de trabalho. Ainda também, importante ressaltar o papel do 802.1X. Junto a solução proposta, o padrão permite que os funcionários não autênticos sejam removidos da rede a qual estão conectados. É por meio do 802.1X que, após a detecção de um usuário não autêntico, este permaneça desconectado e não consiga mais ingressar na rede corporativa até uma ação de terceiros, no caso os responsáveis pela administração de rede e segurança.

É importante ainda ressaltar que o trabalho se limita ao número de características observadas, além de só exercer uma verificação das características por dia. Muitas vezes, esta única verificação pode não ser o suficiente para inibir o acesso indevido de algum elemento malicioso que possam vir a invadir a rede utilizando credenciais de usuários autênticos. Dessa forma, mesmo com a análise

comportamental dos elementos ingressantes na rede, o incidente de segurança pode ocorrer e causar problemas. Futuramente, a aplicação da lógica proposta junto a um intervalo de tempo menor de verificação deve ser apresentada, possibilitando que a verificação não seja efetuada somente uma vez ao final de cada dia observado, mas sim em determinados horários do dia, ou até mesmo em intervalos curtos que possam trazer maior segurança e ação ao sistema. Também, um maior número de características pode ser observado e obtido dos dados utilizados neste trabalho, tais como: informações como o percurso padrão que o funcionário passa diariamente, *sites* favoritos acessados por período, informações de carga de trabalho específicas por período, entre outras, podem trazer maior granularidade na identificação do perfil de comportamento dos funcionários presentes na companhia. Outras informações obtidas por outros serviços disponíveis na rede também podem auxiliar na formação do perfil dos funcionários, junto também a um período maior de observação, próximo a um ano. Ferramentas de identificação comportamental *web*, como *browser fingerprint*, que conseguem analisar o padrão de comportamento de usuários em seu *browser,* podem também auxiliar na identificação. Podem ainda também ser estudados outros dados, obtidos de outras plataformas disponíveis na rede desde que sejam verificadas as respectivas relevâncias e associações com as características estudas neste trabalho.

## 7. REFERÊNCIAS


ALIPOUR, Hamid *et al*. **Wireless Anomaly Detection Based on IEEE 802.11 Behavior Analysis**. Ieee Transactions On Information Forensics And Security, [s.l.], v. 10, n. 10, p.2158-2170, out. 2015. Institute of Electrical and Electronics Engineers (IEEE). http://dx.doi.org/10.1109/tifs.2015.2433898.

BALACHANDRAN, Anand *et al*. **Characterizing user behavior and network performance in a public wireless LAN**. Proceedings Of The 2002 Acm Sigmetrics International Conference On Measurement And Modeling Of Computer Systems - Sigmetrics '02, [s.l.], p.195-205, 2002. ACM Press. http://dx.doi.org/10.1145/511334.511359.

CARMO, Luiz Fernando Rust da Costa; COSTA, Danielle. **Reconhecimento de padrões de comportamento individual baseado no histórico de navegação em um Web Site**. Núcleo de Computação Eletrônica – Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, p.1-12. 2007.

CONGDON, P. *et al*. **IEEE 802.1X Remote Authentication Dial In User Service (RADIUS)**. Rfc 3580, ., p.1-30, set. 2003.

DAVID D. COLEMAN (Indiana). **802.1X/EAP Authentication**. In: COLEMAN, David D.; WESTCOTT, David A.; HARKINS, Bryan. CWSP® Certified Wireless Security Professional Study Guide CWSP-205. 2. ed. Indianapolis: John Wiley & Sons, 2017. Cap. 4. p. 87-150.

DESMOND, Loh Chin Choong *et al*. **Identifying unique devices through wireless fingerprinting**. Proceedings Of The First Acm Conference On Wireless Network Security - Wisec '08, [s.l.], p.46-54, 2008. ACM Press. http://dx.doi.org/10.1145/1352533.1352542.

DRAPER, Norman R.; SMITH, Harry. **Applied Regression Analysis**. 3. ed. New York: Wiley-interscience, 2014.

EMIGH, Aaron. **The Crimeware Landscape: Malware, Phishing, Identity Theft and Beyond**. Journal Of Digital Forensic Practice, [s.l.], v. 1, n. 3, p.245-260, set. 2006. Informa UK Limited. http://dx.doi.org/10.1080/15567280601049985.

EVERITT, B.s.; SKRONDAL, A.. **The Cambridge Dictionary of Statistics**. 4. ed. Cambridge: Cambridge, 2010.

FREITAS NETO, José Alves de; TASINAFO, Célio Ricardo. História Geral e do Brasil. 2. ed. Belo Horizonte: Harbra, 2011.

HWANG, Hyunuk *et al*. **A Study on MITM (Man in the Middle) Vulnerability in Wireless Network Using 802.1X and EAP**. 2008 International Conference On Information Science And Security (iciss 2008), [s.l.], p.164-170, jan. 2008. IEEE. http://dx.doi.org/10.1109/iciss.2008.10.

KEDMA, Gabi *et al*. **Analyzing users' web surfing patterns to trace terrorists and criminals. 2013 Ieee International Conference On Intelligence And Security Informatics**, [s.l.], p.143-145, jun. 2013. IEEE. http://dx.doi.org/10.1109/isi.2013.6578804.

KIM, Jong-moon *et al*. **A Study on Wireless Intrusion Prevention System based on Snort**. International Journal Of Software Engineering And Its Applications, Canada, v. 9, n. 2, p.1-12, 2015.

Kumar, Vinod. (2012). **Signature Based Intrusion Detection System Using SNORT**. International Journal of Computer Applications & Information Technology. 1. 7.

Lackner, Günther & Payer, Udo & Teufl, Peter. (2009). **Combating Wireless LAN MAC-layer Address Spoofing with Fingerprinting Methods**. I. J. Network Security. 9. 164-172

LOPEZ, Marton Andreoni *et al*. BroFlow: **Um Sistema Eficiente de Detecção e Prevenção de Intrusão em Redes Definidas por Software**. Grupo de Teleinformatica e Automação da Ufrj, Rio de Janeiro, p.1-14, out. 2014.

Morettin, L.G., 2010. **Estatística Básica: Probabilidade e Inferência**. São Paulo:

Pearson Prentice. 375 p.

MORETTIN, Pedro A.; BUSSAB, Wilton de O.. **Estatística Básica**. 5. ed. São Paulo: Saraiva, 2004.

PAN, Junshan; HU, Hanping; LIU, Ying. **Human behavior during Flash Crowd in web surfing**. Physica A: Statistical Mechanics and its Applications, [s.l.], v. 413, p.212-219, nov. 2014. Elsevier BV. http://dx.doi.org/10.1016/j.physa.2014.06.085.

PANTIC, Maja et al. **Human Computing and Machine Understanding of Human Behavior: A Survey**. Artifical Intelligence For Human Computing, [s.l.], p.47-71, 2007. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-540-72348-6_3.

SEGURANÇA DE REDES DE COMPUTADORES NA INTERNET. Teresina: Santo Agostinho, v. 1, n. 2, jul. 2012. Semestral. Issn Eletrônico: 2357-9501.

Silva, Anderson & Guelfi, Adilson. (2010). **Sistema para Identificação de Alertas Falso Positivos por meio de Análise de Correlacionamentos e Alertas Isolados**. I2TS - 9th Internartional Information and Telecommunication Technologies Symposium 2009.

TRABELSI, Dr. Zouheir; ALKETBI, Latifa. **Using Network Packet Generators and Snort Rules for Teaching Denial of Services Attacks**. ., United Arab Emirates, p.285-290, ago. 2013.

YU, Shui *et al*. **Discriminating DDoS Attacks from Flash Crowds Using Flow Correlation Coefficient**. Ieee Transactions On Parallel And Distributed Systems, Washington, v. 23, n. 06, p.1073-1079, jun. 2012.