UNIVERSIDADE FEDERAL DO CEARÁ

CAMPUS DE QUIXADÁ

PROGRAMA DE PÓS-GRADUAÇÃO EM COMPUTAÇÃO

CARLOS BRUNO PEREIRA BEZERRA SANTOS

SBIM - UM MODELO DE INVESTIGAÇÃO FORENSE PARA PRÉDIOS INTELIGENTES

QUIXADÁ

2023

CARLOS BRUNO PEREIRA BEZERRA SANTOS

SBIM - UM MODELO DE INVESTIGAÇÃO FORENSE PARA PRÉDIOS INTELIGENTES

Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Computação. Área de Concentração: Ciência da Computação

Orientador: Prof. Dr. Marcio Espíndola Freire Maia

QUIXADÁ

2023

Dados Internacionais de Catalogação na Publicação
Universidade Federal do Ceará
Sistema de Bibliotecas
Gerada automaticamente pelo módulo Catalog, mediante os dados fornecidos pelo(a) autor(a)

S234s Santos, Carlos Bruno Pereira Bezerra.
SBIM - um modelo de investigação forense para prédios Inteligentes / Carlos Bruno Pereira Bezerra Santos. – 2023.
124 f. : il. color.

Dissertação (mestrado) – Universidade Federal do Ceará, Campus de Quixadá, Programa de Pós-Graduação em Computação, Quixadá, 2023.
Orientação: Prof. Dr. Márcio Espíndola Freire Maia.

1. Internet das coisas. 2. Ciência Forense. 3. Modelo Investigativo. 4. Segurança. I. Título.
CDD 005

CARLOS BRUNO PEREIRA BEZERRA SANTOS

SBIM - UM MODELO DE INVESTIGAÇÃO FORENSE PARA PRÉDIOS INTELIGENTES

Dissertação apresentada ao Programa de Pós-Graduação em Computação da Universidade Federal do Ceará, como requisito parcial à obtenção do título de mestre em Computação. Área de Concentração: Ciência da Computação

Aprovada em: _____/_____/_____.

BANCA EXAMINADORA

Prof. Dr. Marcio Espíndola Freire Maia (Orientador)
Universidade Federal do Ceará (UFC)

Prof. Dr. Arthur de Castro Callado
Universidade Federal do Ceará (UFC)

Prof. Dr. Emanuel Bezerra Rodrigues
Universidade Federal do Ceará (UFC)

À minha mãe e tia, por compartilharem minhas esperanças. Mãe, obrigado por todo o seu incentivo, nos momentos de silêncio e nos mais difíceis e improváveis. Tia Mazé, sem você eu nunca teria enxergado pelo que vale a pena tentar mais uma vez.

AGRADECIMENTOS

Ao Prof. Dr. Marcio Espíndola por toda a orientação.

À banca julgadora deste trabalho, por todos os ensinamentos desde o processo de qualificação. Em especial ao professor Arthur, meu melhor professor da graduação, o qual contribuiu para o meu desenvolvimento pessoal e profissional.

À minha namorada, Lívia Maria, por escutar meus problemas de pesquisa e me incentivar a ser uma pessoa melhor.

Ao Iron Maiden, Blind Guardian, Eluveitie, Behemoth e mais algumas bandas que me serviram de companhia no processo de escrita desta dissertação.

Aos professores Jeandro Bezerra e João Marcelo, da UFC Quixadá, por me apresentarem a iniciação à pesquisa, ainda na época da graduação.

A todos os professores do Programa de Pós Graduação em Computação de Quixadá, pelo esforço e comprometimento em manter o Ensino e a Pesquisa mesmo nos momentos mais difíceis que atravessamos durante dois anos pandemia.

Ao amigo de laboratório Leôncio Resende pela companhia e discussões sobre as disciplinas de Mestrado.

Aos meus amigos, que também contribuem para a minha evolução pessoal e profissional.

Agradeço a todos os professores que participaram da minha formação enquanto ser humano e aluno. Este trabalho é uma amostra do quanto vocês podem transformar vidas.

À Fundação Cearense de Apoio ao Desenvolvimento (FUNCAP) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo financiamento desta pesquisa de Mestrado.

E ao Doutorando em Engenharia Elétrica, Ednardo Moreira Rodrigues, e seu assistente, Alan Batista de Oliveira, aluno de graduação em Engenharia Elétrica, pela adequação do *template* utilizado neste trabalho para que o mesmo ficasse de acordo com as normas da biblioteca da Universidade Federal do Ceará (UFC).

"Questions are a burden and answers are a prison for oneself."

(Smith/Dickinson)

RESUMO

Os recentes avanços em *hardware* e Tecnologia da Informação propiciaram a interconexão de bilhões de dispositivos ao redor do globo. Com cada vez mais objetos inseridos na rede mundial, vive-se, hoje, a era da Internet das Coisas. A transferência de dados pela rede sem a interferência humana traz confiabilidade e conveniência a usurários de serviços em saúde, transporte, controle ambiental e automação doméstica. Entretanto, esse cenário abre um novo mundo de oportunidades a invasores. Com isso, é pertinente a aplicação da Ciência Forense para levantar evidências com validade judicial e atribuir responsabilidades. Atualmente, não existe qualquer padrão amplamente aceito para a coleta, análise e apresentação de dados forenses para a Internet das Coisas ou para qualquer uma de suas aplicações, como os Prédios Inteligentes. Neste trabalho é apresentado o *Smart Building Investigation Model* (SBIM), um modelo de investigação forense eficiente para Prédios Inteligentes. O modelo proposto tem a vantagem de estar adequado ao processo investigativo preconizado pelo *National Institute of Standards and Technology* (NIST). Portanto, acredita-se que o modelo proposto, se incorporado com sucesso, facilitará a investigação de crimes e incidentes de Segurança da Informação em Prédios Inteligentes, sobretudo por não haver, até o momento, modelos específicos para eles.

Palavras-chave: internet das coisas; ciência forense; modelo investigativo; segurança.

ABSTRACT

Recent advances in hardware and Information Technology have enabled the interconnection of billions of devices around the globe. With more and more objects inserted in the world wide web, today we live in the age of the Internet of Things. Data transfer over the network without human interference brings reliability and convenience to users of services in healthcare, transportation, environmental control and home automation. However, this scenario opens up a whole new world of opportunity for attackers. With this, it is pertinent to apply Forensic Science to raise evidence with judicial validity and assign responsibilities. Currently, there is no widely accepted standard for the collection, analysis and presentation of forensic data for the Internet of Things or any of its applications such as Smart Buildings. This work presents the *Smart Building Investigation Model* (SBIM), an effective and efficient forensic investigation model for Smart Buildings. The proposed model has the advantage of being adequate to the investigative process recommended by the NIST. Therefore, it is believed that the proposed model, if successfully incorporated, will facilitate the investigation of crimes and Information Security incidents in Smart Buildings, especially as there are no specific models for them so far.

Keywords: internet of things; forensic science; investigative model; security.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE QUADROS

LISTA DE ABREVIATURAS E SIGLAS

SBIM	*Smart Building Investigation Model*
NIST	*National Institute of Standards and Technology*
IoT	*Internet of Things*
BPMN	*Business Process Model and Notation*
DFR	*Digital Forensics Readiness*
ISO	*International Organization for Standardization*
TI	*Tecnologia da Informação*
SDI	*Sistema de Detecção de Intrusão*
ITIL	*IT Infrastructure Library*
LoRaWAN	*Long Range Wide Area Network*
IP	*Internet Protocol*
DHCP	*Dynamic Host Configuration Protocol*
BLE	*Bluetooth Low Energy*

SUMÁRIO

1 INTRODUÇÃO

1.1 Contextualização

Desde que foi projetada, a rede mundial de computadores, ou Internet, tem permitido o surgimento de muitas oportunidades de negócio. Ao longo dos anos, os seus usuários têm utilizado serviços cada vez mais sofisticados, como por exemplo aplicações de monitoramento de saúde, atividades físicas e rastreamento de objetos em tempo real. O crescimento da Internet como um meio de negócio é constante, assim como a quantidade de dispositivos conectados a ela (MANYIKA *et al.*, 2015).

Atualmente, a Internet está evoluindo para interconectar não somente os dispositivos mais convencionais, como os computadores pessoais e *smartphones*, como também coisas diversas do nosso dia-a-dia, como pulseiras, relógios, televisores, carros e micro-ondas. E essas "coisas" estarão muito mais presentes na Internet em poucos anos (ATZORI *et al.*, 2010). Mais que isso, a nossa tendência de estar cada vez mais conectado movimentará um mercado na casa dos bilhões de dólares (RANA *et al.*, 2017).

Com tantas "coisas" interconectando-se para trocar dados e informações de contexto, surge um novo paradigma de comunicação chamado de Internet das Coisas (LIN *et al.*, 2017). Apesar de não existir um padrão amplamente aceito que defina a IoT, ela pode ser entendida como um um meio empregado para conectar objetos físicos e virtuais que são unicamente identificados e dinamicamente configurados (MINERVA *et al.*, 2015). Assim, embora a IoT tenha o potencial de expandir a Internet, o rápido crescimento de novas aplicações e serviços viabilizam o aparecimento de novas vulnerabilidades de segurança da informação que precisam ser cuidadosamente tratadas (HARBI *et al.*, 2021).

Nessa conjuntura, é revelante que as organizações disponham de meios para investigar os acontecimentos que comprometam as políticas de segurança da organização. Nas investigações, é necessário que as evidências sejam coletadas de maneira que permitam a responsabilização devida dos autores (NIST, 2006). Para isso, existem processos que conduzem a investigação mediante fases importantes para a coleta e apresentação de evidências, sobretudo mantendo a integridade das informações.

Tendo em vista que até o momento ainda não existem padrões amplamente aceitos para a busca de evidências digitais em quaisquer das aplicações de Internet das Coisas, sobretudo em Prédios Inteligentes, é necessário um amplo esforço para se chegar a um consenso sobre os

requisitos em IoT que devem ser atendidos ao se propor qualquer processo de investigação em um ambiente tão volátil e heterogêneo.

1.2 Motivação

O principal problema que motivou esta dissertação de mestrado foi a ausência de um modelo eficiente de investigação de evidências em Prédios Inteligentes. O campo emergente de investigação digital em IoT ainda não está consolidado e há falta de consenso nos procedimentos utilizados para a manutenção de evidências digitais (STOYANOVA *et al.*, 2020). Há muitos desafios para as partes interessadas no processo investigativo em Prédios Inteligentes, pois, pela falta de conceitos e procedimentos apropriados, o corpo investigativo torna-se vulnerável ao cometimento de erros que comprometam a validade judicial das provas. Devido a essa falta de padronização, é necessário que novos modelos de investigação sejam propostos na tentativa de se chegar a um consenso (STOYANOVA *et al.*, 2020). Nessa perspectiva, um modelo de investigação para Prédios Inteligentes torna-se imprescindível.

1.3 Objetivos

Esta dissertação de mestrado tem como principal objetivo o estabelecimento do *Smart Building Investigation Model* (SBIM), um modelo eficiente de investigação forense para Prédios Inteligentes. Além disso, este trabalho conta com os seguintes objetivos específicos:

- Fornecer duas fases adicionais ao modelo criado pelo *National Institute of Standards and Technology* (NIST);
- Fornecer uma implementação capaz de padronizar o formato de arquivos de logs das tecnologias Internet Protocol (IP), Long Range Wide Area Network (LoRaWAN) e Bluetooth Low Energy (BLE);
- Diminuir o tempo de consulta aos *logs* por parte da equipe investigativa.

Como resultado, o SBIM possibilita a execução de processos investigativos de maneira eficiente, além de preservar a integridade dos dados coletados. Consequentemente, este trabalho pretende contribuir para a consolidação do campo de investigação forense em Internet das Coisas.

1.4 Organização

Esta dissertação está organizada da seguinte forma: o Capítulo 2 apresenta o conhecimento necessário em Internet das Coisas e o processo forense. Ele inclui a definição e conceitos básicos de IoT e ciência forense, incluindo arquitetura, camadas e caracterização de requisitos importantes como a manutenção da cadeia de custódia. Além disso, ele foca na apresentação dos problemas em segurança no contexto de IoT, na investigação de incidentes por meio da coleta de evidências e no uso do modelo do NIST como processo investigativo, inclusive em Prédios Inteligentes.

O Capítulo 3 apresenta os trabalhos relacionados no campo de investigação forense em Internet das Coisas, inclusive para Prédios Inteligentes. Depois, faz uma análise qualitativa dos trabalhos levantados utilizando características importantes para uma investigação digital eficaz.

O Capítulo 4 apresenta o SBIM: um modelo de investigação forense para Prédios Inteligentes. O processo de investigação SBIM é modelado usando-se a notação *Business Process Model and Notation* (BPMN) [1] e formas de implementar cada uma de suas fases são mostradas.

O Capítulo 5 apresenta quais características do SBIM serão validadas, através de questionários enviados a especialistas, para, de fato, mostrar-se ser um processo eficaz e mais eficiente do que o do NIST.

O Capítulo 6 apresenta a conclusão desta dissertação de Mestrado, trazendo os principais contribuições da pesquisa, as dificuldades e limitações e os trabalhos futuros.

[1] https://www.bpmn.org/

2 FUNDAMENTAÇÃO TEÓRICA

A Internet das Coisas é uma expressão utilizado para designar a troca de informações entre coisas e objetos heterogêneos. Ainda assim, existem diversas definições na literatura sobre do que se trata o termo exatamente, pois não existe um padrão amplamente aceito para conceituá-lo. Como tecnologia, a IoT promete ampliar o número de dispositivos conectados à rede mundial de computadores e oferecer aplicações que viabilizem novos negócios e melhorem a qualidade de vida dos usuários.

Nesse contexto, aspectos de segurança da informação devem ser levantados em um ambiente tão diversificado de dispositivos e tecnologias, os quais aparecem como novas oportunidades para criminosos na internet. Dessa forma, ao estarem potencialmente sujeitas a ataques, as organizações precisam entender, planejar e aplicar ferramentas, métodos e processos que investiguem os incidentes, coletem evidências e identifiquem os envolvidos. Isso não só ajuda a entender suas vulnerabilidades, mas também a corrigi-las.

Neste capítulo, são apresentados os conceitos fundamentais de IoT, tais como seu modelo arquitetural mais presente na literatura e a descrição de suas camadas. Além disso, são mostrados problemas de segurança da informação envolvendo vários aspectos do paradigma de Internet das Coisas. Posteriormente, é trazido o conceito de Forense Digital, como ele é importante na investigação de incidentes e seus desafios inerentes à IoT. Logo depois, são sugeridas as fases de investigação forense do NIST para coleta, análise e apresentação de evidências digitais que podem ser usadas em processos judiciais. Por último, é ressaltada a importância da forense digital em uma aplicação real da Internet das Coisas, chamada Prédio Inteligente, que faz parte do escopo deste trabalho.

2.1 Internet das Coisas

A internet das coisas é um paradigma bem conhecido o qual é constituído por um ambiente com vários dispositivos dinamicamente configurados. A IoT segue basicamente o princípio da comunicação machine-to-machine (M2M), efetua computação baseada no contexto e utiliza a identificação por rádio frequência (RFID), as redes de sensores sem fio, os protocolos da internet e a comunicação móvel para a troca de informações entre objetos na rede (SILVA *et al.*, 2013). Várias "coisas" podem estar conectadas em um ambiente IoT: relógios inteligentes, trancas de porta inteligentes, sensores de ambiente, temperatura, gás ou luz, veículos inteligentes,

drones e aplicações para automação industrial. Como se pode notar, os dispositivos IoT variam desde pequenos até grandes objetos. Eles são equipados com sensores e atuadores que, de maneira inteligente, percebem o arredor e executam ações de forma autônoma (HAMMOUDI *et al.*, 2018). Esses dispositivos têm recursos limitados, tais como pouca memória, baixa capacidade de processamento e pouco poder computacional. A IoT se refere à próxima geração da internet (LI *et al.*, 2015) e também é chamada de Internet of Everything (LEE; LEE, 2015).

Dada essa grande admissibilidade de dispositivos, as previsões para o crescimento da IoT nos próximos anos revelam que o número de dispositivos que estarão conectados na rede representa um impacto positivo considerável no mercado de vendas. Para se ter uma ideia, o mercado de casas inteligentes é esperado que alcance mais de 140 bilhões de dólares até 2023, o que significa um aumento de 17% em comparação ao ano de 2019 (RANA *et al.*, 2017). O conselho de inteligência nacional dos Estados Unidos estima que até 2025 as coisas mais presentes na internet sejam coisas que estão no nosso cotidiano, incluindo pacotes de comida, mobiliária, documentos de papel e muito mais (ATZORI *et al.*, 2010). Além disso, espera-se que o número de dispositivos IoT alcance 1 trilhão até 2025, o que significa que esse mercado tem o potencial econômico de receita, por ano, de 11 trilhões de dólares até 2025 (MANYIKA *et al.*, 2015).

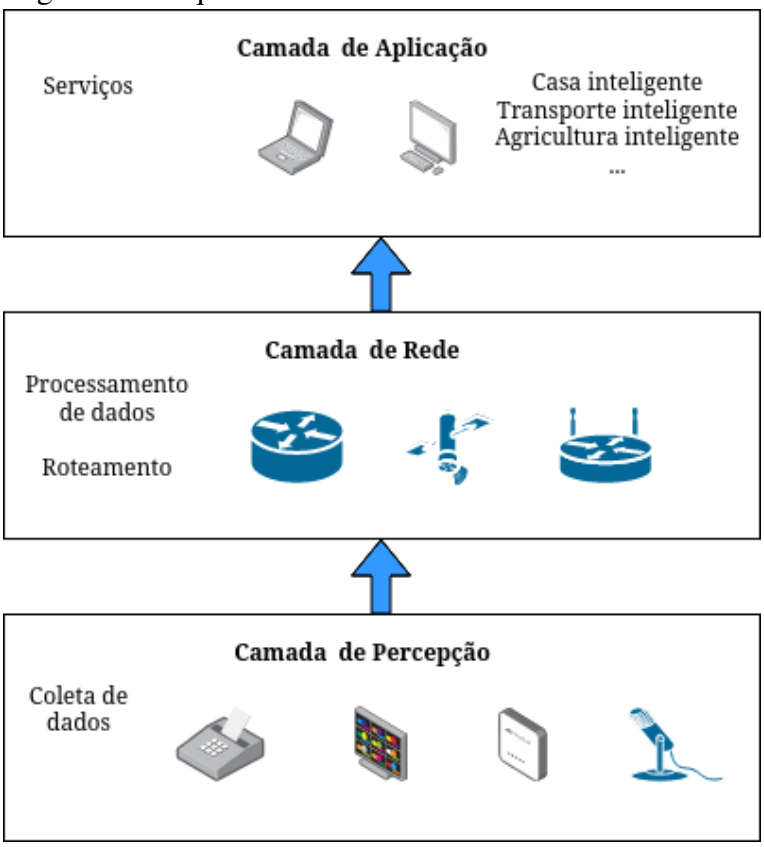
2.1.1 Arquitetura da IoT

Não existe um padrão amplamente aceito de arquitetura para a Internet das Coisas na literatura. Entretanto, tipicamente, a arquitetura IoT é proposta em três camadas: Aplicação, Rede e Percepção (LIN *et al.*, 2017), como mostra a Figura 1.

2.1.1.1 Camada de Percepção

Na camada de percepção, estão presentes diferentes tipos de dispositivos IoT. É nessa camada que é feita a coleta de dados e a interação entre os dispositivos. A coleta de dados geralmente é feita por sensores e pelo uso de Identificação de Frequência de Radio (RFID). A tecnologia RFID permite a identificação, o rastreamento e a monitoração de objetos (JIA *et al.*, 2012). No sistema RFID, existem os marcadores (tags) e os leitores. Os objetos são identificados pelas tags. Já os leitores RFID têm a função de identificar os dispositivos através da leitura das tags por ondas de rádio. Sensores sem fio também têm grande importância no ambiente de Internet das Coisas. As Redes de Sensores sem Fio (WSN) são uma tecnologia que consiste em

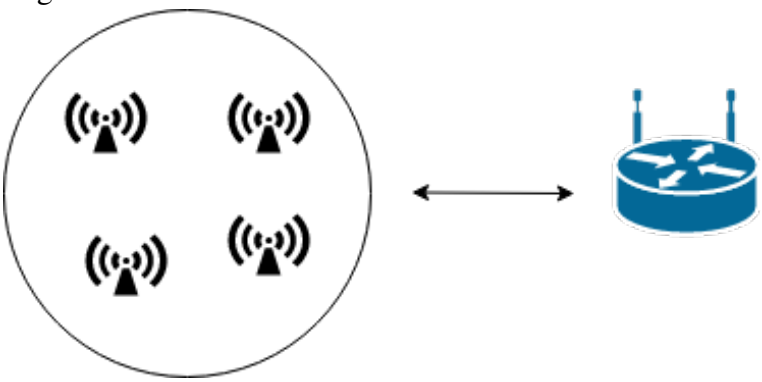
Figura 1 – Arquitetura da IoT



Fonte: Elaborada pelo autor.

vários nós interconectados que trocam informações entre si e entre um gateway (KOCAKULAK; BUTUN, 2017), como mostrado na Figura 2. Tais informações dizem respeito ao que os sensores captam nos arredores: temperatura, umidade e vibração, por exemplo.

Figura 2 – Redes de Sensores sem Fio



Fonte: Elaborada pelo autor.

2.1.1.2 Camada de Rede

A camada de Rede processa os dados coletados na camada de percepção. Esses dados podem ser armazenados ou enviados para a camada de aplicação. Com certeza, esta camada tem extrema importância na arquitetura da IoT, já que ela integra várias tecnologias diferentes de comunicação. Dentre os protocolos mais amplamente utilizados na camada de rede, estão: ZigBee, *Bluetooth low energy* (BLE), IPv6 *over low power wireless personal area networks* (6LoWPAN) e o *long-range wide area network* (LoRaWAN). Cada um desses protocolos aplicam-se a diferentes requisitos de interoperabilidade, segurança, distância da rede e escalabilidade (HARBI *et al.*, 2021).

2.1.1.3 Camada de Aplicação

A camada de Aplicação recebe os dados processados da camada de rede e provê os serviços solicitados pelos usurários. Ela suporta uma infinidade de serviços como Casas Inteligentes, Redes elétricas inteligente, Transportes Inteligentes e vários outros. Os protocolos mais comumente utilizados nesta camada são o *Constrained Application Protocol* (CoAP) e o *Message Queuing Telemetry Transport* (MQTT) (HARBI *et al.*, 2021). Tais protocolo oferecem serviços sem grande consumo de recursos, haja vista a limitação de recursos disponível em dispositivos IoT.

2.1.2 Segurança na IoT

A Internet das Coisas, como um novo paradigma, trouxe uma gama de novas possibilidades para o desenvolvimento de novos produtos e serviços. Hoje em dia, relógios e pulseiras inteligentes já fazem parte da vida das pessoas. Esses dispositivos trazem uma série de benefícios aos seus usuários, como monitoramento de atividades físicas, monitor cardíaco e de sono. É notório que informações a respeito de nossa rotina e saúde são muito importantes para a manutenção de uma vida saudável. Por estarem tão disseminados, os dispositivos incorporaram-se a nossa vida de tal forma que sequer percebemos quando e onde os dados estão sendo processados.

A chegada de cada vez mais dispositivos inteligentes e pervasivos no mercado permite abrir uma discussão relevante sobre a segurança dos usuários. Nesse contexto, algumas dimensões precisam ser consideradas, como a necessidade de evitar o vazamento de informações sensíveis (confidencialidade), a garantia de confirmar a identidade de algo ou alguém (autenticidade), o

tratamento correto das informações, de forma que haja mecanismos que detectem adulteração nos dados (integridade), e a disponibilidade dos dados e serviços (disponibilidade).

Como vimos, a IoT possui uma grande heterogeneidade de dispositivos, que vão desde lâmpadas inteligentes a uma completa rede de energia inteligente. Além disso, muitos dos dispositivos possuem baixa capacidade de processamento, armazenamento e memória, por serem muito pequenos. Todas essas características favorecem o aparecimento de novas tecnologias e protocolos que busquem lidar com essas limitações. O outro lado que devemos enxergar é que com um número muito maior de dispositivos conectados, os atacantes ou usuários maliciosos agora têm uma possibilidade muito maior de alvos. Basta pensarmos na ideia de que marcapassos, injeções de insulina e monitor de pressão arterial estarão interconectados em uma rede inteligente hospitalar. Os riscos são pertinentes e toda brecha de segurança deve ser tratada. A partir da próxima subseção, são elencadas as ameaças mais comumente presentes em cada camada da arquitetura, segundo Harbi *et al.* (2021)

2.1.2.1 Ameaças à camada de percepção

Na camada de percepção temos a presença das redes de sensores sem fio e do RFID. O grande problema das redes de sensores é que elas podem ser dispostas nos mais diversos tipos de ambientes, como ruas, florestas e regiões inóspitas. Portanto, eles são suscetíveis a diversos tipos de ataques (HARBI *et al.*, 2019):

- *Sinkhole*: um ataque *Sinkhole* é um ataque no qual o atacante modifica as métricas do protocolo de roteamento em uma rede para que o tráfego seja todo direcionado a um dispositivo específico. Com todos os dados centralizados em um único ponto, o atacante tenta extrair informações (KIBIRIGE; SANGA, 2015).
- *Sybil*: um ataque contra a identidade em que uma entidade individual se disfarça como múltiplas identidades simultâneas. O ataque *Sybil* é um problema fundamental em muitos sistemas e até agora resistiu a uma solução universalmente aplicável (LEVINE *et al.*, 2006).
- *Denial of Service* (DoS): seu objetivo é esgotar recursos e fazer com que o servidor ou a rede não forneçam serviço a usuários legítimos (CHAO-YANG, 2011).

Similar às redes de sensores, as redes RFIDs também estão suscetíveis a ataques à segurança da informação (HARBI *et al.*, 2019).

2.1.2.2 Ameaças à camada de rede

A respeito do ZigBee, Cao *et al.* (2016) apresentou um ataque que objetiva drenar a energia de nós ZigBee. Os autores em Coppolino *et al.* (2015) avaliaram a vulnerabilidade da rede ZigBee perante um ataque de *Sinkhole*. Já em Morgner *et al.* (2017) mostrou-se que sistemas de luz inteligentes baseados em ZigBee estão suscetíveis a ataques de negação de serviço e injeção de código.

6LoWPAN é um protocolo que permite o uso do protocolo IPv6 em uma rede de dispositivos com recursos limitados, incluindo pouca energia. Ele comprime o cabeçalho original do IPv6 de forma que o processamento e envio do pacote seja mais rápido e consuma menos memória, rede e processamento. Todavia, ele não oferece confidencialidade, autenticação ou preservação de integridade. Como consequência, um indivíduo malicioso pode injetar outros fragmentos de pacote na rede fazendo com que o nó receptor esgote seus recursos (HUMMEN *et al.*, 2013). Repetindo-se o processo de injeção por várias vezes, constitui-se um ataque de negação de serviço (RGHIOUT *et al.*, 2014).

O protocolo LoRaWAN utiliza um algoritmo de 128 bits para garantir a confidencialidade e integridade. Quando um dispositivo deseja entrar na rede, o servidor o envia duas chaves: chave de sessão de rede e chave de sessão de aplicação. Essas chaves são usadas para criptografia e descriptografia dos dados. A principal brecha de segurança em um cenário como esse é o gerenciamento das chaves de sessão. Um intruso pode acessar as chaves por meio de ataque do tipo *side channel*, pois elas são armazenadas nos próprios dispositivos (HARBI *et al.*, 2021).

2.1.2.3 Ameaças à camada de aplicação

Na camada de aplicação, vemos a presença de serviços, usuários e dos protocolos CoAP e MQTT. Os usuários serão sempre um alvo em potencial para os atacantes, já que podem ser manipulados a entregar informações que permitam o acesso a dados sensíveis. As manipulações podem ocorrer através de técnicas de *phishing*, por exemplo. O CoAP é um protocolo desenvolvido para prover os serviços do HTTP no ambiente IoT e, por isso, adota a arquitetura RESTful para a troca de dados. Para prover confidencialidade, autenticidade e integridade, o CoAP utiliza os serviços do datagram TLS (DTLS). Entretanto, existem limitações no DTLS que podem ser consideradas vulnerabilidades, também, no CoAP (RAHMAN; SHAH,

2016).

O MQTT pode utilizar o secure socket layer (SSL) para criptografar e descriptografar seus dados. Entretanto, apesar de o SSL ser amplamente usado, ele ainda pode mostrar vulnerabilidades em um ataque do tipo man-in-the-middle (CYNTHIA *et al.*, 2019). Foi proposta uma versão segura do MQTT chamada SMQTT. Todavia, não existe padrão para os algoritmos de geração das chaves criptográficas (SINGH *et al.*, 2015).

2.2 Forense Digital

O potencial de crescimento da IoT deve facilitar a rotina das pessoas por meio dos serviços oferecidos através de dispositivos interconectados por rede. O novo paradigma de conectar tudo à rede mundial de computadores apresenta diversos desafios. Ao concebermos que praticamente qualquer coisa estará apta a trocar dados via rede, não é difícil concluirmos que a heterogeneidade dos dispositivos é um dos problemas a serem mitigados. Com a presença de vários fabricantes no mercado, necessita-se de uma padronização de comunicação para que haja interoperabilidade em um ambiente tão diversificado. Outro ponto, mas não menos importante, diz respeito à segurança das informações e usuários.

Como se pôde notar, não existe um padrão aceito de forma ampla para o modelo arquitetural de internet das coisas. Geralmente, o que se encontra na literatura é um modelo de três camadas. Sendo assim, ao analisar os protocolos e serviços presentes em cada uma das camadas, pode-se concluir que os principais protocolos de comunicação ainda apresentam falhas de segurança, pelo menos no que diz respeito à confidencialidade, autenticidade e integridade. Para ratificar a importância de se adotar medidas rigorosas de segurança, apenas em 2017 houve um crescimento de 600% nos ataques relacionados a dispositivos IoT (SYMANTEC, 2018).

Como resultado, o cybercrime tornou-se o segundo crime mais reportado do mundo (PWC, 2018). Além disso, os fabricantes geralmente preocupam-se mais com o custo e a usabilidade de seus produtos, enquanto aspectos de segurança são negligenciados. O que mais impressiona é que as empresas utilizam práticas de segurança principalmente com receio de que sua imagem seja afetada negativamente, ou seja, a proteção dos usuários não vem em primeiro lugar (LALLY; SGANDURRA, 2018).

Tão importante quanto a prevenção, a investigação das causas de um incidente de segurança da informação deve ser regularmente planejada e executada. Incidente é qualquer evento com alta probabilidade de comprometer os ativos da organização (ISO, 2018). Nessa

conjuntura, uma poderosa ferramenta científica pode auxiliar na busca de evidências legalmente aceitas: a ciência forense.

De acordo com NIST (2006), a ciência forense é geralmente definida como a aplicação da ciência à lei. Forense Digital (FD), também conhecida como forense de rede ou de computador, tem muitas definições. Geralmente, ela é considerada a aplicação da ciência para a identificação, coleta, exame e análise dos dados, enquanto se preserva a integridade da informação e manutenção da cadeia de custódia para os dados. De acordo com o Request for comment (RFC) 3227, a cadeia de custódia diz repeito ao processo de documentar tudo o que acontece com a evidência: quando, onde e por quem a evidência foi descoberta e coletada. Onde, quando e quem examinou a evidência. Como a evidência foi armazenada e quem ficou responsável por ela por quanto tempo. Quando a evidência trocou de responsável, durante qual período e como ocorreu a transferência. A cadeia de custódia começa quando uma evidência é coletada e termina quando a evidência é apresentada na corte. Além disso, ela prova que o material sendo investigado não foi alterado enquanto estava passando por um processo de investigação (LONE; MIR, 2018; COSIC; COSIC, 2012; ZULKIPLI *et al.*, 2017).

A forense digital na internet das coisas (forense IoT) pode ser vista como um ramo da forense digital tradicional (ZAWOAD; HASAN, 2015). Ela também busca extrair informações que eventualmente possam ser utilizadas em um processo legal. Apesar de objetivos similares, a forense IoT é um ramo novo e pouco explorado na academia (STOYANOVA *et al.*, 2020). Enquanto a forense digital tradicional coleta e examina dados de dispositivos como notebooks, smartphones e servidores, a forense IoT engloba uma variedade muito maior de dispositivos, tais como: sistemas veiculares, semáforos, trancas de porta, drones e até implantes em seres humanos.

A forense IoT pode ser relevantes diante de vários cenários. Como já foi mencionado, com o paradigma IoT, espera-se que um número muito maior de dispositivos estejam interconectados, o que aumenta as possibilidades para usuários mal-intencionados (WANG *et al.*, 2016). Por exemplo, dispositivos IoT podem ser usados para executar um ataque de negação de serviço e até mesmo causar o caos na bolsa de valores (RONDEAU *et al.*, 2019).

Além disso, Alabdulsalam *et al.* (2018) afirmam que a IoT pode ameaçar a vida humana. Os autores falam de um caso que ocorreu nos Estados Unidos no ano de 2017, onde aparelhos que gerenciavam os batimentos cardíacos de pessoas com arritmia tinham vulnerabilidades de segurança.

Assim, a forense IoT deve auxiliar os profissionais de tecnologia da informação a rastrear fatos e provar ou refutar suas hipóteses em cenas de crime ou incidentes. Por exemplo, é possível coletar informações de um detector de incêndio e determinar o exato momento e lugar onde o fogo começou (SERVIDA; CASEY, 2019).

2.2.1 Desafios da Forense Digital em IoT

O NIST (2014) elencou 65 desafios presentes no processo investigativo forense na nuvem e nos dispositivos IoT. Dentre os vários problemas, estão o processo complexo para a coleta de evidências e a natureza *multitenant* presente na nuvem.

Acerca dos demais desafios que podem ser encontrados num ambiente IoT, podemos dividir tais desafios em seis categorias (STOYANOVA *et al.*, 2020), quais sejam: *Forensic Readiness*, Identificação, Coleta, Preservação, Análise e Correlação, Responsabilização de Ataques e Apresentação das Evidências.

2.2.1.1 Forensic Readiness

O termo *Digital Forensics Readiness* (DFR), não traduzido neste trabalho em razão da falta de uma tradução apropriada na literatura, representa a capacidade de "coletar, preservar, proteger e analisar provas digitais para que elas possam ser efetivamente utilizadas, em processos legais, em um Tribunal de Justiça" (ASSURANCE, 2015). Na verdade, isso implica que as abordagens forenses digitais não devem ser usadas apenas em atividades pós-incidente, mas também para aumentar as chances de obter bons resultados e gastar menos recursos em investigações futuras (KEBANDE; VENTER, 2018; ALENEZI *et al.*, 2017). Portanto, o DFR pode ser entendido como uma fase proativa em determinada investigação.

Conforme presente na ISO/IEC 27043 (ISO/IEC, 2015d), os objetivos das práticas DFR incluem:

- Preservar e melhorar o nível de segurança da informação nas organizações;
- Prevenir ou minimizar a interrupção das atividades e processos de negócios da organização;
- Minimizar o custo de conduzir uma investigação forense digital;
- Maximização do valor potencial da evidência forense digital (KEBANDE *et al.*, 2018).

Muitas organizações modernas já reconheceram a necessidade de um processo de DFR. Apesar de ter sido reconhecido como um objetivo altamente recomendado (CHERNYSHEV *et al.*, 2018; KEBANDE *et al.*, 2018), a integração da DFR em sistemas IoT permanece um

desafio.

Dessa forma, a produção de equipamentos de IoT e o fornecimento de serviços prontamente adaptáveis e integrados aos processos digitais atuais ainda é um desafio nas investigações forenses digitais. Mesmo que medidas tenham sido tomadas para abordar os recursos de segurança na IoT, os problemas relacionados ao DFR para sistemas IoT ainda permanecem nebulosos (BAJRAMOVIC *et al.*, 2016).

2.2.1.2 Identificação

O primeiro e mais importante passo no processo forense é a identificação de evidências (CONTI *et al.*, 2018). Em alguns casos os investigadores não sabem nem mesmo onde os vestígios estão armazenados (YAKUBU *et al.*, 2016).

Na forense digital tradicional, é possível delimitar a área de investigação e acelerar o processo como um todo (ORIWOH *et al.*, 2013). No ambiente IoT, entretanto, com a intensa comunicação entre dispositivos heterogêneos, é difícil delimitar o escopo investigativo, o que pode dificultar muito as investigações (CONTI *et al.*, 2018). Podemos citar ainda o caso de arquivos serem permanentemente deletados após o desligamento ou exclusão de máquinas virtuais na nuvem (ALEX; KISHORE, 2017).

Além disso, os dispositivos IoT podem migrar por diferentes redes enquanto em execução. Um atleta em uma maratona pode acessar mais de um provedor de nuvem, enquanto corre de uma cidade para outra, por meio de seu relógio inteligente. Imaginando um cenário mais amplo, os dispositivos IoT podem viajar de um país para outro e encontrar diferentes jurisdições, o que pode limitar a atuação legal dos investigadores (HAMMOUDI *et al.*, 2018).

2.2.1.3 Coleta de Evidências

Depois de identificar os dispositivos envolvidos na investigação, o passo seguinte será coletar as informações em si. Todavia, até o presente não há nenhum método padrão amplamente aceito para a coleta de evidência em um ambiente IoT que seja legalmente correto (CONTI *et al.*, 2018). Ser legalmente correto significa coletar as informações de maneira que elas tenham validade legal e respeitem a cadeia de custódia (MCKEMMISH, 2008). Sendo assim, qualquer erro na parte de coleta de evidência pode afetar o processo investigativo inteiro (ARSHAD *et al.*, 2018). Outros problemas que aparecem nesta categoria:

– Falta de treinamento: segundo o NIST (2006), erros comuns acontecem por falta de

conhecimento, tal como o desligamento de dispositivos antes de criar a imagem do sistema.

– Criptografia dos dados: hoje em dia, mais e mais aplicações dão suporte à criptografia ponta a ponta (CAVIGLIONE *et al.*, 2017). Isso significa que somente o dono dos dados tem acesso a eles, nem mesmo o provedor de nuvem tem a chave para descriptografar. Enquanto a criptografia preserva privacidade dos usuários, ela limita o poder dos investigadores. Um tradeoff precisa ser planejado para que a forense digital possa cumprir sua função (LOSAVIO *et al.*, 2018).
– Hardware e software heterogêneos: é sabido que cada fabricante desenvolve seus dispositivos IoT sem um padrão previamente acordado. O que leva a um conjunto de dispositivos com hardware e software diferentes (SAADEH *et al.*, 2016). Seria importante ter representações externas dos dados para facilitar o trabalho de leitura das ferramentas forenses (STOYANOVA *et al.*, 2020).
– Considerações éticas e de privacidade: muitos provedores de nuvem negam acesso aos dados de determinado usuário pois há o risco de acessar parte da memória compartilhada com outros usuários na nuvem que em nada têm a ver com a investigação (NIETO *et al.*, 2018; O'SHAUGHNESSY; KEANE, 2013). Uma nuvem *multitenant* tem a característica de compartilhar recursos com usuários de maneira simultânea.
– Falta de um modelo forense padrão: não existe um modelo padrão para a aquisição de evidências em dispositivos IoT (CONTI *et al.*, 2018). Diferentes organizações podem usar abordagens completamente distintas uma da outra. A existência de um padrão é imprescindível para manter a validade das evidências mesmo que coletadas por diferentes organizações e em diferentes localidades geográficas. Omissões no processo de coleta podem levar a complicações no tribunal (MCKEMMISH, 2008).

2.2.1.4 Preservação e Proteção de Evidências

Vencida a etapa de coleta de evidência, os investigadores agora terão um novo desafio: como resguardar a integridade dos dados. Existem, pelo menos, os seguintes desafios no que diz respeito à proteção de evidências:

– Protegendo a cadeia de custódia: como se sabe, a cadeia de custódia é um processo de documentação para assegurar que não houve adulteração nos dados analisados durante as etapas intermediárias. Para a manutenção da cadeia de custódia, as normas *International Organization for Standardization* (ISO) 27037:2012 e 10118-2:2010 sugerem a criação

da imagem do sistema a ser investigado e, posteriormente, a utilização de uma função *hash* para garantir a integridade dos dados. Existem, também, propostas que utilizam a tecnologia *blockchain*, como em (LONE; MIR, 2019).

– Tempo de vida: os dispositivos IoT possuem restrições de recursos computacionais. Informações pertinentes podem ser continuamente sobrescritas, pois a capacidade de armazenamento é pouca. Os dispositivos poderiam armazenar os dados em uma base centralizada, mas isso poderia comprometer a cadeia de custódia, já que poderia haver modificação durante a transmissão (YAQOOB *et al.*, 2019).
– Dados na nuvem: quanto maior flexibilidade é dada aos usuários da nuvem, mais responsabilidade eles terão para gerenciar a segurança e, consequentemente, analisar evidências. Por exemplo, no modelo Infraestrutura como Serviço, o usuário é capaz de instalar o Sistema Operacional e analisar seus *logs*. No Software como Serviço, não. Além disso, os equipamentos do provedor de nuvem podem estar espalhados por mais de um país, o que pode parecer problemático já que cada país ou estado mantém uma legislação distinta e pode não conceder o acesso aos dados investigados, embora eles pertençam ao mesmo usuário, pois a juridição é diferente (STOYANOVA *et al.*, 2020).

2.2.1.5 Análise e correlação de evidências

Considerando o cenário de IoT com um grande número de dispositivos dispersos geograficamente, a origem dos dados é incerta, dado que os dispositivos IoT não guardam metadados (STOYANOVA *et al.*, 2020), o que também dificulta a correlação dos dados. Entretanto, pode-se ter uma ideia de onde os dados foram originados dependendo do modelo de nuvem utilizado (O'SHAUGHNESSY; KEANE, 2013). Outro fator pertinente é que uma única informação pode ser fragmentada em dados espalhados por diferentes regiões, com diferentes fusos horários, além de diferentes jurisdições. Sem um relógio sincronizado de maneira global, os investigadores podem somente fazer uma especulação a respeito da relação entre os dados (CONTI *et al.*, 2018). E o fato de cada país ter seus próprios regulamentos também onera o processo investigativo (RANA *et al.*, 2017).

2.2.1.6 Responsabilização de ataques

A maioria dos provedores de serviços de nuvem possibilitam o acesso à sua infraestrutura mesmo com poucos informações de identificação fornecidas pelo usuário (STOYANOV,

2014). Aliado a outras ferramentas de obscuração de informações, isso deve complicar as atividades de identificação de usuários maliciosos (SAADEH *et al.*, 2016; RANA *et al.*, 2017).

Na infraestrutura de computação em nuvem, normalmente os usuários compartilham os mesmos recursos físicos, embora cada máquina virtual pertença a pessoas diferentes. Se um desses usuários praticar algum ato ilegal, os investigadores terão a responsabilidade de encontrar vestígios em um ambiente compartilhado por vários usuários que podem não ter ligação com tal ato (O'SHAUGHNESSY; KEANE, 2013).

2.2.1.7 Apresentação

Há fatores legais e de formatação que podem dificultar a aceitação das evidências coletadas. A respeito de fatores legais, há sistemas que obrigam os investigadores a explicarem a jurados como as evidências digitais foram colhidas e por que elas são válidas (STOYANOVA *et al.*, 2020). Ademais, Hegarty *et al.* (2014) afirmam que o uso de funções analíticas e de agregação de informação podem modificar a estrutura e o significado dos dados.

2.3 Modelo NIST para Forense Digital

O NIST é responsável por desenvolver padrões que ofereçam a adequada segurança da informação para todas operações e ativos de agências americanas (NIST, 2006). Nessa toada, o NIST propõe, na publicação especial 800-86, um processo de investigação forense digital em resposta a incidentes. O processo não é taxativo. Por isso, pode ser adaptado para melhor adaptar-se às necessidades da organização. Além disso, a publicação oferece informações de como proceder em investigação quando os dados forem coletados de arquivos, sistemas operacionais, tráfego de rede e aplicações.

De acordo com o NIST, a forense digital pode ser executada para suprir as seguintes necessidades:

- Busca por erros operacionais: configuração de rede e aplicações;
- Monitoramento de logs: correlação de logs dos sistemas para auxiliar na descoberta da causa raiz de problemas que levaram a incidentes, tais como violação de políticas;
- Recuperação de dados apagados ou modificados;
- Requisitos regulatórios: regulamentos podem prever que a organização mantenha informações de auditoria. O processo forense deve auxiliar no atingimento desses requisitos.

A publicação traz, ainda, aspectos relacionados à formação da equipe de investigação forense, ao tratamento de incidentes e à constituição de uma política que considere os dados sensíveis dos usuários. O documento ressalta, também, a importância de as organizações terem a capacidade forense. Segundo o NIST, a capacidade forense diz respeito à necessidade de praticamente todas organizações serem capaz de executar o processo forense em computadores e redes. Sem essa capacidade, as organizações terão dificuldades em determinar quais eventos causaram determinados incidentes. A publicação especial do NIST diz que os usuários das ferramentas e técnicas forenses são divididos geralmente em três categorias:

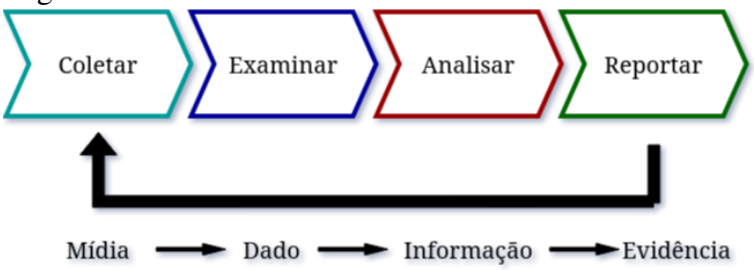
- **Investigadores**: são os responsáveis por investigar alegações de má condutas. Imediatamente assumem a investigação de qualquer evento que seja suspeito de envolver atividade criminal. Fazem uso de muitas ferramentas forenses;
- **Profissional de *Tecnologia da Informação* (TI)**: este grupo inclui administradores de sistema, rede e segurança da organização. Eles usam um número menor de ferramentas forenses;
- **Pessoas que lidam com incidentes**: respondem por incidentes de segurança da informação, tais como acesso não autorizado, infecção de código malicioso e ataques de negação de serviço. Tipicamente, utilizam uma grande variedade de técnicas e ferramentas forenses.

Abaixo estão as principais recomendações para estabelecer e organizar a capacidade forense:

- A organização deve saber executar ferramentas forenses em computadores e na rede;
- A organização deve determinar quais equipes executam quais atividades do processo;
- Os times de incidentes devem ser capazes de executar o processo forense;
- A forense deve ser compartilhada por vários times na organização;
- Todo o processo forense deve estar contido na política da organização; e
- A organização deve manter guias para auxiliar no processo forense.

Segundo NIST (2006), o principal objetivo do processo forense é adquirir melhor conhecimento de um evento, encontrando e analisando os fatos relacionados a ele. Para o processo de investigação forense digital, o NIST propõe quatro fases: coletar, examinar, analisar e reportar. Conforme a figura 3.

Figura 3 – Fases do Processo Forense



Fonte: adaptado de NIST (2006).

2.3.1 Coletar os dados

O primeiro passo do processo forense é identificar as origens dos dados e extraí-los. Os dados podem estar armazenados em pendrives, CDs, DVDs, discos magnéticos, cartões de memória, discos de estado sólido, *logs* e outros. Os investigadores também devem estar cientes de que os dados podem estar armazenados em outros lugares, como um provedor de acesso à Internet ou provedor de nuvem.

Para extrair os dados, algumas considerações devem ser tomadas, tais como: desenvolver um plano de aquisição de dados, de modo que ele traga considerações sobre a volatilidade dos dados, quantidade de esforço requerido para acessar os dados (quando estes estão fora do domínio da organização) e verificação da integridade dos dados. Este último passo é importante para assegurar a cadeia de custódia. Além de uma ferramente que verifique a integridade dos dados a cada acesso, é necessário o mantimento de registros, uma vez que todas as informações de local e posse dos dados são pertinentes à cadeia de custódia.

2.3.2 Examinar os dados

Uma vez que os dados foram coletados, é hora de examiná-los e decidir quais devem ser considerados para a investigação. Muitos dados podem ter passado por processos de ofuscação e/ou criptografia. É nessa fase que os investigadores tentam recuperar tais dados. Também é aqui que a equipe se depara com grande quantidade de dados, tais como *logs* de firewall, sistema operacional e rede, e fazem uso de ferramentas sofisticadas para filtrar o que mais pode ter importância para o processo forense. Isso é importante já que um log de sistema pode ter centenas de milhares de registros. Buscas por palavras-chaves e experiências anteriores podem diminuir o esforço da equipe.

2.3.3 Analisar os dados

Nesta fase, os profissionais devem analisar as informações obtidas para chegar a uma conclusão. O fundamento da ciência forense é usar uma abordagem metodológica para chegar a conclusões baseadas nos dados disponíveis ou determinar que não se pode chegar a qualquer conclusão (NIST, 2006). Aqui também ocorre a identificação de dispositivos, pessoas, lugares e como deve se dar a relação entre eles. Frequentemente, um conjunto de informações são cruzadas para se obter uma correlação entre elas e chegar-se a uma conclusão razoável. Por exemplo, um log do *Sistema de Detecção de Intrusão* (SDI) de rede pode indicar de qualquer host surgiu certa atividade maliciosa. O log do host, por sua vez, pode indicar a conta relacionada ao host que fez determinada ação.

2.3.4 Reportar

A última fase compreende atividades de preparar e apresentar as conclusões obtidas na fase de análise. É pertinente ter em mente que nem sempre as conclusões terão somente uma explicação para cada evento. É necessário que todas as explicações sejam aqui consideradas e que os profissionais utilizem abordagem metodológicas para convencer a audiência sobre suas explanações. Além disso, é importante que o nível de detalhes das explicações atendam as expectativas da audiência.

2.4 Forense Digital em Prédios Inteligentes

Para projetar novos modelos forenses de investigação em IoT, é importante pensar na sua aplicabilidade em aplicações reais da Internet das Coisas, tais como os Prédios Inteligentes.

Os Prédios Inteligentes são construções que possuem a capacidade de reprogramar os dispositivos embarcados presentes no ambiente, com o auxílio de sensores (hardware), software e rede, de maneira que o local se adapte às necessidades das pessoas que ali estão. Luzes, termostato e trancas de portas são exemplos de artefatos que podem ser ajustados para melhor atender à demanda no ambiente. Os prédios inteligentes monitoram o local, observando o padrão de comportamento ali existente, e tentam predizer seus estados futuros (BATOV, 2015). Dessa forma, um edifício inteligente pode determinar a ocupação dos quartos, intensidade de luz, temperatura interna e externa, nível de dióxido de carbono, nível de ruído, detectar vazamento de gás e assim por diante.

Alguns dos potenciais benefícios da concepção de prédios inteligentes são (BATOV, 2015):

- Conforto das pessoas: prédios inteligentes aprendem com o comportamento dos habitantes e procuram maximizar seu conforto;
- Economia de energia: podem reduzir significativamente o consumo de energia. Por exemplo, ao desligar ar-condicionados e luzes quando pessoas não estão presentes;
- Economia de tempo: podem economizar muito tempo ao automatizar rotinas;
- Segurança: podem detectar fogo, vazamento de água e gás. Existem sistemas de auto-diagnóstico que avisam aos responsáveis quando equipamentos apresentam defeito ou o desempenho começa a diminuir;
- Saúde e cuidados: em todas as decisões, a saúde das pessoas tem a maior prioridade. Isso tem reflexo na temperatura apropriada, intensidade de luz, parâmetros do ar condicionado, etc.

A importância de prédios inteligentes remete, pelo menos, ao ano de 2001, quando o prédio do pentágono, após sofrer o ataque terrorista de 11 de setembro, conseguiu controlar o fogo a uma área restrita. O pentágono possuía uma rede de sensores e controladores que permitiram o isolamento das chamas (SNOONIAN, 2003).

2.4.1 A importância da forense digital em prédios inteligentes

Apesar das inúmeras vantagens presentes em prédios inteligentes, riscos em segurança da informação estão presentes, principalmente nas redes que integram os dispositivos em tais prédios (KHAUND, 2015). Tecnologias emergentes, como sensores inteligentes, são fornecidas sem testes apropriados de segurança. Isso porque os fabricantes preocupam-se mais com a facilidade de uso do que com a segurança dos dados (BAJRAMOVIC *et al.*, 2016).

Como se pode perceber, a vulnerabilidade de prédios inteligentes também está relacionada às vulnerabilidades das tecnologias presentes no ambiente. Como discutido anteriormente, ainda não existem padrões para as redes e mecanismos de segurança, como protocolos, que propiciem ambientes inteligentes, tais como casas inteligentes, prédios inteligentes, entre outros. Assim, como em qualquer outro ambiente de IoT, os prédios inteligentes tornam-se alvo em potencial para ataques à segurança da informação. Por isso, é de fundamental importância que existam processos proativos de monitoramento de eventos, coleta e preservação de evidências digitais para que se possa determinar adequadamente as causas e consequências de eventos de

segurança da informação que possam interromper os serviços do negócio.

2.4.2 Requisitos de um modelo investigativo em Prédios Inteligentes

Os prédios inteligentes são construções físicas, como casas, hospitais, empresas e demais empreendimentos, capazes de adaptarem-se às necessidades das pessoas que transitam por eles. Além disso, contam com software, hardware e rede para a troca de informações entre seus diversos dispositivos de IoT. Hospitais que contam com prédios inteligentes, por exemplo, possuem uma gama de tecnologias específicas para suprir a necessidade do negócio. Por isso, constituem um ambiente específico que certamente requer tarefas específicas para a execução de processos forenses.

Ademais, os sistemas informatizados em Prédios Inteligentes precisam estar devidamente protegidos contra invasores, pois existem dados sigilosos dos clientes e de estratégia da organização. Como é sabido, as melhores práticas do mercado indicam o uso de processos de monitoramento e gerenciamento de eventos para que possíveis incidentes não venham comprometer a missão da organização (GËRVALLA *et al.*, 2018). É importante que empreendimentos informatizados, independentemente do tamanho, sigam algumas práticas relacionadas a ações reativas e proativas de monitoramento e gerenciamento de eventos em serviços de TI, como as que constam na biblioteca *IT Infrastructure Library* (ITIL). Desse modo, é importante que os modelos forenses para Prédios Inteligentes tratem do processo de gerenciamento e monitoramento de eventos, já que os vestígios digitais podem ser capturados e preservados mesmo em fases proativas do gerenciamento de eventos. Como já mencionado, essa característica de *forensic readiness* acelera o processo de identificação de causas e barateia os custos. Cabe mencionar, também, que para preservar tal característica, a preservação dos dados deve seguir um procedimento o qual permita a validação dos dados quando apresentados na corte, que diz respeito, consequentemente, à preservação da cadeia de custódia.

Além da fase proativa, as organizações necessitam de tarefas reativas a eventos internos e externos, desde que sejam pertinentes à organização. Para essas tarefas, o NIST elaborou o documento especial 800-86 que provê um modelo forense para o tratamento da informação durante todo o seu ciclo de vida, salvaguardando a cadeia de custódia das informações. Como o modelo do NIST preocupa-se em tornar judicialmente válido o processo forense, é importante que as organizações sigam os conceitos e propostas de órgãos desse tipo, de tal forma que respeitem a execução de cada uma das fases propostas, diminuindo, assim, o risco de não

aceitação de dados como evidências (NIST, 2006). Não só isso, operando desde 1901, o NIST é um instituto que tem como missão a promoção de inovação e competitividade industrial pelo avanço da ciência, padrões e tecnologia, de tal forma que melhore a segurança econômica e melhore a qualidade de vida dos cidadãos (NIST, 2008).

O modelo proposto pelo NIST é holístico, ou seja, passa por todas as fases da investigação forense, desde a coleta dos dados até a apresentação de evidências na corte. Sendo assim, não há impeditivos ao uso de outros guias e melhores práticas cujo escopo limite-se dentro de casa fase. Alguns dos guias são:

- **ISO/IEC 27035 – Part 1**: Principles of incident management (ISO/IEC, 2016);
- **ISO/IEC 27037** – Guidelines for identification, collection, acquisition and preservation of digital evidence (ISO/IEC, 2012);
- **ISO/IEC 27041** – Guidance on assuring suitability and adequacy of incident investigative method (ISO/IEC, 2015b);
- **ISO/IEC 27042** – Guidelines for the analysis and interpretation of digital evidence (ISO/IEC, 2015c);
- **ISO/IEC 27043** – Incident investigation principles and processes (ISO/IEC, 2015d);
- **ISO/IEC 30121** – Governance of digital forensic risk framework (ISO/IEC, 2015a);

Quando do estabelecimento de modelos (genéricos ou específicos) e fases, é imprescindível que em algum momento todos devam ser testados na prática. Isso é importante para validar as propostas de cada um. O mesmo vale para modelos específicos para prédios inteligentes.

Outro fator não menos importante, é que o escopo da investigação em dispositivos IoT não se limita a dispositivos presentes somente na rede local. Segundo Zawoad e Hasan (2015), as zonas de investigação dividem-se em dispositivo, rede local e nuvem. Isso é essencial para a condução de uma investigação forense mais abrangente.

Além disso, é preciso que as organizações tenham pessoal treinado e capaz de executar as tarefas em um processo forense. Deve haver guias e manuais de instruções para que os responsáveis possam usar procedimentos e ferramentas necessárias para a coleta dos dados pertinentes (NIST, 2006).

2.5 Conclusão

Neste capítulo, foram apresentados os conceitos básicos, arquitetura e camadas da Internet das Coisas. Depois, foram discutidos desafios de segurança da informação inerentes a um ambiente IoT. Posteriormente, foi ressaltada a importância de se investigar as causas dos incidentes de segurança utilizando ferramentas, como a ciência forense, para a coleta de evidências que possam ser aceitas na corte. Em seguida, foi apresentado o conceito de forense digital junto com as fases do modelo proposto pelo NIST para coleta, análise e apresentação de evidências digitais.

Finalmente, o capítulo conceituou Prédios Inteligentes, uma aplicação real da Internet das Coisas, e como a utilização de processos investigativos forenses podem auxiliar na descoberta e responsabilização dos envolvidos em eventos dentro da organização.

O próximo capítulo traz um levantamento bibliográfico a respeito dos modelos forenses existentes para ambientes IoT, inclusive para Prédios Inteligentes.

3 TRABALHOS RELACIONADOS

Sabe-se que, na ocorrência de incidentes em segurança da informação, é importante conhecer métodos e ferramentas que auxiliem na coleta de evidências para o descobrimento das causas de cada evento. Além disso, é imprescindível que as informações coletadas sejam preservadas de modo que seja possível atribuir responsabilidades de maneira devida. Embora existam processos que viabilizem a condução da investigação forense em ambientes heterogêneos, como a IoT, nem todos seguem um modelo cujas fases assegurem a preservação da cadeia de custódia e, assim, a manutenção da validade judicial das evidências coletadas.

Neste capítulo, é realizado um levantamento bibliográfico acerca dos modelos de investigação forense, especialmente para a Internet das Coisas. Após a discussão de cada modelo apresentado, é feita uma avaliação qualitativa dos trabalhos discutidos. Depois, descreve-se como os requisitos dessa avaliação aplicam-se a modelos investigativos para Prédios Inteligentes.

3.1 Levantamento bibliográfico

Zawoad e Hasan (2015) definem um modelo para o processo forense na IoT em três níveis: dispositivo, rede e nuvem. Cada nível representa um escopo diferente e limita o local de investigação. No escopo de dispositivo, a atenção deve ser direcionada para os periféricos locais da rede interna, tais como computadores pessoais, impressoras e dispositivos de armazenamento. No nível de rede, a investigação deve ocorrer nos dispositivos de borda da rede, como roteadores e firewalls. No nível da nuvem, os investigadores devem recorrer à análise de dispositivos que estão fora do domínio da organização, provavelmente o processo forense ocorrerá em um provedor de nuvem.

A definição de cada uma dessas fases é baseada em uma definição do NIST de forense digital. Esse modelo proposto também oferece um repositório central para tratar a cadeia de custódia. Segundo os autores, este foi o primeiro trabalho que definiu formalmente o termo forense IoT, que é a aplicação da ciência forense ao paradigma da Internet das Coisas. Entretanto, a eficácia deste modelo de forense digital é difícil de ser verificada, pois os autores limitaram-se à teoria.

O modelo de processo *The Next Big Thing* (ORIWOH *et al.*, 2013) foi proposto com foco na identificação de dispositivos que contenham evidências. Ele baseia-se nas zonas 1, 2 e 3. Na zona 1, é identificada a pessoa que causou o incidente e, possivelmente, produziu as

evidências. A zona 2 cobre todos os dispositivos na rede interna, tais como roteadores, firewall, switches, SDI e gateways. Todos os dispositivos fora da rede interna são identificados na zona 3. O desafio existente neste modelo diz respeito a capacidade de implementação e teste, pois os investigadores podem não ter permissão de acesso aos dispositivos do provedor de nuvem. O modelo também não considera a volatilidade dos dados armazenados fora dos domínios da organização, o que significa que a investigação pode ser realizada com uma quantidade de dados menor do que o esperado. Possui as fases de preparação, aquisição, investigação e reportar e armazenar. Apesar de ser um processo abrangente, ao considerar dispositivos locais e remotos, ele não trata a cadeia de custódia, podendo, assim, comprometer a validade das evidências apresentadas na corte.

Perumal *et al.* (2015) integraram as zonas apresentadas no *The Next Big Thing*. Este modelo começa com as fases de autorização e planejamento, na forma de um Procedimento de Operação Padrão. Depois, segue pela fase de identificação de dispositivo e comunicação M2M.

Uma vez que as mídias forem identificadas, os investigadores seguem para a fase de triagem. Essa fase precisa ser executada de maneira cuidadosa, pois nela existem dados estruturados e não estruturados. Segundos os autores, os dispositivos comuns nessa fase seriam os roteadores, gateways e plataforma de nuvem. Após o processo de extração de dados, o modelo segue os processos de preservação da cadeia de custódia, resultado e arquivamento.

Este modelo tenta resolver o problema de como lidar com a preservação de dados voláteis, e como isso deve ajudar os investigadores. Todavia, não há como assegurar que ele realmente funcione, já que os autores não o implementaram em um ambiente real.

Quick e Choo (2018) utiliza a automatização no processo de coleta de dados, que analisa rapidamente quais dados deverão ser considerados para a investigação. Segundos os autores, é necessário coletar dados de fontes variadas e realizar análises rápidas em diversas estruturas de dados, que auxiliem na identificação de evidências, em tempo hábil. Conforme demonstrado nesta pesquisa, o uso de um processo de redução de dados e análise semiautomática, com o software Bulk Extractor, permite a análise oportuna de uma ampla gama de dados díspares. Os autores seguem o Digital Forensic Intelligence Analysis Cycle (QUICK; CHOO, 2017) para a definição das fases.

Apesar da proposta de diminuir o tempo de análise dos dados como um todo, o experimento mostrado no artigo ainda demorou bastante. Os autores também relatam a dificuldade que existe no processo de extração de dados, pois fabricantes costumam utilizar

estruturas de dados diferentes. Outro ponto negativo é que o artigo se propôs a analisar dados de dispositivos dissimilares, entretanto os resultados não mostraram a quantidade desses dispositivos. O escopo do Bulk Extractor também poderia ser expandido para outros tipos de dados específicos.

Já o modelo *Mobility Forensics* (RAHMAN *et al.*, 2016) considera um ambiente inteligente com dispositivos móveis. Ele faz considerações desde o processo de coleta dos dados. Além disso, oferece uma discussão detalhada de um cenário de ataque com o modelo que lida com o incidente. O modelo proposto procura determinar o que aconteceu, quando aconteceu, quem ou o que causou, por que aconteceu e quais dados foram coletados. Essas são perguntas que permeiam o processo convencional de investigação forense digital. Todavia, o modelo procura responder as questões com a perspectiva direcionada a um ambiente de Internet das Coisas.

Dentre algumas limitações do modelo, pode-se verificar que ele utilizou somente dados de dispositivos inteligentes específicos; ele não foi implementado ou testado. Também foi assumido que o modelo será escalável a um grande número de dispositivos, o que pode não ser verdade. Os autores consideraram apenas um tipo de dispositivo, o que não reflete a realidade de ambientes heterogêneos, tais como a IoT.

Zia *et al.* (2017) propuseram um modelo forense investigativo no qual os dados são extraídos, examinados e analisados. Esse modelo é constituído de três grandes domínios: casas inteligentes, cidades inteligentes e wearables.

Segundo os autores, um dos maiores desafios na IoT é lidar com a natureza heterogênea das "coisas" que trazem fragilidades de segurança, tornando-as vulneráveis a invasões e ataques. Isso inspirou a necessidade de medidas forenses digitais exclusivas que podem abordar a coleta, exame, análise e relato de evidências em sistemas IoT de aplicações específicas.

Nesse artigo, foi apresentado um modelo forense digital específico para certas aplicações, apontando os artefatos de maior importância forense na IoT. Adicionalmente, foi mostrada uma abordagem forense holística que abrange as melhores práticas existentes na indústria forense digital. Esse artigo define também um cenário para o desenvolvimento de processos, diretrizes e ferramentas forenses digitais, específicos a certas aplicações, que seriam benéficos em investigações corporativas.

Vale ressaltar que os autores não propuseram uma abordagem prática do modelo. Assim, esse trabalho é somente teórico. Ademais, o modelo não considerou a presença de protocolos de segurança. Como, por exemplo, os que oferecem confidencialidade por meio de criptografia. Isso é um grande desafio, haja vista a disponibilidade de dados apenas criptografados.

Em Harbawi e Varol (2017) foi proposto um modelo baseado no processo *The Next Big Thing*. O algoritmo busca determinar por onde os investigadores devem começar as buscas, que é no último dispositivo que apareceu na cadeia de comunicação. Isso gera economia de recursos e tempo, já que os profissionais não terão que analisar todas as zonas.

Com base em vários estudos da literatura, recursos online e conhecimento dos autores, um procedimento para o modelo de aquisição de evidência digital para forense IoT foi fornecido na primeira fase do estudo. O primeiro e o mais importante passo no procedimento é a implantação do algoritmo *Last-on-Scene*, que melhora a rastreabilidade e reduz a sobrecarga, bem como as complicações da análise forense digital. Além disso, foi feita uma revisão para forense IoT com base na estrutura teórica proposta, a fim de garantir a usabilidade do procedimento de aquisição.

Algumas das limitações desse framework são a falta de uma aplicação prática do modelo e o não tratamento da cadeia de custódia.

Em Zulkipli *et al.* (2017) os autores propuseram duas abordagens para conduzir investigações forenses em um ambiente IoT. Na abordagem de DFR, são tratados elementos que assegurem que o ambiente esteja pronto para passar por processos de investigação, antes mesmo do incidente acontecer. Na abordagem de tempo real, o modelo conta com componentes que são capazes de detectar atividades anormais e, posteriormente, iniciar as fases de pré-investigação, além de executar etapas de identificação, coleta e preservação de evidências concorrentemente.

Esse modelo parece ser muito geral porque considera que o processo convencional de forense digital seja aplicado à IoT, ou seja, não leva em conta que os dados podem não estar disponíveis em tempo oportuno. Assim, essa abordagem não funcionará se o investigador não tiver acesso total à rede ou dispositivos investigados. Além disso, nenhuma parte prática foi feita pra ilustrar a proposta dos autores.

Nieto *et al.* (2017) apresentaram o modelo PRoFIT para conduzir investigações forenses digitais em ambientes de IoT. Ao contrário das abordagens anteriores, o modelo PRoFIT integra requisitos de privacidade (ISO / IEC 29100: 2011) como parte da metodologia. Segundo os autores, o objetivo de considerar a privacidade é promover a colaboração voluntária de dispositivos IoT pessoais e não pessoais em investigações forenses digitais. Apesar de a metodologia proposta ter sido aplicada a um cenário de caso de uso realista (propagação de *malware* em uma cafeteria com dispositivos IoT), não existe implementação do modelo.

Em Meffert *et al.* (2017), os autores apresentaram o framework FSAIoT e ilustraram

sua viabilidade implementando uma prova de conceito usando a aplicação OpenHAB, bem como scripts para validar a viabilidade do framework. Mostraram que podem coletar dados de estado de dispositivos IoT de maneira confiável usando três modos diferentes: controlador para dispositivo, controlador para nuvem e controlador para controlador. Mesmo considerando as limitações, foi afirmado que a estrutura proposta e a prova de conceito são etapas essenciais na evolução de uma metodologia geral para obter evidências forenses valiosas em um ambiente diversificado de dispositivos de IoT.

A maior limitação desse trabalho está em acessar dados históricos e dados deletados. O segundo desafio diz repeito à necessidade de acesso físico aos diferentes dispositivos IoT. Esse é um desafio comum na foresense digital, pois nem todos os dispositivos podem ser acessados fisicamente. Uma outra limitação pertinente no trabalho está relacionada à conexão a diferentes dispositivos com tecnologias variadas. Dentre os métodos comuns de conexão wireless estão o bluetooth, o Zigbee e o Zwave. Para conseguir acessar dados via Zigbee ou Zwave, por exemplo, é necessário ter hardware que suporte essas tecnologias de comunicação. Apesar de não ser comum em computadores pessoais tradicionais, módulos que suportem essas tecnologias estão disponíveis no mercado. Além disso, o framework não garante a cadeia de custódia, o que abre a possibilidade de perda de integridade das evidências.

Kebande *et al.* (2017) discutem os desafios para aquisição de evidências em cenários que buscam proteger a anonimidade dos usuários e confidencialidade das informações. O principal objetivo desse artigo foi propor uma estrutura segura para isolar Big data como evidência forense em infraestruturas IoT baseadas em nuvem.

Os autores afirmam que: “ainda não existem padrões ou estruturas aceitas para a realização de investigação forense digital em infraestruturas de IoT baseadas em nuvem”. Como resultado, eles propuseram o CFIBD-IoT, por meio do qual evidências forenses puderam ser extraídas de ambientes de IoT na nuvem.

Os autores apresentaram isso por meio de duas abordagens, abordagem de alto nível e uma abordagem detalhada. Além disso, facilitaram a visualização do processo CFIBD-IoT usando um diagrama de atividades da UML, que mostrou ao leitor o fluxo de eventos da estrutura proposta. Os autores afirmaram, ainda, que se a estrutura for totalmente implementada, ela suportará a criação de ferramentas IoT baseadas em nuvem e também poderá oferecer suporte a futuras técnicas de investigação na nuvem.

Além de não existir aplicações práticas do modelo, os autores afirmar que as atuais

ferramentas de extração de evidências em grandes conjuntos de dados possuem grande limitação porque não existem padrões para isso. Sendo assim, mesmo que o framework proposto seja adotado, a aplicabilidade torna-se limitada como consequência das restrições de tais ferramentas.

Segundo Babun *et al.* (2018), dispositivos inteligentes e sensores presentes em ambientes inteligentes têm acesso a dados que podem ser usados para fins forenses. No entanto, as plataformas atuais de programação de aplicativos inteligentes não oferecem nenhum recurso para rastrear informações relevantes para a forense digital. Além disso, as soluções atuais de análise forense não usam informações de aplicativos inteligentes e/ou dispositivos inteligentes para realizar investigações forenses.

Nesse trabalho, foi apresentada a IoTDots, uma nova estrutura usada para extrair logs de aplicativos inteligentes relevantes para fins forenses e para analisá-los de maneira automática. A estrutura tem dois componentes principais: IoTDots-Modifier e IoTDots-Analyzer. O modifier realiza análise de código-fonte de aplicações inteligentes, detecta dados relevantes para a forense no código-fonte e insere logs específicos em tempo de compilação. Em seguida, em tempo de execução, os logs são enviados para um servidor IoTDots remoto. Em um caso de investigação forense, o Analyzer aplica técnicas de processamento de dados e aprendizado de máquina para extrair informações forenses dos logs do IoTDots.

De acordo com os resultados, o IoTDots atinge mais de 98% de precisão na detecção de atividades de usuário e mais de 96% de precisão na detecção do comportamento de usuários e aplicações inteligentes. Além disso, o IoTDots executa com mínima ou nenhuma sobrecarga nos dispositivos inteligentes testados e possui baixa sobrecarga para o servidor de nuvem IoT.

Algumas de suas limitações são a especificidade, já que nem todos dispositivos podem rodar as aplicações propostas pelos autores, e a falta de mecanismo para garantir a cadeia de custódia.

Em Hossain *et al.* (2018) os autores propõem o Probe-IoT, uma estrutura de investigação forense para sistemas baseados em IoT, usando um livro-razão digital público. O Probe-IoT armazena interações de dispositivo com dispositivo, dispositivo com usuário e dispositivo com nuvem como evidência, em um livro-razão digital público. O Probe-IoT garante confidencialidade, anonimato e não repúdio às evidências publicamente disponíveis. O Probe-IoT também fornece interfaces para aquisição de evidências e um esquema para verificar a integridade durante a investigação de um incidente criminal.

Nesse trabalho, Os autores poderiam ter explicado melhor como as características de

segurança são mantidas por meio do livro-razão digital. Ainda, a validação experimental dessa proposta ficou fora do escopo do artigo.

Em Sadineni *et al.* (2019) é argumentado que devido à diversidade de dispositivos, redes e aplicativos, várias soluções forenses digitais ad hoc foram desenvolvidas para ambientes específicos de Internet das Coisas. Por isso, um modelo forense digital holístico, que cobre diversos ambientes de IoT, é necessário para eliminar a sobrecarga imposta pelas soluções existentes.

O modelo forense da Internet das Coisas apresentado nesse artigo é holístico e cobre todo o ciclo de vida forense. O modelo, que é baseado no padrão internacional ISO/IEC 27043, é personalizável, configurável e oferece suporte a diversas aplicações da Internet das Coisas. Não obstante, os autores também ficaram devendo a implementação e o teste do modelo em domínios de aplicações IoT.

Já em Hossain *et al.* (2017) foi proposto um framework para investigar casos criminais nos sistema de Internet of Vehicles (IoV). Os autores identificaram peças de evidências que poderiam ser usadas para encontrar fatos sobre incidentes criminais e propuseram um serviço para coletar e armazenar evidências de maneira segura. A integridade das evidências foram asseguradas usando uma cadéia de proveniência à prova de violações. Também foi apresentado um algortimo que permite que os investigadores verifiquem a integridade das evidências. Finalmente, foi apresentado um estudo de caso e executada uma análise de segurança do sistema proposto. A análise mostrou que o sistema funciona bem em cenários com fortes adversidades. Um protótipo do Trust-iov foi implementado usando um simulador para verificar a aplicabilidade do modelo em um sistema IoV.

Os autores de Kebande e Ray (2016) afirmam que atualmente não existe nenhuma estrutura forense digital aceita que possa ajudar a conduzir uma investigação sobre os ambientes de IoT. Como resultado, eles propuseram a estrutura DFIF-IoT que tem uma abordagem holística e capacidade forense que planeja e prepara o ambiente IoT antes que potenciais incidentes de segurança possam ocorrer nele.

Além disso, os autores também incluíram processos concorrentes nesse modelo, de acordo com a norma ISO/IEC 27043:2015, de modo que os processos possam ser executados continuamente e permitam a garantia de admissibilidade na corte. Ademais, uma comparação com os modelos existentes foi feita, o que revelou ainda mais a eficácia da estrutura DFIF-IoT proposta. Entretanto, algumas afirmações feitas nesse trabalho só podem ser verificadas usando

um protótipo funcional, o que não foi oferecido pelos autores.

FoBI (AL-MASRI *et al.*, 2018) é um framework que utiliza um software instalado em um dispositivo da rede que atua como gateway entre a *fog computing* e a *edge computing*. Ele coleta e publica informações sobre dispositivos da rede. Dessa forma, ele auxilia no rastreio de ameaças. Basta perceber que informações sobre ataque a determinado dispositivo sejam publicadas na rede. Assim, as consequências do ataque podem ser evitadas ou minimizadas. Entretanto, o fato de um software adicional precisar ser instalado, pode causar problemas na aceitação dos dados na corte. Isso porque não há garantia que os dados não sofreram mutações.

Em Kebande *et al.* (2018), foi proposta uma estrutura integrada de investigação forense digital para ambientes IoT. Os autores delinearam uma abordagem promissora que mostrou os componentes que foram usados para projetar o *framework* IDFIF-IoT. Posteriormente, os autores foram capazes de propor uma estrutura completa que pode auxiliar nas abordagens de detecção pré-incidente em ambientes IoT. Isso foi feito com a conveniência que permite que o ambiente baseado em IoT seja preparado judicialmente para possíveis incidentes de segurança cibernética. Os processos pré-incidente foram propostos em conformidade com o padrão internacional ISO/IEC 27043. Não possui protótipo de implementação do modelo para que aspectos críticos do processo forense possam ser identificados.

Para Bajramovic *et al.* (2016), embora o DFR seja opcional para o processo de investigação digital, é altamente recomendável que seja implementado para sistemas digitais em um mundo inteligente. Conforme indicado pelos autores, as evidências mostram que uma quantidade esmagadora de incidentes de segurança cibernética não pode ser resolvida devido à falta de artefatos forenses. Portanto, o DFR e os testes de segurança cibernética devem ser estabelecidos em conjunto para prédios inteligentes, a fim de melhorar a resposta a incidentes de segurança. Os autores também propuseram um processo de DFR para prédios inteligentes, com uma visão um pouco detalhada de como o processo deveria ser implementado. O modelo forense proposto nesse trabalho foi o único encontrado com relação especificamente a Prédios Inteligentes. Apesar disso, ele é muito superficial, sem descrição das atividades do processo nem aplicação prática.

Os autores Dorai *et al.* (2018) desenvolveram uma ferramente para automatizar o processo de análise de evidências em dispositivos móveis na IoT. Os autores documentaram uma abordagem para recuperar artefatos de dispositivos móveis. Eles foram capazes de encontrar dados relevantes, em base de dados recuperadas de dispositivos móveis, os quais revelaram a data

e hora que certos eventos aconteceram. Esses eventos denotaram quando um termostato foi recalibrado, se um usuário estava em casa em certo momento ou se uma câmera foi intencionalmente desligada em algum momento, por exemplo.

Apesar dos resultados, a ferramenta utilizada está disponível apenas para algumas versões do iPhone, o que não garante a eficácia dos resultados em versões futuras do sistema operacional. Além disso, a ferramente funciona apenas quando os dados estão disponíveis, ou seja, não trata a volatilidade ou não disponibilidade dos dados.

Mais recentemente, o trabalho de Alam e Kabir (2023) propôs um modelo de investigação específico para casas inteligentes, que é composto por cinco fases: identificação e coleta de múltiplas fontes, análise, interpretação, apresentação das evidências e conclusão. Apesar de o modelo preocupar-se com tarefas importantes como a cadeia de custódia e as fases bem conhecidas da investigação forense, não existe qualquer implementação. Também foi apresentado um estudo de caso muito simples.

Já em Kim *et al.* (2023). foi implementada uma ferramenta de prova de conceito que mostra a conexão entre dispositivos IoT. Além disso, foi proposta uma arquitetura chamada SIIFF para investigar incidentes associados a serviços e dispositivos IoT. Ademais, duas fases distintas aos modelos tradicionais foram fornecidas para aprimorar as atividades forenses de IoT. No entanto, o trabalho tem limitações, pois os autores concentraram-se apenas em artefatos extraídos de um dispositivo baseado em Android.

3.2 Avaliação Qualitativa

Com base nos trabalhos relacionados levantados até aqui, foi elaborada uma tabela comparativa entre eles, considerando oito requisitos que se mostram relevantes para modelos de investigação forense em IoT. A importância e significado de cada um dos requisitos foram discutidos no Capítulo 2. Os requisitos são identificados por siglas, sem ordem de importância, conforme o Quadro 1.

1. **RQ1**: Preserva a cadeia de custódia;
2. **RQ2**: Executa todas as fases do NIST;
3. **RQ3**: Planeja a Capacidade Forense;
4. **RQ4**: Segue outros modelos, normas, melhores práticas ou padrões;
5. **RQ5**: Existe aplicação prática do modelo;
6. **RQ6**: Cobre os três níveis da forense para IoT;

7. **RQ7**: *Digital Forensic Readiness*;
8. **RQ8**: Representação padronizada das evidências.

Quadro 1 – Comparação entre modelos forenses para IoT

Modelo	RQ1	RQ2	RQ3	RQ4	RQ5	RQ6	RQ7	RQ8
Oriwoh *et al.* (2013)		✓				✓		
Zawoad e Hasan (2015)	✓	✓				✓		
Perumal *et al.* (2015)	✓	✓				✓		
Rahman *et al.* (2016)				✓				
Kebande e Ray (2016)	✓	✓		✓		✓	✓	
Bajramovic *et al.* (2016)							✓	
Zia *et al.* (2017)	✓	✓						
Harbawi e Varol (2017)						✓		
Zulkipli *et al.* (2017)	✓	✓					✓	
Nieto *et al.* (2017)	✓	✓		✓			✓	
Meffert *et al.* (2017)					✓			
Hossain *et al.* (2017)	✓				✓			
Kebande *et al.* (2017)	✓			✓				
Quick e Choo (2018)	✓			✓	✓			
Babun *et al.* (2018)					✓			
Hossain *et al.* (2018)	✓				✓	✓		
Al-Masri *et al.* (2018)				✓			✓	
Dorai *et al.* (2018)					✓			
Kebande *et al.* (2018)	✓	✓		✓		✓	✓	
Sadineni *et al.* (2019)	✓	✓		✓		✓	✓	
Alam e Kabir (2023)	✓	✓		✓		✓		
Kim *et al.* (2023)	✓	✓		✓	✓	✓		
Smart Building Investigation Model	✓	✓	✓	✓	✓	✓	✓	✓

Fonte: Elaborada pelo autor.

Como se pode perceber, os modelos estão divididos em duas categorias: modelos mais genéricos, que consideram um ambiente IoT com grande heterogeneidade de dispositivos, e modelos mais específicos, que objetivam nortear o processo investigativo em ambientes mais específicos, sem grande variedade de dispositivos presentes, como é o caso dos modelos *Mobility Forensics* (RAHMAN; SHAH, 2016), *Application-specifc Digital Forensics* (ZIA *et al.*, 2017), *IoT-dots* (BABUN *et al.*, 2018), e *Trust-iov* (HOSSAIN *et al.*, 2017). A grande maioria dos modelos existentes trata do processo forense de maneira genérica. Um possível problema para abordagens desse tipo é que os modelos devem ser capazes de lidar com um grande número de estrutura de dados na IoT, já que não existem padrões aceitos amplamente e cada fabricante pode definir seus próprios tipos de dados. Por outro lado, os modelos mais específicos podem ser validados mais facilmente, pois se aplicam a situações mais restritas. Não obstante, isso não quer dizer que sejam modelos menos eficazes àquilo que se propõem, como é o caso dos modelos *Mobility Forensics* (RAHMAN; SHAH, 2016) e *Application-specifc Digital Forensics* (ZIA *et al.*, 2017).

O primeiro requisito avaliado na tabela é a preservação da cadeia de custódia, conceito já discutido na Seção 2.2. O segundo requisito avalia a presença de fases que se confundem com as do NIST, mostradas na Seção 2.3. O terceiro requisito preocupa-se com avaliar a presença de atividades proativas no planejamento da capacidade forense, cujo conceito é apresentado na Subseção 4.2.2. A importância de tais atividades foi elucidada na Seção 2.3. O quarto requisito, por sua vez, verifica se a proposta dos autores segue outros modelos ou melhores práticas que não as do NIST. Sejam modelos de execução do processo investigativo ou modelos de representação de dados coletados. Seguir modelos, padrões ou melhores práticas auxilia numa melhor comunicação entre as partes interessadas e no êxito do processo. Já o quinto requisito verifica se o modelo proposto é apenas teórico ou se existe uma aplicação prática no todo ou em parte. O sexto requisito mostra qual proposta satisfaz os três níveis da forense em ambiente IoT, conforme Zawoad e Hasan (2015). Essa divisão do escopo de investigação forense é bem conhecida na literatura por estar presente em mais de um trabalho. Por sua vez, o requisito sete verifica a existência de práticas de DFR nos trabalhos relacionados. A DFR é definida e discutida na Subseção 2.2.1.1. Finalmente, o requisito 8 identifica a presença de alguma etapa de padronização dos dados elencados como possíveis evidências no processo investigativo, os quais podem ser *logs* ou quaisquer outros artefatos relevantes.

Como se pode notar, o SBIM é o único modelo que atende a todos os requisitos, executando atividades proativas e contínuas para o melhoramento do processo forense como um todo.

3.3 Conclusão

Neste capítulo, foi realizada uma apresentação de trabalhos relacionados a modelos de investigação forense para Internet das Coisas. Os trabalhos foram brevemente discutidos e avaliados qualitativamente considerando-se um conjunto de requisitos desejáveis em processos desse tipo. Com a comparação realizada, notou-se, principalmente, a ausência de processos que se preocupem com o planejamento da capacidade forense da entidade executora das atividades investigativas.

Outro fato importante foi a percepção de que, basicamente, as contribuições até o momento propuseram processos genéricos (que se aplicam à qualquer aplicação de IoT) ou específicos (que se aplicam a aplicações específicas de IoT). Nesse contexto, concluiu-se que inexistem trabalhos, até o momento, específicos para a aplicação do processo forense em Prédios

Inteligentes.

No próximo capítulo, é apresentado o SBIM, um modelo de investigação forense eficaz e eficiente para prédios inteligentes.

4 SBIM: UM MODELO DE INVESTIGAÇÃO FORENSE PARA PRÉDIOS INTELIGENTES

A Internet das Coisas possibilitou o crescimento da quantidade de dispositivos interconectados através da rede mundial de computadores. Em um ambiente tão diversificado, surgem aplicações com novas características. Como junção de muitas tecnologias e aplicações, são concebidas as cidades inteligentes. Tais cidades são capazes de interconectar pessoas e objetos, oferecendo novos meios de interação entre as pessoas e o mundo real. Com isso, busca melhorar a qualidade de vida dos indivíduos ao prestar serviços que antes seriam, em grande parte, inviáveis.

Dentro do escopo de uma cidade inteligente, existem os Prédios Inteligentes, os quais são construções físicas que suportam o uso pessoal ou de negócio. A partir de sensores e atuadores, eles podem detectar e prever o comportamento dos residentes, ajustando o ambiente de acordo com suas necessidades. Os Prédios Inteligentes propiciam diversos benefícios, tais como automação de tarefas e barateamento de custos.

Como mostrado no Capítulo 2, o aumento da quantidade de aplicações e dispositivos interconectados favorece o crescimento de ataques à segurança da informação. Por isso, é imprescindível que as organizações em Prédios Inteligentes possuam mecanismos proativos e reativos para apurar as causas de um incidente através de evidências digitais.

Para conduzir o processo investigativo digital, o responsável precisa assegurar que as evidências coletadas tenham validade legal. Senso assim, é de extrema importância seguir um processo que se proponha a produzir evidências que tenham validade legal. Para isso, existe o modelo do NIST, conforme mencionado em capítulos anteriores.

A escolha de um modelo forense depende das necessidades inerentes ao negócio. Conforme o Capítulo 3, as propostas podem dividir-se em modelos genéricos e modelos específicos. A grande vantagem de trabalhar com modelos específicos é a definição de cenários mais restritos na condução do processo forense. Por exemplo, um modelo utilizado em ambiente de Prédio Inteligente não precisa, necessariamente, tratar de requisitos presentes em processos forenses quando executados em uma rede de carros inteligentes, como em Hossain *et al.* (2017). Nesse contexto, o principal problema que motivou o desenvolvimento deste trabalhou foi a ausência de um modelo específico para investigação forense em Prédios Inteligentes.

Dessa forma, este capítulo apresenta o SBIM, um modelo de investigação forense para prédios inteligentes, incluindo a descrição de suas seis fases: *Forensic Readiness*, Capaci-

dade Forense, Coletar, Traduzir, Analisar e Examinar e Apresentar.

4.1 Introdução

Apesar do uso de um processo forense ser opcional, ele é altamente recomendável, sobretudo porque 38% dos incidentes de segurança em infraestrutura de TI têm um vetor de infecção desconhecido, ou seja, as causas dos incidentes não são conhecidas (SECURITY, 2015).

Espera-se que em Prédios Inteligentes as atividades de monitoramento sejam executadas, em grande parte, sem nenhuma interação humana (BAJRAMOVIC *et al.*, 2016). As atividades de rede e sistemas podem ser continuamente observadas a partir de registros gravados em *logs*.

Os eventos de *log* acontecem sempre que os sistemas digitais sofrem modificação. Seja se o sistema executa de forma apropriada ou se a execução encerra com erros, tais eventos são gravados em arquivos de *log* (IBRAHIM *et al.*, 2011). Em ambientes digitais, usuários também deixam seus rastros de atividades, assim como suas intenções, gravados em *logs* (SOMMER, 2012). No contexto de investigação forense em Prédios Inteligentes, os *logs* apresentam uma grande vantagem, pois eles oferecem um registro inteiro, passo a passo, de todos os eventos que aconteceram em um sistema (IBRAHIM *et al.*, 2011). Ao examinar *logs*, os investigadores podem possivelmente confirmar se um ataque à segurança da informação obteve êxito (IBRAHIM *et al.*, 2011), já que muitas aplicações geram eventos de *log* por padrão.

Depois de gerados, os *logs* precisam estar seguros e obedecer à cadeia de custódia. Isso porque muitos atacantes removem os *logs* quando invadem sistemas (PASCUCCI, 2013). Para contornar essa vulnerabilidade, um sistema gerenciador de *logs* centralizado é apropriado para manter os *logs* seguros. A cada intervalo regular, por exemplo, os sistemas e aplicações enviam seus *logs* ao servidor centralizado. Adicionalmente, os *logs* em trânsito devem estar criptografados para assegurar a sua confidencialidade (PASCUCCI, 2013). Para que os *logs* não sejam alterados, pode-se utilizar softwares bloqueadores de escrita; e para verificar sua integridade, usam-se funções *Hash* (NIST, 2006).

Nessa conjuntura, o principal objetivo deste trabalho é propor o SBIM, um modelo baseado em *logs* para investigação forense em Prédios Inteligentes. O SBIM foi concebido como um modelo específico para conduzir o processo investigativo digital em empresas e organizações públicas ou privadas situadas em Prédios Inteligentes. A partir do levantamento bibliográfico, percebeu-se a inexistência de modelos de processo que conduzam a investigação de evidências

digitais em Prédios Inteligentes, considerando, principalmente, a capacidade forense de cada organização, o gerenciamento de eventos proativo e reativo, a abrangência da execução do processo nos níveis de dispositivo, rede e nuvem, e a validação prática do modelo. O trabalho de Bajramovic *et al.* (2016) é o único relacionado a prédios inteligentes, porém possui o foco apenas no processo proativo de DFR.

Por outro lado, o SBIM abrange todas as fases proativas e reativas no contexto de investigação forense digital em Prédios Inteligentes. Esse modelo considera o monitoramento e gerenciamento de eventos da organização como um processo fundamental para os investigadores, pois é nele que incidentes e eventos são reportados ou previstos e os dados são proativamente coletados através de outro processo chamado *digital forensics readiness*. Vencida a etapa proativa, o modelo segue para suas tarefas reativas de coletar, traduzir, examinar e analisar e reportar as evidências. Todas essas fases estão de acordo com o que preconiza o documento especial 800-86 do NIST. Isso quer dizer que todas as evidências coletadas têm uma boa probabilidade de serem legalmente aceitas no tribunal, sobretudo pela preservação da cadeia de custódia. Adicionalmente às fases propostas pelo NIST, o SBIM contém três novas fases: Capacidade Forense, *Forensic Readiness* e uma fase de Tradução dos *logs* coletados.

As principais contribuições deste trabalho são:

- Incluir uma fase de tradução no modelo investigativo, a qual será constituída por uma camada de software capaz de representar, de forma padronizada, os *logs* coletados de diferentes tecnologias de comunicação, tais como *IP*, *LoRaWAN* e *Bluetooth Low Energy*. Essa camada de software busca diminuir o esforço da equipe de investigação, já que os investigadores não precisam lidar com formatos distintos de *log*, pois aprendem uma única representação;
- Uma fase de execução contínua para planejar a capacidade forense, cuja responsabilidade é preparar a equipe de investigação para a execução do processo forense;
- Uma fase de *Forensic Readiness*, com a função de centralizar, gerenciar e garantir a integridadade dos logs coletados.

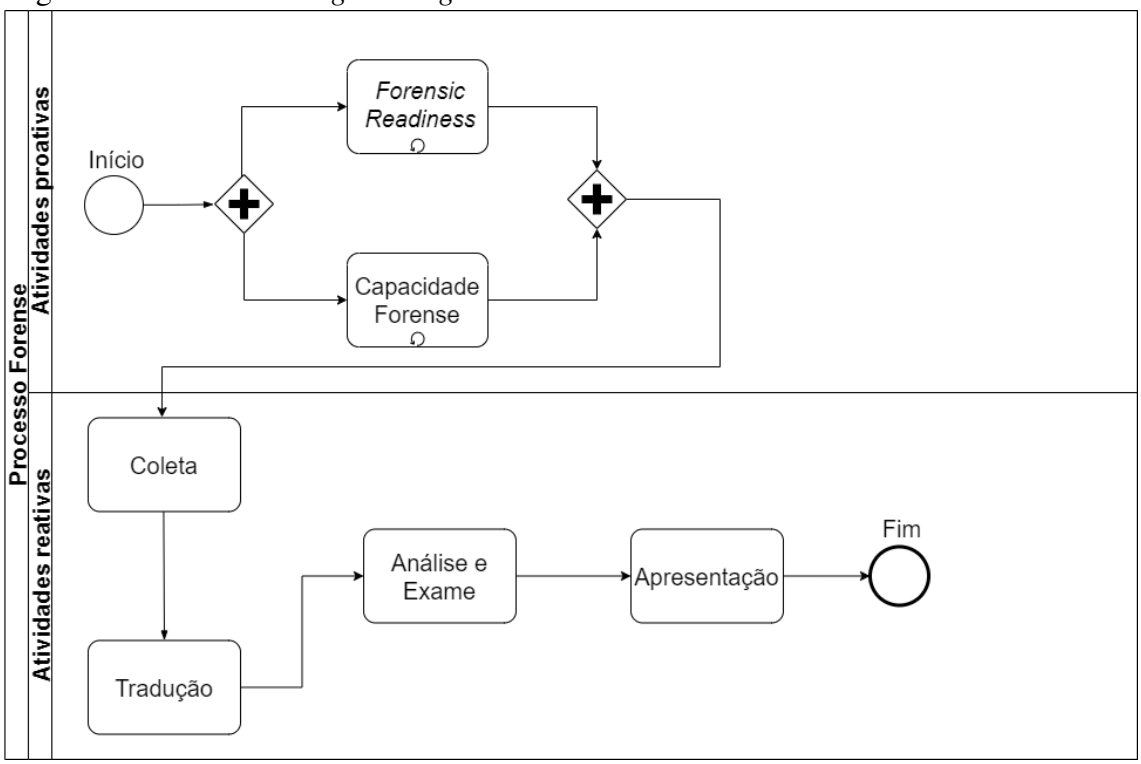
4.2 Fases do SBIM

O SBIM é um modelo de processo composto por fases que se propõem a coletar, interpretar e apresentar evidências digitais, de forma que sejam legalmente aceitas em uma corte. Por isso, o SBIM pode ser ajustado à realidade de cada organização, dentro do limite

da tecnologia disponível dentro de cada Prédio Inteligente. O modelo proposto não é taxativo, ou seja, não limita o uso de ferramentas dentro de cada fase nem receita como as organizações devem realizar cada fase. O processo foca em "o quê" e não "como" deve ser feito. Não obstante, para que as organizações possam coletar evidências com menor custo (*Forensics Readiness*), com pessoal bem capacitado (Capacidade Forense) e que as evidências tenham validade judicial, é recomendado que todas as fases do processo sejam executadas.

Para modelar o processo investigativo forense digital, foi utilizada a notação BPMN [1], na sua versão 2.0. A BPMN apresenta um conjunto robusto de símbolos para modelagem de diferentes aspectos de processos de negócio. Seus símbolos descrevem relacionamentos claramente definidos, tais como fluxo de atividades e ordem de precedência.

Figura 4 – *Smart Building Investigation Model*



Fonte: Elaborada pelo autor.

A Figura 4 representa a modelagem do SBIM em duas raias do BPMN, que separam as fases ou atividades reativas das proativas. As descrições de cada uma das seis fases ou atividades estão dispostas a seguir.

[1] https://www.bpmn.org/

4.2.1 Forensic Readiness

Forensic Readiness é um termo que se refere à capacidade proativa e pré-incidente de coletar e preservar os *logs* para que sejam reutilizados nas fases reativas ou pós-incidente. No SBIM, o DFR se dá por meio do armazenamento centralizado dos *logs* gerados pelas aplicações e sistemas presentes no Prédio Inteligente. Assim, o principal objetivo desta fase é coletar e preservar os *logs* de dispositivos em um servidor de *logs* centralizado. Além disso, é nesta fase que sistemas para representação padrão de *logs* são desenvolvidos e mantidos pela equipe de desenvolvimento. Esses softwares são essenciais para a fase de tradução do SBIM.

A ideia principal do armazenamento centralizado é ter um único local, como um servidor, que é dedicado a receber e armazenar *logs* de várias fontes, de modo que todos os dados estejam disponíveis em um único local (VEGA *et al.*, 2017; KAVIS, 2014). Em uma rede de vários servidores, é mais fácil acessar os *logs* quando eles estão em um único local em vez de separados em várias máquinas, que é a principal motivação para o armazenamento centralizado. Dessa forma, os *logs* também não seriam perdidos se ocorresse qualquer problema com os dispositivos que geram os *logs*.

Uma solução de armazenamento centralizado de *logs* consiste principalmente em múltiplos geradores de *logs* e um sistema coletor *logs* (OLINER *et al.*, 2012). Os geradores de *logs* incluem todas os dispositivos que estão gerando os dados de *log* que devem estar acessíveis em um único local. O coletor de *log* pode ser um ou vários servidores que recebem os *logs*. Vários coletores de *log* podem ser usados para suportar um volume maior de dados. Outras razões para ter vários coletores de *log* incluem a tolerância a falhas e a alta disponibilidade em caso de falhas de algum deles (VEGA *et al.*, 2017). Ademais, usar apenas um único coletor de *log* introduziria um único ponto de falha, algo que pode ser inaceitável.

Geralmente, os dados de *log* podem ser considerados estruturados (quando existe uma relação semântica entre eles) ou não estruturados. É possível armazenar os dados assim como foram recebidos, porém existem formas mais adequadas de armazenar os dados se eles estiverem em um formato bem estruturado (JAYATHILAKE, 2012). Por exemplo, um arquivo de *log* bem estruturado conteria uma única entrada de *log* em cada linha e cada entrada conteria a data, hora e um nível de gravidade da mensagem, além da própria mensagem. Esse tipo estruturado de *log* pode ser facilmente armazenado em um banco de dados SQL com colunas separadas para registro de data e hora, gravidade e mensagem.

Para dados semiestruturados, um banco de dados NoSQL pode funcionar melhor do

que armazenar os dados de *log* como estão. Isso pode facilitar a pesquisa de dados de *log* de um determinado intervalo de tempo ou *logs* com certa gravidade, por exemplo. No armazenamento de *logs*, também existem certas características que devem ser levadas em consideração ao se planejar um sistema de *log* centralizado, tais como a persistência, capacidade de armazenamento e segurança (OLINER *et al.*, 2012). Deve ser decidido por quanto tempo os dados devem ser mantidos de forma persistente, pois isso afetará a capacidade total necessária junto com a velocidade na qual os *logs* são gerados. Se o custo de armazenamento for um fator limitante, ele determinará a quantidade de armazenamento disponível e também por quanto tempo os *logs* antigos podem ser mantidos. Vale ressaltar que esse tempo pode ser estendido compactando os *logs* mais antigos ou mantendo apenas alguns dados mais importantes, em vez de armazenar tudo. Outro ponto não menos importante é que o envio de *logs* ao coletor de *log* será limitado pela largura de banda e latência da conexão de rede. Se a quantidade de *logs* gravados for capaz de esgotar a largura de banda, o coletor de *log* pode ser movido para mais perto dos geradores de *log* na rede ou o sistema do coletor de *log* pode ser escalado horizontalmente com novas máquinas, se isso for permitido pelo sistema existente.

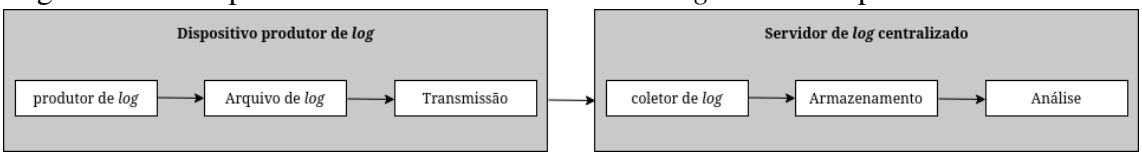
O método de transmissão dos *logs* ao coletor também tem certos aspectos a serem considerados, tais como compactação, criptografia, autenticação e integridade (TRENWITH; VENTER, 2013). Quando os *logs* são transmitidos pela rede, pode ser uma boa ideia compactar os dados antes de enviá-los para economizar largura de banda. Como os *logs* geralmente são texto simples, eles devem ser compactados de maneira muito eficaz. A compactação e a descompressão introduzem alguma carga adicional na CPU e podem causar um pequeno atraso antes que os dados de registro se tornem disponíveis no servidor centralizado. Portanto, se todo o sistema de *logs* estiver dentro de uma rede local em que a largura de banda não seja um problema, pode ser que não seja necessário utilizar a compressão.

Já a criptografia e a autenticação são geralmente importantes na comunicação via rede e provavelmente também devem ser implementadas para a transmissão dos *logs*, o que seria especialmente importante se eles contiverem dados confidenciais. A criptografia é necessária para que terceiros não possam ler os *logs* transmitidos pela rede. A autenticação pode ser usada para assegurar a identidade do coletor de *log* e, dessa forma, impedir a coleta por um agente malicioso. Da mesma forma, a autenticação pode impedir que os dados enviados ao coletor sejam de aplicações não autorizadas. Assim como acontece com a compactação, pode ser possível deixar a criptografia e a autenticação de fora para economizar recursos da CPU e problemas de

configuração se todo o sistema estiver dentro de uma rede local devidamente protegida, embora camadas adicionais de segurança sejam sempre uma boa ideia. Por último, a integridade dos *logs* transmitidos também pode ser uma preocupação. Se o TCP estiver sendo usado para transmissão, ele já trata da integridade.

A Figura 21 apresenta um exemplo de configuração para um sistema centralizado de *log* e seus diferentes componentes, incluindo a parte de análise de *logs*. O sistema pode consistir em vários dispositivos produtores de *log*, todos executando o software que produz os dados de *log*, e o sistema de transmissão, o qual envia os *logs* para os servidores centralizados. Os componentes do servidor centralizado podem estar em execução em máquinas diferentes, pois não é necessário que o componente coletor *logs*, o de armazenamento e a análise estejam em execução na mesma máquina.

Figura 5 – Exemplo de um sistema centralizado de *log* e seus componentes



Fonte: adaptado de Vainio (2018).

Existem várias soluções de *software* no mercado que implementam o armazenamento centralizado de *logs*. A primeira delas, possivelmente uma das maneiras mais simples de obter um serviço centralizado, seria configurar uma tarefa no *cron* (IEEE; GROUP, 2018), usando algum utilitário para a transmissão dos arquivos, tais como o SCP (RINNE; YLöNEN, 2013) ou o *rsync* (TRIDGELL, 2018), para copiar periodicamente os arquivos de *log* para a máquina servidora centralizada. A configuração desse tipo de sistema na máquina produtora de *log* pode ser feita facilmente adicionando uma linha ao arquivo *crontab*, como a seguir:

Figura 6 – Configuração do arquivo *Crontab*

```
0 * * * * rsync -r /home/ubuntu/diretoriolog ubuntu@nome_servidor_centralizado:~
```

Fonte: Elaborado pelo autor.

Esse exemplo copia recursivamente, a cada hora, tudo que está dentro do diretório `diretoriolog`. Essa abordagem é rápida e não necessita de instalação de programas adicionais. Entretanto, ela apresenta algumas limitações. Provavelmente a maior delas é que as ferramentas utilizadas (serviço *crontab* e *rsync*) podem não estar disponíveis para todos os dispositivos disponíveis no Prédio Inteligente. A segunda limitação, seguindo o exemplo, seria o envio

de dados feito somente a cada hora, o que limita atuações de coleta em tempo menor que esse. Se o dispositivo for desligado antes de atingir a hora configurada, os dados são perdidos imediatamente, antes de serem enviados ao servidor. Finalmente, se os dispositivos enviarem *logs* de mesmo nome, eles poderão ser sobrescritos no servidor. Apesar da limitações, as ferramentas rsync e scp possuem mecanismos de segurança e autenticação.

A segunda solução, *syslog daemons*: *Syslog* (GERHARDS, 2009), tornou-se o sistema de *log* padrão em sistemas Unix, porém não está disponível no *Windows* por padrão. Ele define um formato de mensagem padrão que faz com que o software grave *logs* em um formato padrão, facilitando as ferramentas de análise de *logs*.

Para escrever os *logs* de acordo com o *syslog*, geralmente isso deve ser feito explicitamente em alguma linguagem de programação. Existem implementações em muitas linguagens, tais como C e *Python*. Exemplos em C e *Python* podem ser executados desta forma:

Figura 7 – Exemplo em C

```c
#include <syslog.h>
syslog(LOG_INFO, "log message");
```

Fonte: Elaborado pelo autor.

Figura 8 – Exemplo em *Python*

```python
import syslog
syslog.syslog(syslog.LOG_INFO, "log message");
```

Fonte: Elaborado pelo autor.

Um *daemon syslog* que está sendo executado na máquina local lida com o recebimento de mensagens de *log* e o armazenamento delas, geralmente em arquivos. Também é possível rotear essas mensagens para outros *daemons syslog* em execução noutros hosts e, dessa forma, permitir a transmissão de *logs* para um local central.

O *daemon syslog* original é bastante limitado, entretanto. Uma vez que ele só aceita mensagens de *sockets* Unix, não pode rotear com base em expressões regulares e pode não estar disponível para todas as plataformas em um ambiente inteligente. Portanto, foram criados alguns *daemons* alternativos para o *syslog* que podem substituir o *daemon* original, como o Syslog-ng (LLC, 2018) e o Rsyslog (GERHARDS, 2017).

Uma outra forma seria utilizar os projetos de código aberto *Filebeat* (BV, 2018b) ou *Logstash* (BV, 2018c), mantidos pela empresa *Elasticsearch BV*. Além dessas duas soluções,

a empresa também fornece o software de banco de dados *Elasticsearch* (BV, 2018a) que pode ser usado para armazenar os dados de *log*. O *Filebeat* e *Logstash* não fornecem nenhum armazenamento de *log*, concentrando-se apenas na transmissão deles. Eles buscam manter um armazenamento mais estruturado e distribuído. O banco de dados *Elasticsearch* é bem suportado no *Filebeat* e *Logstash*, já que é da mesma empresa. O Logstash é frequentemente citado na literatura (HE *et al.*, 2016; VEGA *et al.*, 2017) e é ocasionalmente usado para soluções em IoT, inclusive em propostas para prédios inteligentes (BAJER, 2017; DHARUR; SWAMINATHAN, 2018).

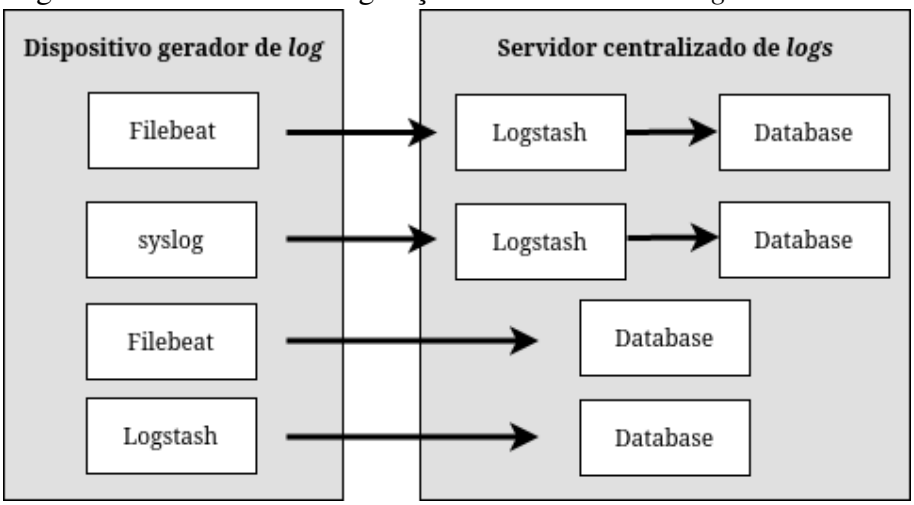
O *Filebeat* e o *Logstash* são muito semelhantes entre si, pois podem ler *logs* de várias fontes diferentes, filtrar e aprimorar os dados de *log* e enviar os dados para destinos específicos. A diferença geral é que o *Filebeat* é muito mais leve e o *Logstash* oferece suporte a mais opções e até tem *plug-ins* para diferentes filtros. A configuração sugerida para um sistema de *logs* centralizado usando essas soluções é instalar o *Filebeat* para atuar como um encaminhador de *logs* em cada dispositivo. Os clientes *Filebeat*, então, enviam os dados de *log* para um ou mais servidores *Logstash* que, por sua vez, processam os *logs* antes de enviá-los ao armazenamento *Elasticsearch*.

Existem também outras maneiras de usar essas soluções, dependendo do caso de uso e dos requisitos: seria possível usar o *syslog* no lugar do *Filebeat* para enviar os *logs* dos clientes para o servidor *Logstash*. Também é possível fazer menos processamento de *logs* removendo a camada do servidor *Logstash* completamente e enviando os *logs* diretamente para o armazenamento *Elasticsearch*. Além disso, os clientes poderiam encaminhar os *logs* usando o *Logstash* em vez do *Filebeat*, o que incluiria maior *overhead* de processamento no lado cliente. Por último, é possível substituir o armazenamento de *log* do *Elasticsearch* por qualquer outra solução de banco de dados, especialmente ao usar o *Logstash*, pois ele pode suportar vários outros sistemas com os seus *plug-ins*. A Figura 9 mostra as opções de configuração de um sistema centralizado de *logs* usando as tecnologias apresentadas.

O *Filebeat* e o *Logstash* também suportam compactação, criptografia, autenticação e integridade.

Já as ferramentas *Fluent Bit* (DATA, 2018) e o *Fluentd* (FOUNDATION, 2018) também são de código aberto, mas mantidas pela *Treasure Data*. *Fluentd* é um projeto hospedado na *Cloud Native Computing Foundation*. O *Fluent Bit* e o *Fluentd* são muito semelhantes ao *Filebeat* e *Logstash*, respectivamente, constituindo, essencialmente, um sistema de transmissão

Figura 9 – Diferentes configurações com *Filebeat* e *Logstash*



Fonte: adaptado de Vainio (2018).

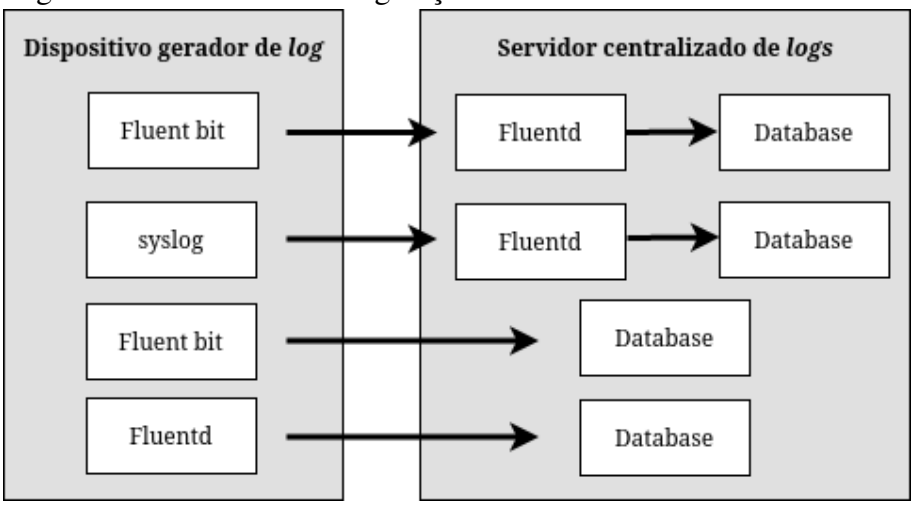
de *logs* que requer algum sistema de banco de dados para armazenamento dos dados, a menos que a saída padrão esteja configurada para salvar os *logs* como arquivos.

A diferença entre o *Fluent Bit* e o *Fluentd* é que o *Fluent Bit* é muito mais leve, mas só oferece suporte a certas entradas e saídas de *log*, enquanto os recursos do *Fluentd* podem ser estendidos com *plug-ins* para oferecer suporte a diferentes fontes de *log* e sistemas de banco de dados adicionais. O *Fluentd* também tem melhores recursos de filtragem e processamento de *logs* em comparação ao *Fluent Bit*.

Assim como no caso do *Filebeat* e *Logstash*, existem várias maneiras diferentes de configurar um sistema de *log* centralizado usando o *Fluent Bit* e o *Fluentd*. A maneira recomendada é usar o *Fluent Bit* para coletar os *logs* nos dispositivos geradores de *logs* e enviá-los às máquinas servidoras que executam o *Fluentd* que, por sua vez, processam os dados e os enviam para algum sistema de armazenamento de *log* (VAINIO, 2018). Todos os métodos alternativos já listados para *Filebeat* e *Logstash* também funcionam com *Fluent Bit* e *Fluentd*. Isso significa que o *Fluent Bit* nas máquinas geradoras de *log* pode ser substituído por uma solução *syslog*, a camada do servidor com *Fluentd* pode ser removida, pois o *Fluent Bit* pode gravar diretamente no bancos de dados. Além disso, o *Fluentd* pode ser usado no lugar do *Fluent Bit* para coletar os *logs*. Como o *Fluentd* usa menos memória que o *Logstash* (PERI, 2015), usar apenas o *Fluentd* em todos os dispositivos para ser uma opção viável. *Fluent Bit* e *Fluentd* também podem suportar compressão, criptografia, autenticação e integridade. A Figura 10 ilustra as variadas configurações.

Finalmente, uma versão gratuita da solução *Splunk* (INC, 2018; JAYATHILAKE, 2012) também está disponível, mas em comparação com outros sistemas centralizados de *logs*

Figura 10 – Diferentes configurações com *Fluent bit* e *Fluentd*



Fonte: adaptado de Vainio (2018).

gratuitos, é muito mais limitado. Ao contrário das soluções apresentadas, os produtos da *Splunk* não são de código aberto, sendo principalmente soluções comerciais.

O *Splunk* é frequentemente mencionado quando se fala sobre soluções de *log* em geral (VEGA *et al.*, 2017; OLINER *et al.*, 2012; JAYATHILAKE, 2012; HE *et al.*, 2016), portanto, dentro das opções comerciais, ele é o que geralmente recebe mais atenção. O *Splunk* também é ocasionalmente usado para outras soluções, incluindo abordagens em IoT (CHEN; CHIEN, 2017).

Para facilitar o trabalho dos investigadores, é recomendável que os *logs* sejam organizados nos coletores de forma que seja possível identificar a propriedade dos *logs*, ou seja qual dispositivo foi responsável pela sua transmissão ao armazenamento central. Além disso, é importante que diferentes aplicações tenham seus *logs* organizados de maneira separada.

Para isso, o administrador do coletor central pode criar diretórios para cada um dos dispositivos que necessitem ter os *logs* coletados, ou os diretórios podem ser criados automaticamente no coletor quando os *logs* forem transmitidos. A ferramenta *rsync* consegue solucionar esse problema com o comando abaixo.

Figura 11 – Ferramenta *rsync*

```
$ rsync -a --relative /var/logs/./dhcp/ user@remote:/deviceid/
```

Fonte: Elaborado pelo autor.

Outra forma seria, antes de executar o comando *rsync*, sempre conectar-se ao coletor via *SSH* para criar o diretório. Os seguintes comandos servem de exemplo:

Os comandos apresentados têm o mesmo objetivo: enviar o *log* da aplicação *dhcp*

Figura 12 – Ferramenta ssh

```
$ ssh user@server mkdir -p deviceid/dhcp
```

Fonte: Elaborado pelo autor.

Figura 13 – rsync e dhcp

```
$ rsync -a ~/var/logs/dhcp user@remote_host:deviceid/dhcp
```

Fonte: Elaborado pelo autor.

ao diretório *deviceid/dhcp* no coletor central. Nesse caso, *deviceid* representa uma identificador único que será atribuído a cada dispositivo produtor de *log* no prédio inteligente.

Vale ressaltar que para usar a ferramenta *SSH* para conexões remotas, as devidas permissões devem ser concedidas aos dispositivos que queiram conectar-se ao servidor.

4.2.2 Capacidade Forense

Esta atividade está de acordo com as recomendações do NIST (2006). Seu principal objetivo é estabelecer e capacitar a esquipe responsável por conduzir o processo reativo de investigação.

A importância dessa atividade está no fato de ela preparar a organização para executar o processo investigativo reativo nas aplicações e na rede. Após um incidente, por exemplo, os *logs* devem ser coletadas e analisados de forma que a equipe não comprometa a integridade das evidências, assegurando, assim, sua validade judicial. Além disso, a equipe deve estar ciente de que nem todas as informações devem ser analisadas, por existir variados graus de confidencialidade. Um e-mail, por exemplo, não deve ser analisado sem as devidas autorizações. Dessa forma, a atividade de capacidade forense deve ensinar o que se pode fazer e o que não se pode fazer durante todo o processo investigativo. Existem grupos específicos na organização que devem ser treinados para desenvolver a capacidade forense desde a etapa pré-incidente. Segundo o NIST (2006), esses grupos dividem-se em:

- **Investigadores**: são os responsáveis por investigar alegações de má condutas. Imediatamente assumem a investigação de qualquer evento que seja suspeito de envolver atividade criminal. Fazem uso de muitas ferramentas forenses;
- **Profissionais de Tecnologia da Informação (TI)**: esse grupo inclui administradores de sistema, rede e segurança da organização. Eles usam um número menor de ferramentas forenses;

– **Pessoal que lida com incidentes**: responde por incidentes de segurança da informação, tais como acesso não autorizado, infecção de código malicioso e ataques de negação de serviço. Tipicamente, utilizam uma grande variedade de técnicas e ferramentas forenses.

Muitas organizações contam com uma combinação de sua equipe interna com partes externas para executar tarefas forenses. Por exemplo, algumas organizações executam elas mesmas tarefas mais simples e usam terceiros apenas quando é necessário realizar assistência especializada (NIST, 2006). Mesmo as organizações que desejam realizar todas as tarefas forenses sozinhas geralmente terceirizam as mais especializadas, como, por exemplo, o envio de uma mídia física danificada para uma empresa de recuperação de dados. Essas tarefas normalmente requerem o uso de software especializado, equipamentos, instalações e conhecimentos técnicos que para a maioria das organizações representa um alto custo (NIST, 2006).

Quando da decisão de terceirizar as atividades forenses, a organização deve levar em conta o seguinte (NIST, 2006):

– **Custo**: Software, hardware e equipamentos usados para coletar e examinar dados podem acarretar custos significativos (por exemplo, preço de compra, atualizações e upgrades de software, manutenção) e também podem exigir medidas de segurança adicionais. Outras despesas envolvem treinamento de pessoal e custos de mão de obra, que são particularmente significativas para especialistas forenses. Em geral, as ações forenses que raramente são necessárias podem ser executadas de forma mais econômica por uma parte externa, ao passo que as ações que são necessárias com frequência podem ser executadas internamente.
– **Tempo de resposta**: O pessoal localizado na organização pode iniciar a atividade forense mais rapidamente do que o pessoal externo. Para organizações com filiais geograficamente dispersas, por exemplo, agentes externos localizados perto de filiais distantes podem responder mais rápido do que o pessoal localizado na sede da organização.
– **Dados sensíveis**: a empresa pode relutar ao conceder acesso de seus dados para que terceiros executem o processo forense. Por outro lado, se o incidente envolver alguém do pessoal do processo de incidentes da organização, por exemplo, é mais interessante que o processo seja executado por uma entidade externa.

As pessoas que lidam com incidente precisam conhecer os princípios forenses, guias, procedimentos, ferramentas e técnicas, assim como ferramentas antiforense que dificultam o processo investigativo. Também é interessante que as pessoas que lidam com eventos, tais como os incidentes, conheçam os sistemas operacionais mais comuns, como também sistemas de

arquivos, aplicações e protocolos de rede dentro da organização. Essas pessoas devem possuir um conhecimento abrangente dos sistemas presentes na empresa, já que isso pode acelerar a ação de resposta a incidentes (NIST, 2006). Por terem vasta experiência em incidentes de segurança da informação, o pessoal do gerenciamento de eventos e de incidentes pode ministrar cursos aos administradores de rede e de sistemas da organização (NIST, 2006).

Os grupos que executam o processo forense devem passar por treinamentos teóricos e práticos em laboratórios planejados pela empresa. Cada membro deve ser capaz de executar o processo inteiramente, pois na falta de alguém na equipe, a organização ainda será capaz de promover uma investigação forense. O pessoal deve manter-se ativo e atualizado com as novas ferramentas de investigação forense digital (NIST, 2006).

Como não é viável que cada membro do time domine todas as tecnologias presentes na organização, é extremamente recomendável que haja uma política de interação entre os diferentes times. Em um possível incidente numa base dados, por exemplo, os administradores de banco de dados devem estar acessíveis para facilitar o trabalho dos investigadores. Para facilitar a comunicação entre os times, cada time deve estabelecer um ou mais pontos de contato, que será a porta de entrada para a troca de informação entre eles. A organização deve manter uma lista de contatos, como e-mail e telefones, a qual tenha contatos padrões e contatos para emergência (NIST, 2006).

Algumas atividades podem ser desenvolvidas para mensurar o quanto a equipe está capacitada para executar o processo forense. As avaliações podem ser individuais ou em equipe e incluem ferramentas como pesquisas sobre atitudes dos membros da equipe em determinada situação. Essa é uma oportunidade para determinar se procedimentos foram seguidos; avaliações específicas, tais como avaliações sobre determinada política de segurança ou tecnologia da empresa; entrevistas estruturadas, nas quais os entrevistadores podem avaliar a postura e o condicionamento emocional de cada membro; testes de habilidades, os quais podem ser executados em laboratórios usando ferramentas específicas em simulações de um cenário real; e grupos de discussão, que são um oportunidade de avaliação e troca de ideias com o propósito de se ter um olhar introspectivo na equipe e determinar o que os membros podem aprender com as experiências passadas, o que deve ser melhorado e o que deve ser mantido em investigações futuras. Essas ferramentas podem melhorar a compreensão, a confiança, o compromisso e as comunicações entre os membros da equipe, e fazer com que as equipes tornem-se mais produtivas (PMI, 2013).

Finalmente, NIST (2006) traz algumas recomendações importantes para estabelecer-se a capacidade forense de uma organização:

- **A organização deve saber executar ferramentas forenses em computadores e na rede**. *O processo investigativo é necessário para várias tarefas dentro de uma organização, incluindo investigação de crimes e comportamentos inadequados, reconstrução de incidentes de segurança, solução de problemas operacionais, análise de logs de auditoria e recuperação de danos acidentais. Sem esse recurso, uma organização terá dificuldade em determinar quais eventos ocorreram em seus sistemas e redes, como exposições de dados confidenciais protegidos. Além disso, lidar com as evidências da maneira adequada coloca os tomadores de decisão em uma posição em que podem tomar ações que sejam legalmente aceitas.*
- **A organização deve determinar quais equipes executam quais atividades do processo**. *A maioria das organizações conta com uma combinação da equipe interna com partes externas para executar tarefas forenses. As organizações devem decidir quais partes devem cuidar de quais tarefas com base em habilidades e capacidades, custo, tempo de resposta e confidencialidade de dados.*
- **Os times de incidentes devem ser capazes de executar o processo forense**. *Mais de um membro da equipe deve ser capaz de realizar cada etapa do processo forense. Exercícios práticos e cursos de treinamento em forense podem ser úteis na construção e manutenção de habilidades, assim como demonstrações de novas ferramentas e tecnologias.*
- **A forense deve ser compartilhada por vários times na organização**. *O pessoal que executa as ações forenses deve ser capaz de contatar outras equipes e indivíduos dentro da organização, sempre que necessário, para obter assistência especializada. Exemplos de equipes que podem fornecer assistência incluem os profissionais de TI, gerenciamento, consultores jurídicos, pessoal de recursos humanos, auditores e equipe de segurança. Os membros dessas equipes devem compreender suas funções e responsabilidades em perícia, receber treinamento e educação sobre políticas, diretrizes e procedimentos forenses e estar preparados para cooperar e auxiliar outras pessoas nas ações forenses.*
- **Todo o processo forense deve estar contido na política da organização**. *As políticas devem permitir que o pessoal autorizado monitore sistemas e redes e execute investigações por motivos legítimos em circunstâncias apropriadas. Todos devem compreender a política forense. As considerações adicionais são as seguintes: a política forense deve definir*

claramente as funções e responsabilidades de todas as pessoas que executam ou auxiliam nas atividades forenses da organização. A política deve incluir todas as partes internas e externas; as políticas, diretrizes e procedimentos da organização devem explicar claramente quais ações forenses devem e não devem ser executadas em circunstâncias normais e especiais e devem abordar o uso de ferramentas e técnicas anti-forenses; incorporar considerações forenses no ciclo de vida do sistema de informação pode levar a um tratamento mais eficiente e eficaz em muitos incidentes. Os exemplos incluem a realização de auditorias em hosts e rede.

– **A organização deve manter diretrizes para auxiliar no processo forense**. *As diretrizes devem incluir metodologias gerais para investigar um incidente usando técnicas forenses. As diretrizes e procedimentos devem apoiar a admissibilidade das provas nos processos judiciais. Como os registros eletrônicos e outros registros podem ser alterados ou manipulados, as organizações devem estar preparadas, por meio de suas políticas, diretrizes e procedimentos, para demonstrar a confiabilidade e integridade de tais registros. As diretrizes e procedimentos também devem ser revisados e mantidos regularmente para que sejam precisos.*

4.2.3 Coletar os dados

Após o armazenamento proativo dos *logs* ocorrer, o Prédio Inteligente executará atividades reativas à medida que eventos significantes para a organização ocorrerem. Como mencionado anteriormente, eventos como tentativas de acesso, ataques à rede e a computadores e data e hora de acesso a sistemas podem ser registrados em *logs* de rede e sistemas mantidos pela organização. Essas informações são fundamentais para o processo investigativo, pois possibilitam a validação de hipóteses que estabeleçam uma ordem de acesso aos sistemas e qual dispositivo estava ativo em determinada data e hora. O cruzamento das informações em *logs* pode revelar evidências importantes no processo de investigação forense digital.

A atividade de coleta reativa dos dados é composta pela identificação de onde os dados estão armazenados e pela aquisição desses dados (NIST, 2006).

4.2.3.1 Identificar as fontes dos dados

A atividade de coleta primeiro preocupa-se com identificar onde os dados estão armazenados. Tratando-se do SBIM, os dados sempre estarão armazenados no coletor de *logs*

centralizado. Isso é um ponto positivo do modelo proposto, pois os profissionais responsáveis pelo processo forense terão um escopo bem definido para iniciar o processo investigativo, mesmo que haja mais de um coletor na organização, independentemente se estejam ou não na rede local.

4.2.3.2 Aquisição dos dados

Depois de identificar onde os dados estão armazenados, os analistas devem extrair os dados de lá. A aquisição dos dados é executada seguindo três passos (NIST, 2006): desenvolver um plano para a aquisição dos dados, extrair os dados e verificar a integridade dos dados.

- **Desenvolver um plano de aquisição de dados**. Desenvolver um plano é um importante primeiro passo na maioria dos casos. Os analistas devem criar um plano que determine o objetivo da investigação, estabelecendo uma ordem de importância que priorize a aquisição de alguns dados em relação a outros. Alguns fatores importantes para a priorização incluem: *o valor do dado*. Baseado no entendimento dos analistas e experiências anteriores, eles podem estimar o valor de cada evidência, dizendo qual é mais importante que a outra; *volatilidade*. Dados voláteis são aqueles que são perdidos quando o sistema é desligado. Portanto, podem ter maior prioridade no processo de aquisição; *quantidade de esforço requerido*. Considera o esforço para a aquisição dos dados no coletor de *logs*. O esforço pode variar, já que mais de um coletor pode existir no Prédio Inteligente ou fora dele.
- **Adquirir os dados**. o processo de aquisição de dados envolve o uso de ferramentas para coletar dados voláteis e criar cópias de fontes de dados não voláteis. A aquisição de dados pode ser realizada localmente ou em uma rede. Embora seja geralmente preferível adquirir dados localmente porque há maior controle sobre o sistema e os dados, a coleta de dados local nem sempre é viável. O NIST (2006) recomenda que, antes de se obter qualquer evidência, seja criada uma imagem da mídia na qual as potenciais evidências estejam armazenadas. Essa prática facilita o processo de verificação da integridade dos dados que serão analisados, pois ao copiar arquivos, como os *logs*, enquanto o sistema estiver funcionando, pode ocorrer novas escritas nos *logs* com a ocorrência de novos eventos. Isso faz com os dados copiados estejam diferentes dos dados do sistema. Ao criar uma imagem, os analistas têm uma fotografia do sistema em um determinado ponto no tempo. Uma imagem nada mais é do que uma cópia da mídia original, incluindo as partições e sistema de arquivos (NIST, 2006). Para criar a imagem de um disco inteiro no ambiente Linux, pode-se usar a ferramenta *dd* (SPECIFICATIONS, 2018) da seguinte forma:

Figura 14 – Ferramenta dd

```
# dd if=/dev/sda conv=sync,noerror bs=64K | gzip -c  > /caminho/backup.img.gz
```

Fonte: Elaborado pelo autor.

No comando acima, */dev/sda* representa o dispositivo para o qual se deseja criar a imagem. *sync* e *noerror* são parâmetros que fazem com que o *dd* continue sua operação mesmo que ocorram erros de leitura. E *bs=64* define o tamanho dos blocos de leitura no processo de cópia do disco. A saída do comando *dd* será direcionada para o programa *gzip* através do comando *pipe* (|). O *gzip* comprime os dados em um arquivo chamado *backup.img.gz*.

Já para restaurar o sistema, pode-se utilizar o seguinte comando:

Figura 15 – Ferramenta *gunzip*

```
# gunzip -c /caminho/backup.img.gz | dd of=/dev/sda
```

Fonte: Elaborado pelo autor.

Aqui, *of=/dev/sda* representa onde o sistema deverá ser recuperado.

Existem outras ferramentas que alcançam o mesmo propósito, tais como a *sfdisk* (LINUX.DE, 2018a) e a *sgdisk* (LINUX.DE, 2018b), também disponíveis em distribuições Linux. Em resposta ao aumento do número de ferramentas disponíveis que se propõem a fazer imagens de dispositivos, O NIST criou um projeto chamado *Computer Forensics Tool Testing* (CFTT), o qual desenvolve testes para validar ferramentas forenses. O projeto procura validar as ferramentas, principalmente, verificando quatro requisitos: a ferramenta cria uma imagem de disco ou partição; a ferramenta não altera o disco original; a ferramenta deve registar a ocorrência de erros de E/S; a ferramenta deve estar bem documentada. Apesar de o *dd* estar entre as ferramentas que passaram pelo teste, ainda há poucas delas que foram testadas, conforme em NIST (2017).

Vale ressaltar que todo o processo de criação de imagem e montagem deve ser documentado para que o processo possa ser replicado a qualquer momento, demonstrando, assim, que as evidências não sofreram modificação a partir da constituição da imagem.

– **Verificar a integridade dos dados**. Depois que a imagem é criada, é importante verificar se os dados copiados são uma duplicata exata dos dados originais. O cálculo do resumo da mensagem dos dados copiados pode ser usado para garantir a integridade dos dados. Um resumo da mensagem é um código *hash* que identifica os dados de forma exclusiva e tem

a propriedade de que caso altere-se um único bit nos dados, faz-se com que um resumo da mensagem completamente diferente seja gerado (NIST, 2006). Existem muitos algoritmos para calcular o resumo da mensagem de dados, mas os dois mais comumente usados são o MD5 e os da família *Secure Hash Algorithm* (SHA) (NIST, 2006). Esses algoritmos tomam como entrada dados de comprimento arbitrário e produzem resumos de mensagens de tamanho fixo. Como o SHA é um algoritmo aprovado pelo *Federal Information Processing Standards* (FIPS) e o MD5 não, as organizações devem dar preferência àquele. No Linux, pode-se utilizar a ferramenta *sha512sum* para produzir um resumo de mensagem. Por exemplo:

Figura 16 – Ferramenta *SHA512sum*

```
$ sha512sum backup.img.gz
```

Fonte: Elaborado pelo autor.

A ferramenta pode ter como entrada arquivos de qualquer tamanho e produz um código *hash* de 512 *bits*.

Ademais, deve-se gerar o resumo da mensagem da mídia original antes de iniciar o processo de geração da imagem. Após a imagem ser gerada, o código *hash* da cópia deve ser calculado e comparado com o *hash* da mídia original. Os dois devem ser iguais. O resumo da mensagem da mídia original deve então ser computado novamente para verificar se o processo de imagem não alterou a mídia original. Todos os resultados devem ser documentados (NIST, 2006).

4.2.4 Tradução

A fase de tradução consiste em propor uma representação padrão dos *logs* que serão analisados. Esta fase será implementada por *software* e o seu principal objetivo é buscar diminuir o esforço da equipe que conduz a investigação, uma vez que os analista terão que aprender somente um único formato de *log* para cada tipo de *log*. Nesse contexto, os *logs* podem ser de muitos tipos, a depender de cada aplicação usada na rede.

Em uma rede de computadores, por exemplo, é comum haver um conjunto de protocolos implementados em sistemas que registram suas ações em arquivos de *log*. Dado o número de soluções disponíveis para alguns desses protocolos, é possível que existam *logs* com formatos diversos, com sintaxe e semântica à escolha dos seus desenvolvedores. Em sistemas

Linux, somente para o protocolo DHCP, existem diferentes aplicações que não seguem um padrão ao gravar os dados em seus *logs*, ou seja, para representar um mesmo evento, as aplicações utilizam mensagens diferentes ao escrever seus *logs*. Para exemplificar, no Linux, algumas das aplicações clientes DHCP são o dhcpd e o isc dhclient, além de gerenciadores de rede que possuem seus próprios clientes DHCP, tais como o ConnMan, netcli, NetworkManager, systemd-networkd e o Wicd. Já para servidores DHCP, existem o dhcpd, o dnsmasq e o Kea. Como não existe um padrão de arquivo de *log* DHCP (NIST, 2014), os desenvolvedores formatam seus arquivos à maneira que desejar.

Em um Prédio Inteligente, a capacidade de representar os *logs* de uma maneira padronizada torna-se ainda mais importante, pois nesse ambiente podem existir diferentes tecnologias para endereçar e interconectar sensores e atuadores. Dentre essas tecnologias incluem-se o NB-IoT (GSMA, 2017b), LTE-M (GSMA, 2017a), LoRaWAN (ALLIANCE, 2017), Buetooth Low Energy (PROPRIETARY, 2021a), SigFox (INC., 2010) e Symphony Link (LABS, 2020). Essas tecnologias possuem a capacidade de interconectar centenas de dispositivos e transmitir dados com baixo consumo de energia, características importantes no contexto da IoT.

Dessa forma, a fase de tradução irá viabilizar uma representação única, por exemplo, de todos os *logs* de aplicações responsáveis pela distribuição automática de endereçamento dos dispositivos na rede. Numa rede IP, o protocolo padrão para distribuição de IP é o DHCP, já numa rede LoRaWAN o mecanismo de endereçamento é diferente. Assim, o software de tradução deverá traduzir os *logs* DHCP e LoRaWAN em uma representação padrão. Essa capacidade poderá ser expandida gradualmente às demais aplicações à medida que elas difiram na representação dos seus *logs*.

Neste trabalho, a fase de tradução será implementada considerando a diversidade de tecnologias de comunicação presentes em um ambiente de Internet das Coisas, sobretudo em Prédios Inteligentes. Para isso, foram consideradas três tecnologias bem conhecidas, cuja principal função seja a conexão e endereçamento de dispositivos na rede: *Long Range Wide Area Network* (LoRaWAN), *Internet Protocol* (IP) e o Bluetooth BLE.

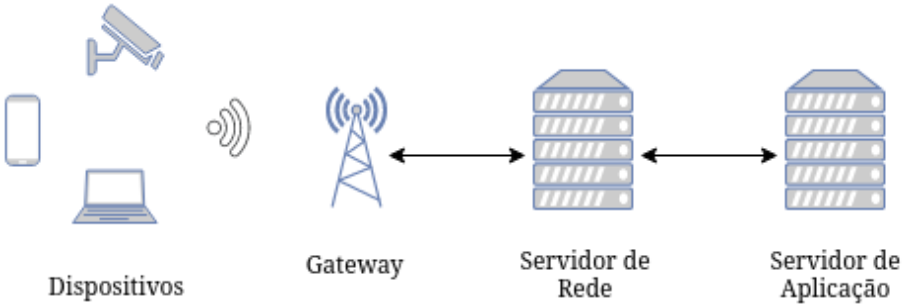
4.2.4.1 LoRaWAN

A especificação LoRaWAN é um protocolo de baixo consumo energético projetado para conectar “coisas” à internet, formando redes regionais, nacionais ou globais. É uma

tecnologia que visa atender alguns dos principais requisitos da IoT, tais como a comunicação bidirecional, segurança ponta-a-ponta, mobilidade e serviços de localização (ALLIANCE, 2015).

As redes LoRaWAN normalmente são dispostas em uma topologia estrela na qual os *gateways* retransmitem mensagens entre os dispositivos finais e um servidor de rede central. O servidor de rede roteia os pacotes de cada dispositivo da rede para o servidor de aplicação associado (ALLIANCE, 2017), como mostra a Figura 17. O protocolo também conta com um servidor de associação que gerencia as chaves criptográficas de sessão.

Figura 17 – Arquitetura de uma rede LoRaWAN



Fonte: Elaborada pelo autor.

Para associar-se a uma rede LoRaWAN, cada dispositivo final deve ser ativado. A ativação de um dispositivo final pode ser realizada de duas maneiras, por meio de ativação pelo ar (OTAA) ou via ativação por personalização (ABP) (ALLIANCE, 2017).

Na ativação pelo ar, os dispositivos finais devem seguir um procedimento de associação com um servidor antes de realizar a troca de dados. Os dispositivos sempre devem passar por um novo procedimento de associação quando perderem as informações de contexto da sessão (ALLIANCE, 2017).

Na ativação por personalização, o identificador de dispositivo e as chaves de sessão são armazenados diretamente no dispositivo, sem prévia interação com o servidor. Assim, o dispositivo final está equipado com as informações necessárias para participar de uma rede LoRa específica assim que ela é iniciada.

Entretanto, antes mesmo da ativação, as seguintes informações devem estar armazenadas nos dispositivos:

– *JoinEUI* - é um identificador de aplicação global que identifica unicamente o servidor de associação;
– *DevEUI* - é um identificador único de cada dispositivo. Tem aplicabilidade global, identificando os dispositivos através de várias redes. Na ativação pelo ar, este identificador

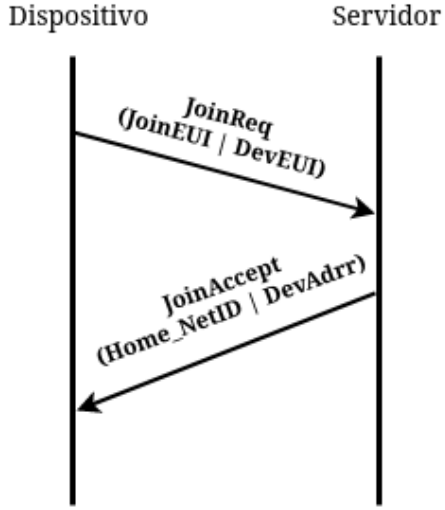
deve ser armazenado no dispositivo antes do processo de associação. Na ativação por personalização, não é necessário que isso ocorra, embora seja recomendado, já que o próprio *Join Server* poderá alocar o DevEUI.

Após a ativação, o *DevAddr* é armazenado no dispositivo final, o qual é um identificador de 32 bits que tem significado apenas na rede local. Esse identificador é alocado pelo servidor da rede.

Após o processo de ativação, inicia-se a fase de associação na rede, que é sempre iniciada pelo dispositivo final ao enviar uma mensagem de *join-request*. Essa mensagem contém os identificadores *JoinEUI* e *DevEUI*.

Ao permitir a associação do dispositivo na rede, o servidor responde com uma mensagem de *join-accept*, que contém um identificador da rede (*NetID*) e um endereço de dispositivo (*DevAddr*). A Figura 18 ilustra as principais mensagens no processo de associação de um dispositivo final à rede.

Figura 18 – Processo de associação simplificado do LoRaWAN



Fonte: Adaptado de Haxhibeqiri *et al.* (2018)

Cabe mencionar que nem todos os dados presentes nas mensagens do protocolo LoRaWAN são relevantes para a fase de tradução implementada neste trabalho. O principal objetivo da proposta é fornecer uma representação de *log* padronizada, com sintaxe e semântica únicas, de tecnologias que tratem do endereçamento de dispositivos de uma rede IoT. Por isso, os campos que tratam de chaves de sessão e criptografia foram desconsiderados.

Os exemplos de *logs* LoRaWAN foram coletadas a partir do AWS IoT Core [2].

[2] https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html

4.2.4.2 Internet Protocol

O IP é um protocolo da camada de rede do modelo TCP/IP cujas principais funções são o roteamento entre redes e o endereçamento lógico de dispositivos (KUROSE; ROSS, 2016). Tecnologias comumente presentes em redes locais, tais como o WiFi (ELECTRICAL; (IEEE), 1997) e a Ethernet (ELECTRICAL; (IEEE), 1983), utilizam os serviços oferecidos pela tecnologia IP. Entretanto, o protocolo IP é carente de um mecanismo nativo para a distribuição e gerenciamento de endereços. Nesse contexto, foi proposto o *Dynamic Host Configuration Protocol* (DHCP), um protocolo amplamente utilizado que automatiza a distribuição e o gerenciamento de endereços IP na rede local, usando a arquitetura cliente-servidor (KUROSE; ROSS, 2016). Dessa forma, uma maneira de capturar informações de endereçamento IP dos diapositivos finais é analisar os *logs* gerados por aplicações clientes DHCP.

Para a implementação da fase de tradução, foram utilizados os *logs* da aplicação cliente *NetworkManager*, uma ferramenta de configuração DHCP padrão no Linux com suporte a servidores, computadores pessoais e dispositivos móveis (PROJECT, 2018).

4.2.4.3 Bluetooth Low Energy

O *Bluetooth Low Energy* (BLE) foi projetado para operações que demandam pouquíssima energia. Transmitindo dados em 40 canais na banda de frequência não licenciada de 2,4 GHz, o rádio *Bluetooth LE* oferece aos desenvolvedores uma enorme flexibilidade para construir produtos que atendam aos requisitos de conectividade exclusivos de seu nicho de mercado. O BLE oferece suporte a várias topologias de comunicação, tais como a ponto a ponto e, mais recentemente, a *mesh*, permitindo que a tecnologia *Bluetooth* suporte a criação de redes de dispositivos em grande escala. Embora inicialmente conhecido por seus recursos de comunicação de dispositivo, o *Bluetooth LE* agora também é amplamente usado como uma tecnologia que também atende à crescente demanda por serviços de localização de alta precisão (PROPRIETARY, 2021b).

Para a implementação deste trabalho, foi utlizado o *log* do *Bluetooth* de um aparelho da marca *Xiaomi* com Sistema Operacional *Android* na versão 9. Primeiramente, para que o dispositivo registre os *logs Bluetooth*, é necessário habilitar o modo desenvolvedor nas configurações do aparelho e, posteriormente, habilitar a opção *Bluetooth HCI snoop log*. Depois de efetuar operações com o *Bluetooth*, o o dispositivo precisar ser reiniciado para que o *log* apareça

no armazenamento do *Android*.

Após a coleta, alguns campos do *log* devem ser identificados, sobretudo o campo de endereçamento dos dispositivos presentes na rede *Bluetooth*. De acordo com Proprietary (2021a), a tecnologia *Bluetooth* utiliza um identificador de 48 *bits* para endereçar os dispositivos, o qual pode ser encontrado no campo *bd_addr* do *log* analisado.

4.2.4.4 Comparando os logs coletados

A fim de que o *software* da fase de tradução gere uma representação padrão de *log*, deve-se primeiro correlacionar as entradas dos *logs* que se deseja representar. Essa tarefa pode ser executada consultando cada *log* individualmente e entendendo a semântica e sintaxe dos arquivos, geralmente com o auxílio de uma documentação técnica, a depender da tecnologia geradora do *log*.

Neste trabalho, foram coletados *logs* DHCP, LoRaWAN e BLE, de modo a gerar uma representação de *log* envolvendo três tecnologias comumente presentes em um ambiente IoT. Ao analisar os três arquivos de *log*, é possível correlacionar entre eles algumas informações importantes para o processo de investigação forense, tal como identificar o momento que determinado dispositivo estava conectado na rede, através dos campos de data e hora e de seu identificador, seja seu endereço IP, endereço físico ou qualquer outro esquema de identificação. A tabela 1 mostra os campos que podem ser correlacionados entre os *logs* DHCP (*NetworkManager*), LoRaWAN e BLE.

Tabela 1 – Correlação entre os campos dos *logs* coletados

Descrição	DHCP	BLE	LoRaWAN
ID local do dispositivo.	*set-hw-addr*	*bd_addr*	*devaddr*
ID global do dispositivo.	*address*		*deveui*
ID do servidor na rede.	*uuid*		*joineui*
Houve conexão.	*connected_global*	*btl2cap.cmd_code*	*event, logLevel*
Houve desconexão.	*disconnected*	*btl2cap.cmd_code*	

Fonte: Elaborada pelo autor.

O campo *devaddr* do protocolo LoRaWAN possui o mesmo objetivo do campo *set-hw-addr* do *log* DHCP e do *bd_addr* do *log* Bluetooth. Eles têm significado apenas no contexto local da rede. O *joineui* é o identificador do servidor de rede do LoRaWAN e pode ser comparado ao campo *uuid* do *log* DHCP. Esse campo possui a finalidade de identificar o servidor DHCP na rede. Adicionalmente, O *deveui* é o identificador global do dispositivo dentre todas as redes, e

tem equivalência semântica ao endereço lógico IP, que está presente no *log* DHCP. Já os campos *disconnected* e *connected_global* do DHCP identificam, respectivamente, se um dispositivo desconectou-se ou conectou-se na rede. Da mesma forma, o campo *btl2cap.cmd_code* do BLE e os campos *event* e *logLevel* do LoRaWAN possuem informações que podem determinar se houve conexão de determinado dispositivo.

4.2.4.5 Representação Padrão dos logs

Para construir um modelo padrão para os *logs* coletados, é necessário analisar a semântica das entradas neles gravadas, estabelecendo uma correlação entre todos os *logs*. Essa atividade pode ser onerosa, haja vista a necessidade de consultar a documentação de cada tecnologia envolvida para melhor entendimento dos registros. Sem uma representação externa e independente, a dificuldade de análise dos *logs* é diretamente proporcional à diversidade de aplicações geradoras de *logs*.

Dessa forma, ao envolver as tecnologias de conectividade IP, LoRaWAN e BLE, e seus *logs*, foi projetada uma representação que abarcasse os campos em comum entre todas elas, tendo em mente, principalmente, a manutenção de informações relevantes no processo de investigação forense, como a data e hora da ocorrência de cada evento e o identificador de cada dispositivo.

O modelo proposto é formado por cinco colunas, cada coluna possui sua própria semântica. A Tabela 2 fornece uma visualização dos cabeçalhos presentes no modelo. Abaixo de cada cabeçalho, constarão as entradas correspondentes. Como regra geral, os *logs*, para permitir a investigação forense digital, devem registrar, pelo menos, informações sobre o que aconteceu, quando aconteceu e quem iniciou determinado evento (MARTY, 2011). Como um todo, a representação poderá consistir de cinco colunas e muitas linhas. Apesar de informações de cabeçalho facilitarem a leitura e interpretação de dados, elas não estão presentes nos *logs* do *NetworkManager* nem do *BLE*.

Tabela 2 – Representação do *log*

GLOBAL_ADDR	LOCAL_ADDR	SERVER_ID	CONNECT	DISCONNECT

Fonte: Elaborada pelo autor.

A coluna **GLOBAL_ADDR** é o identificador global do dispositivo. Já a coluna **LOCAL_ADDR** é o identificador local. **SERVER_ID** é o identificador do processo servidor

DHCP ou de rede do *LoRaWAN*. **CONNECT** mostra o horário da conexão e **DISCONNECT** o horário da desconexão. Com essas informações, os investigadores têm como saber qual dispositivo estava conectado na rede e por qual período.

4.2.5 Análise dos dados

Esta fase recebe como entrada um *log* padrão e tem como objetivos a extração de informações relevantes do *log* e a análise de tais informações para se chegar a uma conclusão, já que o fundamento da ciência forense é usar uma abordagem metodológica para chegar a conclusões baseadas nos dados disponíveis ou determinar que não se pode chegar a qualquer conclusão (NIST, 2006). Aqui também ocorre a identificação de dispositivos, pessoas ou lugares. Nesta fase, um conjunto de dados são cruzados para se obter novas informações.

Nesta fase, a depender do incidente envolvido, os investigadores podem ter acesso a uma grande quantidade de dados e precisarão de ferramentas específicas para auxiliar nas buscas. Por exemplo, para descobrir data e hora que determinado dispositivo sofreu tentativas de acesso, basta analisar os *logs* de somente um dispositivo, os quais estarão identificados no servidor centralizado. Já para casos mais complexos, é imprescindível analisar um número maior de *logs*. Para determinar quem estava na sala de suporte quando um furto aconteceu, é necessário que mais de um arquivo de *log* seja levado em consideração, o que pode incluir *logs* de DHCP, que registram data e hora da distribuição de endereços lógicos na rede, de sistemas operacionais e de aplicações específicas.

No contexto do SBIM, as informações estarão basicamente contidas em *logs* de dispositivos, os quais são arquivos de texto simples, geralmente constituídos por muitas linhas. Cada entrada conta com informações importantes, tais como data e hora de um determinado evento e o significado dele, podendo indicar êxito ou falha ao se executar determinada ação. A notação utilizada na fase de Tradução do SBIM facilita bastante esta fase de Análise, pois a notação naquela fase é independente de qualquer tecnologia.

As buscas em arquivos textuais podem ser realizadas por meio de ferramentas específicas. Para isso, os investigadores precisam estar cientes do que estão procurando. Nesse ponto, é importante que eles conheçam previamente a estrutura dos *logs*, ou seja, sua sintaxe e semântica, para procurar por *strings* específicas. A fase de tradução é essencial para que os investigadores usem aqui o padrão que foi sugerido.

Para realizar as buscar textuais, existem algumas ferramentas *open source* bem

conhecidas. Dentre elas está o *grep* (GNU, 2021a), uma ferramenta que como padrão já vem instalada na maioria das distribuições Linux. A sintaxe do *grep* é bem simples, e pode ser utilizada da seguinte maneira:

Para encontrar o endereço IP *192.168.13.45* no arquivo *log.txt*:

Figura 19 – Ferramenta *grep*

```
$ grep "192.168.13.45" log.txt
```

Fonte: Elaborado pelo autor.

Uma outra opção é gravar a saída do comando *grep* em um arquivo separado, de modo a facilitar a investigação das informações.

Figura 20 – saída da ferramenta *grep*

```
$ grep -ni "192.168.13.45" log.txt >> arquivo.txt
```

Fonte: Elaborado pelo autor.

Todas as linhas com a *string GNU* serão gravadas no arquivo *arquivo.txt* para posterior análise. Isso é uma prática que pode agilizar o trabalho da equipe, já que o número de entradas a serem analisadas vai ser menor.

O *grep* também suporta o uso de Expressões Regulares, as quais podem formar novos padrões de busca em arquivos textuais. As Expressões regulares oferecem uma forma flexível de identificar *strings*, como caracteres particulares, palavras ou padrões de caracteres.

O *grep* também fornece uma maneira de fazer uma análise cruzada dos *logs*. Isso pode ser feito porque o *grep* permite que uma *string* possa ser procurada em vários arquivos simultaneamente.

Figura 21 – *grep* em vários arquivos

```
$ grep -n "192.168.1.45" /var/logs/*
```

Fonte: Elaborado pelo autor.

O comando acima procura por entradas que contenham o endereço lógico de rede *192.168.1.45* em todos os arquivos dentro do diretório *logs*. Se existir subdiretórios, basta acrescentar a opção *-R* para permitir uma buscar recursiva.

Além do *grep*, existem outras ferramentas que permitem a busca de cadeias de caracteres em arquivos de *log*, tais como o *Vim* (MOOLENAAR, 2021) e o *nano* (GNU, 2021b)

que, apesar de serem editores de texto, possuem funções de busca. Como esta fase trata de busca em arquivos de texto, aplicações de edição de texto convencionais, como o *Microsoft Word* (MICROSOFT, 2021) e o *LibreOffice* (FOUNDATION, 2021), também podem ser utilizadas aqui, apesar de serem aplicações que consomem mais recursos computacionais pois possuem muitos recursos gráficos.

Após a extração das informações mais relevantes, faz-se necessário analisá-las. Uma vez que os *logs* foram coletados, é possível realizar o cruzamento de informações e chegar a conclusões razoáveis, como: determinar quais dispositivos estavam conectados à rede e por qual período; identificar quais usuários praticaram determinadas ações; e correlacionar dois ou mais eventos que indiquem a causa de um terceiro.

A primeira informação relevante que é possível levantar utilizando a representação proposta é a data e hora que os dispositivos estavam conectados à rede. A partir desses dados, é possível direcionar a investigação aos dispositivos apropriados. Se um mecanismo de segurança, tal como *firewalls* ou Sistema de Detecção de Intrusão, detectou um comportamento que comprometeu a segurança da informação por algum período, a equipe deve dar prioridade aos dispositivos que estavam conectados à rede durante o mesmo período. Olhando para os *logs* coletados, pode-se perceber a presença de alguns campos que indicam quando o dispositivo conectou-se e desconectou-se da rede, respectivamente identificados pelas colunas de conexão e desconexão na representação implementada. No BLE, para desconectar-se, o dispositivo escravo envia uma requisição ao dispositivo mestre com o código de comando (*command code*) *Disconnection Request*. Após receber uma reposta, o dispositivo escravo encerra a conexão. De forma parecida, o dispositivo escravo envia para conectar-se o comando *btl2cap.cmd_code* seguido do valor 0x02 e espera por uma resposta do dispositivo mestre, o que indicará o momento da sua conexão à rede *Bluetooth*. Para desconectar-se da rede BLE, o valor de *btl2cap.cmd_code* é mudado para 0x06. Já no DHCP, existem duas mensagens que ajudam a diagnosticar o tempo de conectividade de um dispositivo. a *CONNECTED_GLOBAL*, que informa quando o dispositivo conectou-se à rede e a *DISCONNECTED*, que informa que o dispositivo perdeu a conexão. Ou seja, o intervalo de tempo entre as duas mensagens revela por quanto tempo o dispositivo manteve-se conectado na rede.

Com a lista de dispositivos que possivelmente estavam conectados à rede no momento do incidente, é possível, também, identificar quais usuários estavam logados nesse mesmo momento. Mas para isso, pode ser necessário analisar os *logs* do Sistema Operacional. Dessa

forma, o cruzamento dos *logs* facilita o descobrimento de novas informações.

Um outro resultado da análise de dados é a possibilidade de correlacionar dois ou mais eventos da mesma natureza que indiquem a causa de um terceiro. Em uma rede organizacional, se em certos momentos ela sofre com o congestionamento de dados, impedindo a entrega de valor ao cliente, e esses momentos sempre coincidem quando certos dispositivos estão conectados à rede, o que se pode afirmar olhando a hora de conexão e desconexão, então existe uma relação a ser investigada nesse contexto.

4.2.6 Apresentar

A última fase compreende atividades de preparar e apresentar as conclusões obtidas na fase de análise. A apresentação das conclusões pode ser documentada em forma de relatórios físicos ou digitais. O documento final deve reunir todos os dados pertinentes à investigação e o passo a passo das atividades que foram executadas, de modo que o processo possa ser replicado por inteiro. Algumas informações importantes devem vir no relatório: versão das aplicações utilizadas na investigação, nome das aplicações, sistema operacional, algoritmos utilizados para a preservação da cadeia de custódia e a identificação de todo o pessoal envolvido no processo investigativo. Os dados utilizados devem ser resguardados pelo período determinado pela organização. O tempo deve ser o suficiente para que todas as partes interessadas possam certificar-se da validade do processo. Segundo (NIST, 2006), muitos fatores afetam a fase de apresentação, incluindo considerações sobre o público que consumirá a apresentação. Se o processo investigativo reunir evidências que tramitem em um processo legal, a apresentação deve reunir relatórios detalhados de toda a informação reunida. De outro lado, se o destinatário for um administrador de redes, por exemplo, ele vai querer ver informações sobre o tráfego da rede e dados estatísticos. Um gerente da organização, por sua vez, precisa somente de uma visão de alto nível do que foi reunido e de como o processo foi executado. Isso permite que o gerente determine o que pode ser melhorado no futuro.

4.3 Conclusão

Este capítulo definiu o SBIM e descreveu todas as suas fases proativas: *Forensic Readiness* e Capacidade Forense; e reativas: Coleta, Tradução, Análise e Apresentação, incluindo formas de implementação para cada uma delas em um ambiente de Prédios Inteligentes. Espera-se

que as duas primeiras fases, ao executarem de maneira contínua, assegurem a eficácia do modelo, pois elas promovem, respectivamente, a disponibilidade dos *logs* através do seu gerenciamento centralizado e a capacidade de a organização executar o processo forense. A fase de Tradução, por sua vez, preocupa-se em aumentar a eficiência do processo forense, oferecendo uma notação única para a representação dos *logs* coletados a partir de diferentes tecnologias. As demais fases foram descritas de acordo com o NIST e, com isso, espera-se que sejam eficazes.

O próximo capítulo apresenta o processo conduzido para validar o SBIM.

5 VALIDAÇÃO DA FASE DE TRADUÇÃO

O modelo SBIM agrega a fase de tradução que busca aumentar a eficiência da equipe de investigação forense. Por eficiência, entende-se a dimensão de desempenho do processo considerando os custos ou esforços de execução envolvidos (Sá, 2020), ou seja, uma equipe pode ser mais eficiente ao utilizar menos tempo e menos recursos para executar corretamente um processo.

Espera-se que a fase de tradução aumente a eficiência da equipe de investigação por alguns motivos, quais sejam:

- Diminuição do tempo de consulta a documentações externas;
- Menor esforço de aprendizagem;
- Menor tempo de consulta aos *logs*;
- Melhor compreensão dos *logs*.

Em uma organização ausente de representações padrões para os *logs* de seus sistemas, os profissionais devem buscar as documentações de cada tecnologia para entender a sintaxe e a semântica dos *logs* envolvidos. Quando existe somente uma única representação provida pela organização, a consulta de documentos externos torna-se desnecessária. Isso ocasiona, invariavelmente, um menor esforço de aprendizagem por parte da equipe, já que os indivíduos preocupar-se-ão com entender somente as representações providas e documentadas pela própria organização. Além disso, é bastante provável que a equipe gaste menos tempo analisando os *logs* pertinentes, pois as representações condensam apenas as informações mais relevantes para o processo forense.

Diante disso, a validação quanto a eficiência do SBIM será conduzida por meio da implementação da fase de Tradução, responsável por gerar uma representação única para cada tecnologia de comunicação no Prédio Inteligente, e através de questionários respondidos por pessoas com notórios conhecimentos na área ou pesquisadores. Os questionários foram elaborados com itens que avaliem a entrega de eficiência do modelo. Os questionários foram criados e distribuídos utilizando o *Microsoft Forms* [1]. Neste capítulo, o processo de validação da pesquisa é detalhado, a estrutura dos formulários é apresentada, explicando como os itens foram projetados e, depois, os resultados são discutidos.

[1] https://forms.microsoft.com/

5.1 O processo de validação

Antes da realização da validação, é importante seguir uma abordagem sistemática que defina um conjunto de passos a serem cumpridos. Neste trabalho, é conduzida uma avaliação com o objetivo de determinar o tempo de resposta e a corretude dos formulários de pesquisa submetidos pelos participantes. O resultado dessa avaliação pode revelar o real ganho de eficiência ao usar o modelo SBIM. Os passos para o estudo de avaliação de desempenho dos respondentes foram definidos da seguinte forma:

1. Listar os objetivos do estudo;
2. Selecionar as métricas;
3. Selecionar a carga;
4. Listar os parâmetros;
5. Selecionar os fatores;
6. Analisar, interpretar e apresentar os resultados.

O principal objetivo deste processo de validação consiste em demonstrar qual impacto positivo a fase de Tradução proporciona ao modelo SBIM, ou seja, em termos de eficiência, espera-se que essa fase traga benefícios para a equipe investigativa.

As métricas utilizadas para atestar o aumento de eficiência quando da presença da fase de Tradução é o tempo levado para responder as perguntas juntamente com a corretude das mesmas. Essas métricas são constituídas pelo tempo que os respondentes dos questionários levam para analisar os itens propostos e submeter suas respostas, e se as respostas estão corretas ou não. Basicamente, o tempo de consulta é a soma dos tempos de leitura dos *logs* mais o tempo de acesso a informações externas, tais como buscas por documentações na Internet.

A carga dos *logs* usados no processo de validação foram gerados localmente, pelas aplicações instaladas em cada dispositivo, com exceção dos *logs* LoRaWAN, os quais foram gerados por aplicação desenvolvida pelos autores deste trabalho, seguindo o modelo disponibilizado pela AWS IoT Core [2]. Foram gerados *logs* para aproximadamente 100 dispositivos LoRaWAN. Já o *log* do *Bluetooth* foi extraído de um aparelho da marca Xiaomi com Sistema Operacional *Android* na versão 9. Primeiramente, para que o dispositivo registre os logs Bluetooth, é necessário habilitar o modo desenvolvedor nas configurações do aparelho e, posteriormente, habilitar a opção Bluetooth HCI snoop log. Depois de efetuar operações com o Bluetooth, o dispositivo precisar ser reiniciado para que o *log* apareça no armazenamento do Android. O *log* contém

[2] https://docs.aws.amazon.com/iot/latest/developerguide/what-is-aws-iot.html

exatamente 655 entradas registradas ou linhas. Uma informação relevante é que o Android versão 9 escreve seus *logs* em arquivos com a extensão ".cfa". Sendo assim, necessita-se de aplicação adicional para que se efetue a leitura dos registros, já que um editor de texto convencional não é capaz de entender arquivos com tal extensão. A fase de tradução propõe-se a resolver problemas dessa natureza, poupando a equipe investigativa de pesquisar e aprender ferramentas que em nada tenham a ver com a atividade fim do processo forense digital. O *log* IP, por sua vez, foi gerado pela aplicação *NetworkMananager* de um notebook com sistema operacional *Arch Linux*. O *log* original contém aproximadamente 60.000 entradas, com eventos de conexões, desconexões, endereço lógico e físico.

O parâmetro considerado na avaliação, ou seja, a informação que pode afetar a métrica de tempo de resposta do respondente, é a quantidade de linhas presentes no *log*. Aqui, cabe destacar que a fase de Tradução pode reduzir em até mil vezes a quantidade de entradas em um *log* mantendo as informações pertinentes para um processo forense.

Já os fatores utilizados, ou seja, a alteração do parâmetro identificado, são três variações no tamanho dos *logs*, a saber: grandes, médios e pequenos. Os *logs* grandes não sofrem modificações; os *logs* médios possuem metade do número de linhas de um *log* grande, da sua respectiva tecnologia; e os pequenos possuem metade do número de linhas dos médios.

O último passo da abordagem consiste em analisar, interpretar e apresentar os resultados buscando validar a maior eficiência do modelo oferecido. Nessa fase, será analisada a variação do tempo de resposta dos formulários que contêm perguntas sobre os *logs* grandes, médios e pequenos das tecnologias IP, BLE e LoRaWAN em comparação com os formulários que contêm perguntas sobre o *log* padrão gerado pela fase de Tradução. Ou seja, será que a representação proposta mantém a eficiência diante de *logs* grandes, médios e pequenos?

5.2 Estrutura dos formulários

Foram submetidos 14 formulários de pesquisa, totalizando 70 perguntas. Com exceção de dois, que indagavam sobre informações profissionais e aspectos subjetivos, todos correspondiam a perguntas sobre uma das três tecnologias utilizadas em Prédio Inteligentes, quais sejam: IP, LoRaWAN e BLE. Os formulários foram submetido a 27 participantes, os quais deveriam responder a todas as perguntas, a fim de validarem a eficiência do modelo proposto. Foram elaboradas perguntas de natureza investigativa conforme exemplos abaixo.

Para os *logs* IP:

- Que horas o dispositivo com IP 192.168.1.9 conectou-se pela última vez?
- O dispositivo com IP 10.0.0.108 passou quanto tempo conectado na rede no dia 9 de Março?
- Que horas o dispositivo com IP 192.168.1.3 conectou-se à rede?
- O dispositivo IP 192.168.1.3 conectou-se alguma vez à rede “Oi 31EC”?
- O IP 10.0.0.107 conectou-se primeiro à rede “Asgard” do que o IP 10.0.0.105.

Para LoRaWAN:

- Que horas o dispositivo com o endereço local “cee11d9f-81e6-a5ab-1f1c-51c1ca13777e” conectou-se à rede?
- Qual o endereço local do dispositivo com endereço global “111435912f75ff55”?
- Que dia, mês e ano o dispositivo “dcba2ff14e8bd9e5” conectou-se à rede?
- O dispositivo “bd9d862e4d6c51b1” conectou-se alguma vez à rede?
- O dispositivo “4e19b3cfe1b39e36” conectou-se primeiro à rede em comparação ao dispositivo “16ac427a8ef9e2c3”.

Para BLE:

- Que horas o dispositivo com o endereço local “48:2c:a0:ba:15:35” conectou-se à rede pela primeira vez?
- Quando o dispositivo “30:c0:1b:7e:b1:a0” desconectou-se pela primeira vez?
- Quando o dispositivo “48:2c:a0:ba:15:35” desconectou-se pela segunda vez?
- Quando o dispositivo “30:c0:1b:7e: b1:a0” conectou-se pela penúltima vez?
- A primeira desconexão entre os dispositivos “48:2c:a0:ba:15:35” e “30:c0:1b:7e:b1:a0” ocorreu no dia 10 de dezembro.

5.3 Resultados

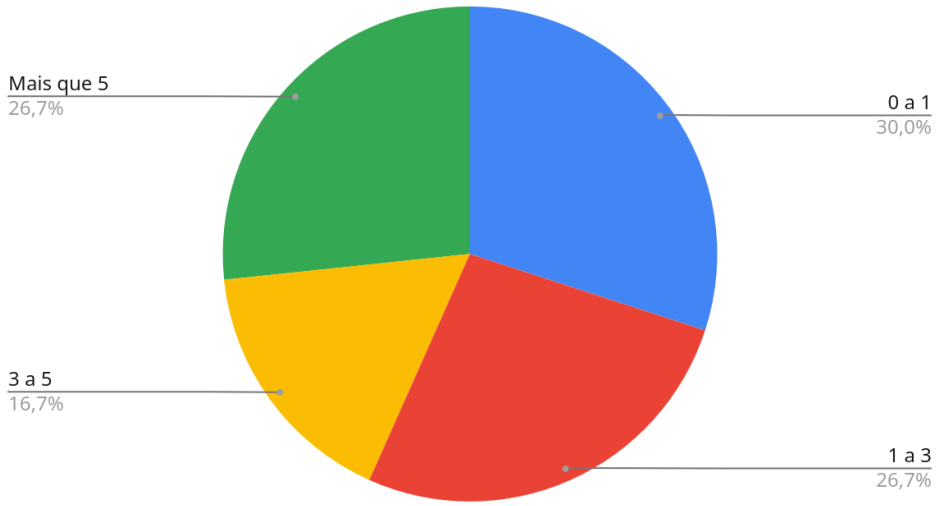
Os formulários foram submetidos a dois públicos distintos, porém pertencentes a mesma área fim, Tecnologia da Informação, com ênfase em auditoria, segurança da informação e/ou forense computacional. O primeiro grupo de respondentes foi composto por Auditores Fiscais de Tecnologia da Informação da Receita Estadual do Ceará e Analistas de TI dessa mesma Secretaria da Fazenda. O segundo grupo foi composto por estudantes de pós-graduação na área de segurança da informação. No total, 27 pessoas participaram da pesquisa.

Antes mesmo de introduzir as perguntas de cunho puramente técnico, os formulários também captaram informações profissionais, tais como: a quantidade de anos de experiência em

TI, se o respondente tinha prévia experiência com ferramentas de processamento de texto, qual o Sistema Operacional utilizado em suas tarefas profissionais e qual o seu campo de atuação dentro da Área de Tecnologia da Informação.

Sobre os anos de experiência de cada um, foram dispostas quatro opções no formulário, das quais somente uma poderia ser escolhida: 0 a 1 anos de experiência, de 1 a 3 anos de experiência, de 3 a 5 anos de experiência e mais do que 5 anos de experiência. Como se pode ver na Figura 22, os resultados mostram uma boa distribuição dos respondentes em cada um dos intervalos de experiência.

Figura 22 – Anos de experiência



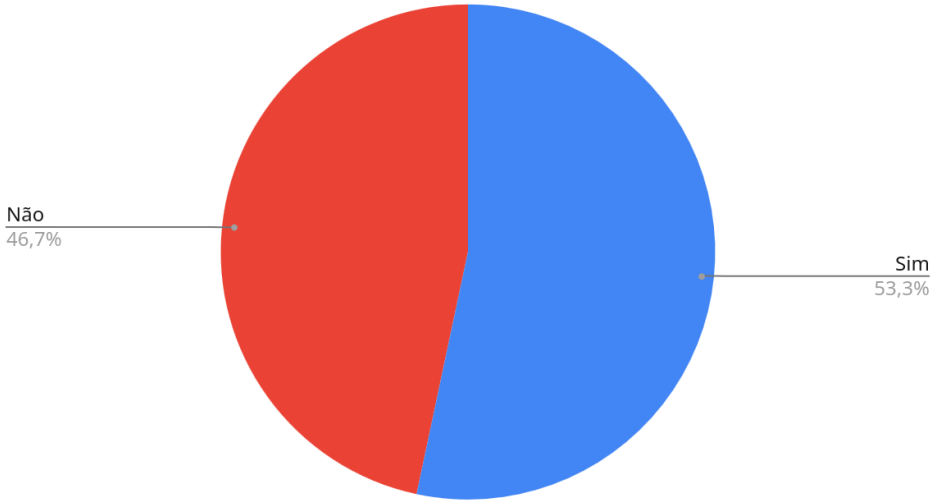
Fonte: Elaborada pelo autor.

A respeito do conhecimento prévio em ferramentas de processamento de texto, essa também foi uma informação importante levantada. Pois, existem ferramentas capazes de realizar busca por padrões e otimizar o tempo de análise. No formulário, como exemplos, foram citadas *grep* e *awk*. Conforme mostram a Figura 23, mais da metade dos respondentes detinham conhecimentos em tais ferramentas. Embora o uso desses conhecimentos não tenha sido requisito obrigatório para participar da pesquisa, eles podem potencializar a atividade forense.

Além disso, foi perguntado qual Sistema Operacional o respondente tinha mais familiaridade. Já na figura, pode-se observar que mais de 70% das pessoas preferiam o Sistema Operacional Windows da Microsoft; 23% utilizam distribuições GNU/Linux para as suas atividades profissionais; e outros 3% escolheram como preferido outro Sistema Operacional não explicitamente elencado na pesquisa, conforme a Figura 24.

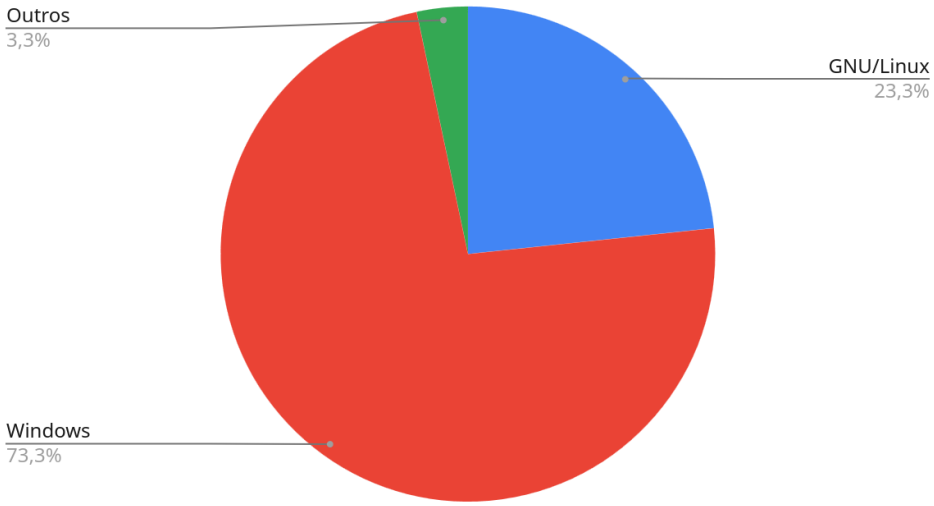
Por último, também foram trazidas informações sobre a área de atuação profissional

Figura 23 – Ferramentas de processamento de texto



Fonte: Elaborada pelo autor.

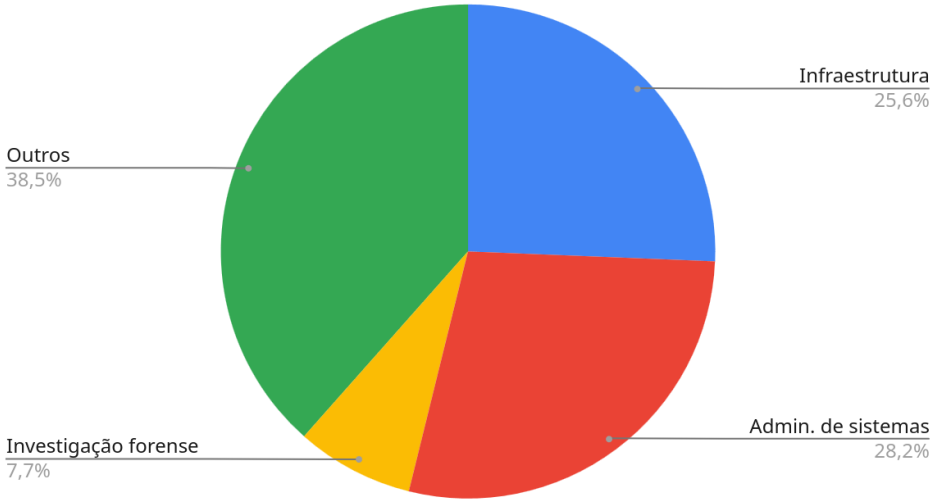
Figura 24 – Sistema Operacional



Fonte: Elaborada pelo autor.

de cada pessoa que respondeu o formulário. Dentre as áreas elencadas, estava a área de Infraestrutura, que engloba toda a parte de configuração de ativos de rede e conectividade, a área de Administração de Sistemas, que envolve a administração de Sistemas Operacionais *desktop* e servidores e a área de investigação forense. Ademais, o respondente poderia marcar a opção "Outros", quando, obviamente, não se encaixasse em nenhuma das áreas mencionadas. Já sobre as áreas explicitamente elencadas no formulário de pesquisa, acredita-se fortemente que são áreas cujos profissionais detêm conhecimentos técnicos para extrair informações relevantes em processos investigativos. Como a análise ocorre basicamente em logs, certamente o corpo técnico de Infraestrutura e Administração de Sistemas de uma organização detém a expertise necessária para executar a recuperação de tais logs dos ativos.

Figura 25 – Área de experiência



Fonte: Elaborada pelo autor.

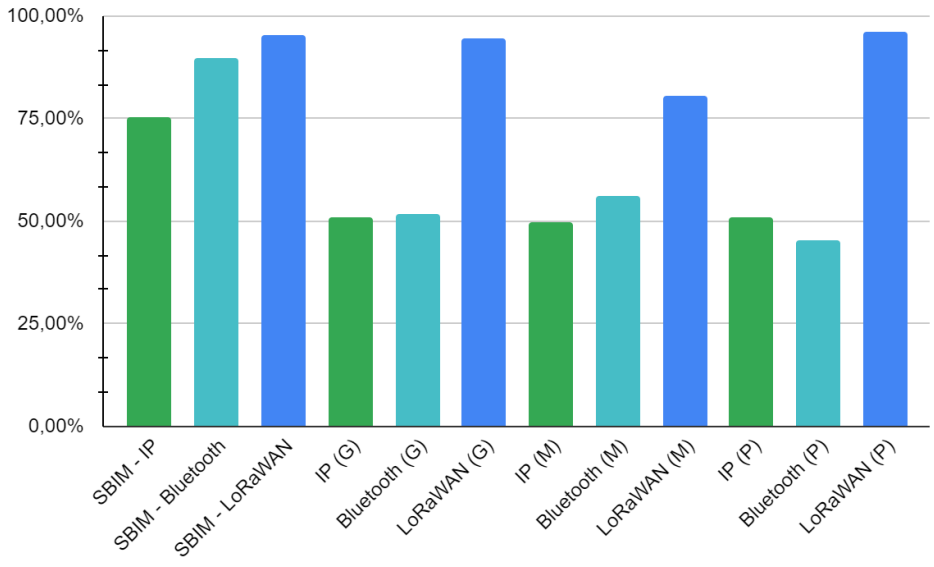
De acordo com a Figura 22, 25% dos respondentes tinham experiência em infraestrutura, 7% em investigação forense, 28% em administração de sistemas e 38% tinham experiência em outras áreas, tais como desenvolvimento, *DevOps*, segurança da informação, etc. É importante ressaltar que era possível escolher mais de uma área de atuação, se necessário.

Com a intenção de mensurar a eficiência do processo proposto, escolheu-se verificar o número de respostas corretas, assim como o tempo necessário para submeter cada uma delas. Dessa forma, o parâmetro de comparação foram os formulários que foram respondidos com base em análise de arquivos alheios à representação sendo proposta. Ou seja, com a coleta dos dados, esperou-se responder se o SBIM é capaz de entregar melhores resultados na análise de logs em ambientes IoT. Mais especificamente, se foi possível aumentar o número de respostas corretas e diminuir o tempo de busca nos logs.

Analisando-se a Figura 26, pode-se notar um incremento considerável de respostas corretas quando da utilização da representação de logs do SBIM. É importante, assim, comparar os ganhos em acertos para cada uma das tecnologias em questão. Quais sejam: IP, LoRaWAN e Bluetooth. Já as letras (G) de grande, (M) de médio e (P) de pequeno denotam os tamanho dos logs analisados para responder às perguntas dos formulários.

Analisando os logs no contexto da tecnologia IP, é notório que a quantidade de acertos aumentou significativamente quando os respondentes analisaram as informações de log de acordo com o SBIM. Seja comparando com os logs grandes, médios ou pequenos, a representação proposta sempre se sobressai em itens respondidos corretamente. Por conseguinte, é necessária a comparação entre o resultado do número de acertos entre os quatro formulários

Figura 26 – Porcentagem de acertos



Fonte: Elaborada pelo autor.

que compreendem a tecnologia IP. No gráfico, a porcentagem de acertos em tais formulários está representada por barras verdes.

Inicialmente, pode-se notar um aumento de 24,44% na quantidade de respostas corretas a favor do SBIM, quando comparado a arquivos de logs IP maiores. Além disso, o SBIM apresentou aumento de 25.92% em respostas corretas, quando comparado a logs médios. Já para os logs pequenos, a proposta apresentou uma melhoria de 24.45%. Dessa forma, enxergando pouca variação de acertos, mesmo variando o tamanho dos arquivos de log, pode-se concluir que a quantidade de respostas corretas podem não ter relação com o tamanho do log em si, mas com a maneira como esses logs são apresentados sintaxe e semanticamente aos analistas. Por exemplo, realizar uma busca textual por determinado endereço lógico, pode ser facilmente executada utilizando ferramentas para esse fim, tais como o grep ou awk, ou até mesmo utilizando editores de texto. Todavia, o menos trivial é determinar o significado de cada entrada no log a qual contém a cadeia de caracteres sendo buscada.

Em seguida, analisando o desempenho do modelo quanto à próxima tecnologia, Bluetooth, pode-se constatar que houve um aumento ainda maior do número de respostas corretas. Ou seja, o SBIM possibilitou uma taxa de acertos próxima dos 90% (89,62%). Ao passo que os logs de tamanhos grande, médio e pequeno apresentaram, respectivamente, taxas de acertos de 51,84%, 56,29% e 45,18%. Dito isso, o ganho sobre logs grandes foi de 37.78%, sobre logs médios correspondeu a 33.33% e sobre pequenos, 44.44%. Da mesma forma que os IP, o tamanho dos logs não teve importância relativa ao número de acertos. Haja vista a quantidade

de itens corretos sobre logs pequenos terem sido até menor quando comparada aos logs médios e grandes. Já quando se compara os médios e grandes, a variação é na casa dos 4%.

Em certo, existem algumas características intrínsecas aos logs Bluetooth que merecem atenção e podem ser a causa do maior número de erros pelos analistas. Quais sejam: o grande número de entradas geradas no arquivo de log, mesmo com a presença de apenas dois dispositivos e conectados por um curto período de tempo, a existência de campos em hexadecimal que representam estados de conexão, e o uso de extensão .cfa, em especial aos dispositivos Android. Embora, na pesquisa, os respondentes tenham tido a opção de analisar logs bluetooth no formato XML, o formato de arquivo originário era em .cfa, o qual não é interpretado da maneira correta por ferramentas convencionais de processamento de texto. Isso muito provavelmente deve influenciar no tempo de análise dos logs, já que as ferramentas apropriadas devem ser primeiramente encontradas.

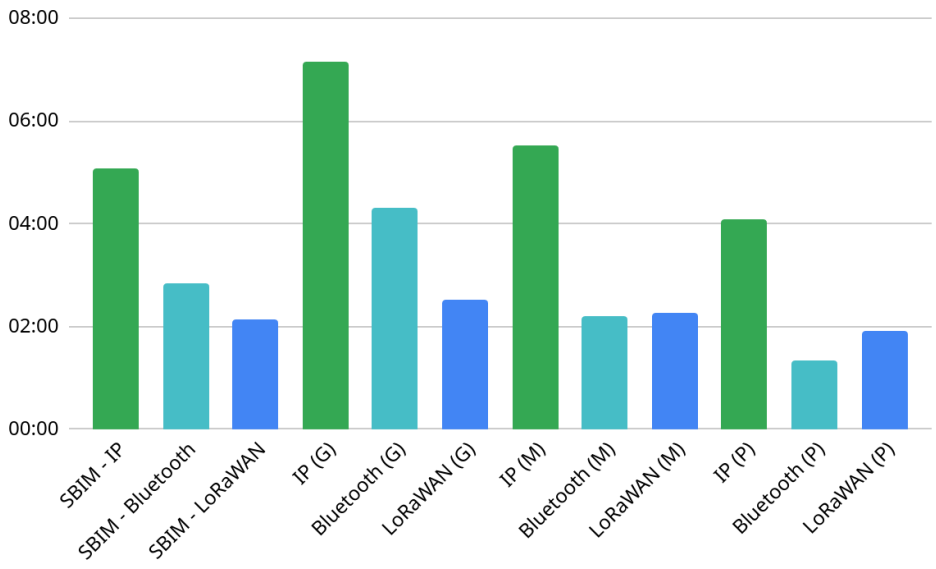
Por último, há que se discutir os resultados alcançados a respeito de logs LoRaWAN. É notório que nesse contexto a abordagem proposta obteve os menores ganhos em termos de acertos. Entretanto, cabe, primeiramente, expor fator que explica satisfatoriamente os resultados alcançados.

Surge-se daí, o fato de que os logs usados para essa tecnologia já eram previamente formatados, ou seja, seguiam um padrão bem conhecido no mercado: JSON. Esses logs foram gerados dinamicamente e seguiram a mesma semântica proposta pela AWS, em seu ambiente de IOT para LoRaWAN. A escolha de prosseguir com a pesquisa mesmo utilizando o JSON como comparativo, se deu por conta da dificuldade de encontrar logs nativos de dispositivos LoRaWAN em tempo hábil para a execução deste projeto. Em contrapartida, a escolha viabilizou um exame pragmático ao comparar os resultados a partir de uma abordagem já bem estabelecida no mercado. E, assim, reforçar a hipótese de que são necessárias representações padronizadas no campo de análise forense em logs no ambiente IoT.

Como evidência, registrou-se o resultado de 95.55% em acertos, quando analisada a proposta do SBIM, 94.81% em logs grandes, 80.73% em logs médios e 96.29% em logs pequenos. Como se vê, o modelo proposto pouco variou quando comparado a logs pequenos e logs grandes. Destoando apenas quando se compara aos logs médios, contribuindo a uma taxa de aproximadamente 15% em acertos.

O que talvez explique essa divergência entre os logs médios e os demais, seja a interpretação equivocada de algum enunciado no formulário de pesquisa, já que, de fato, essa

Figura 27 – Tempo gasto



Fonte: Elaborada pelo autor.

discrepância não era esperada. Pois, os resultados dos logs pequenos e grandes mantiveram-se bastante próximos.

Dessa forma, os resultados obtidos com os logs LoRaWAN foram bastante promissores. Pois, teve-se a oportunidade de comparar o modelo proposto no SBIM com o JSON e verificar as altas taxas de acertos em ambos. O que somente reforça a aplicabilidade dessas representações em ambientes IoT e a importância que ambas podem ter nesse contexto.

Além da corretude das respostas, foi possível medir o tempo que os respondentes gastaram em cada um dos formulários. Essa métrica de avaliação, quando utilizada em conjunto com o número de acertos, pode determinar, por exemplo, se um modelo é mais eficiente que outro. Nesse sentido, de acordo com a Figura 27, pode-se observar uma comparação entre os tempos medianos de resposta dos formulários submetidos. Aqui, foi escolhida a mediana dos tempos de resposta como forma de proteção a valores extremos.

Como os formulários consideraram tempo de resposta todo o intervalo desde o acesso a eles até o momento da submissão, existiu a possibilidade de algum respondente deixar o formulário aberto enquanto fazia outra atividade não relacionada às respostas. E, também, houve a possibilidade de alguns terem submetido as respostas muito rapidamente, por realmente não entenderem a semântica dos logs ou não quererem prolongar o tempo de busca.

Começando pelos logs IP, pode-se verificar um aumento no tempo de análise em torno de 41% em logs grandes, quando não foi utilizada a fase de Tradução. Já em logs médios, a variação foi de apenas alguns segundos, o que representou aproximadamente um acréscimo

de 9% no tempo de análise. E, por fim, os logs pequenos, os quais obtiveram tempo de análise inferior à proposta deste trabalho.

Entretanto, é relevante mencionar que a métrica de tempo de resposta pode não ter importância alguma quando analisada isoladamente de outros fatores, como, por exemplo, o número de acertos das questões. Senão, o que agregaria um modelo que se preocupa apenas com a diminuição do tempo de análise de logs e não com a corretude/precisão daquilo que está sendo buscado?

Nos logs bluetooth grandes, sem a fase de Tradução, houve um aumento significativo de 50% no tempo de análise e uma queda de mais de 37% na taxa de acertos. Ou seja, o SBIM, além de acelerar o processo de análise, também foi bastante eficaz. Já nos logs médios e pequenos, por sua vez, não houve ganhos com relação ao tempo de análise quando se utilizou a fase de Tradução. Entretanto, cabe ressaltar que o desempenho em respostas corretas foi bastante comprometido quando da ausência da fase proposta pelo SBIM.

Por fim, os logs LoRaWAN grandes, médios e pequenos, que, por suas características já mencionadas, mantiveram-se sempre próximos ao desempenho do SBIM, tanto em relação ao tempo como na eficácia das análises. Nesse contexto, os logs grandes apresentaram um aumento do tempo de análise em 17%; os logs médios apresentaram uma subida de tempo em torno dos 5%, embora esses tenham sido os que apresentaram pior desempenho em número de acertos; e os logs pequenos, que propiciaram uma análise mais rápida, em torno dos 10%, além de um desempenho em número de acertos levemente superior.

5.3.1 Informações Adicionais

Conforme mencionado anteriormente, a pesquisa também captou a opinião dos participantes sobre o modelo proposto. Essa parte foi importante, pois permitiu verificar se, na representação padronizada dos *logs*, informações sobre conexões e desconexões são compreendíveis, se as informações do modelo são úteis para uma investigação forense, se a sintaxe e semântica da representação são fáceis de entender, se os respondentes enxergam alguma necessidade de representação única para os *logs* e se eles, os participantes da pesquisa, considerariam utilizar o SBIM em investigações futuras.

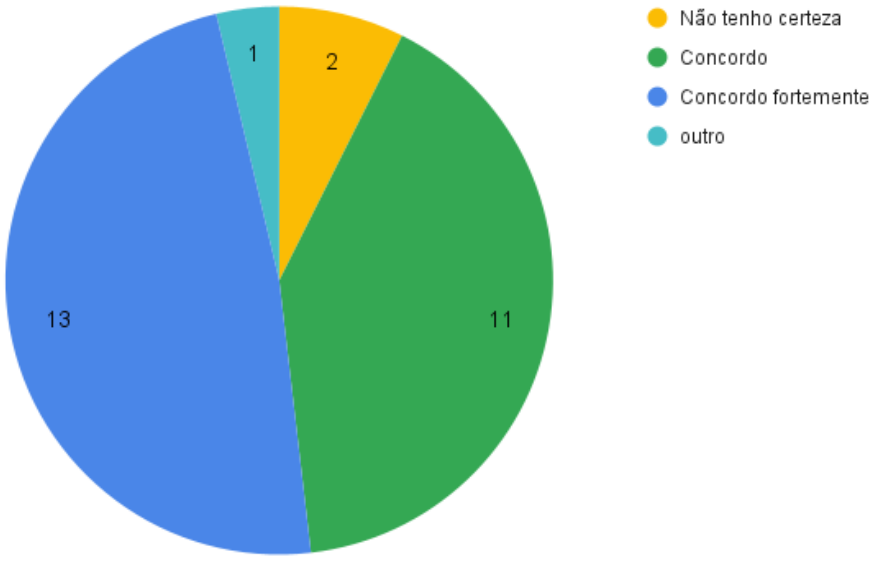
Nesse sentido, os questionamentos da pesquisa buscaram levantar níveis de concordância por parte dos participantes. Dos seis itens a serem respondidos, cinco foram objetivos e um, subjetivo, o qual requisitava considerações ou comentários adicionais sobre a proposta. Para

cada um dos cinco itens objetivos, o respondente só poderia escolher um. E para cada uma das perguntas, existiam seis respostas possíveis, com exceção de duas, que aceitavam somente “sim” ou “não” como resposta. As seis possíveis respostas estão elencadas a seguir:

1. Não concordo fortemente;
2. Não concordo;
3. Não tenho certeza;
4. Concordo;
5. Concordo fortemente;
6. Outro.

Com o primeiro item, buscou-se verificar o nível de compreensão do participante a respeito das informações de conexão e desconexão dos ativos em *log*. Informação de grande relevância em investigação forense. Os resultados do questionamento estão ilustrados na Figura 28. Dentre os que responderam, 88,8% concordaram em diferentes níveis que sim, as informações sobre os eventos de conexão e desconexão dos dispositivos estão apresentadas de forma clara.

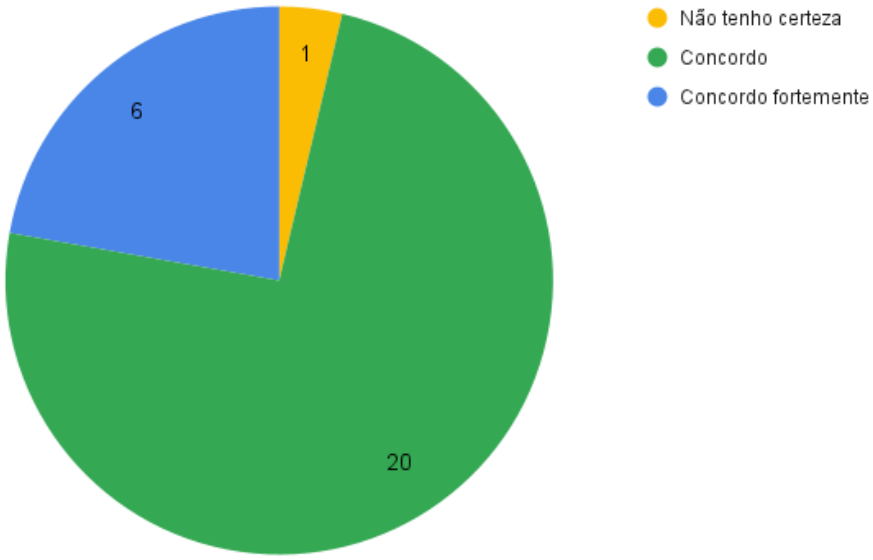
Figura 28 – Compreensão das informações de conexão e desconexão



Fonte: Elaborada pelo autor.

A segunda pergunta do questionário teve a intenção de coletar as opiniões sobre a utilidade das informações mostradas nos *logs* em um processo de investigação de incidentes. Como mostra a Figura 29, apenas uma pessoa não teve certeza daquilo que estava sendo perguntado. Isso quer dizer que 96,2% das pessoas concordaram que as informações são relevantes no contexto de incidentes.

Figura 29 – Utilidade das informações



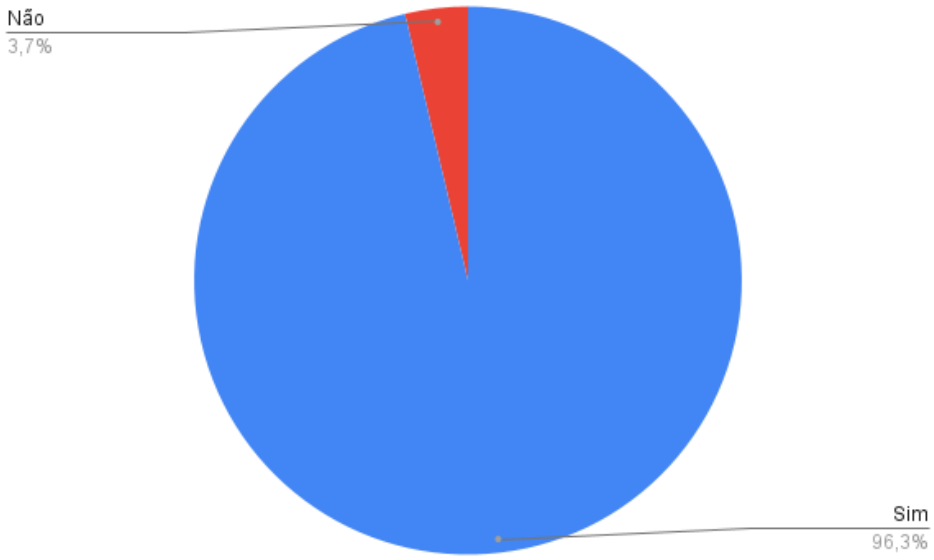
Fonte: Elaborada pelo autor.

Os participantes também foram indagados sobre a facilidade de entendimento no que diz respeito à sintaxe e semântica da representação proposta. Assim, 81,4% concordaram que os *logs* possuem uma boa sintaxe e semântica, conforme a Figura 30. Já 96,3% reconheceram a necessidade de os *logs* seguirem algum padrão para representação de informações em um cenário de investigação de incidentes (Figura 31). Ademais, 88,9% responderam que consideraria utilizar a representação proposta em investigações futuras (Figura 32).

Figura 30 – Sintaxe e semântica fáceis de compreender

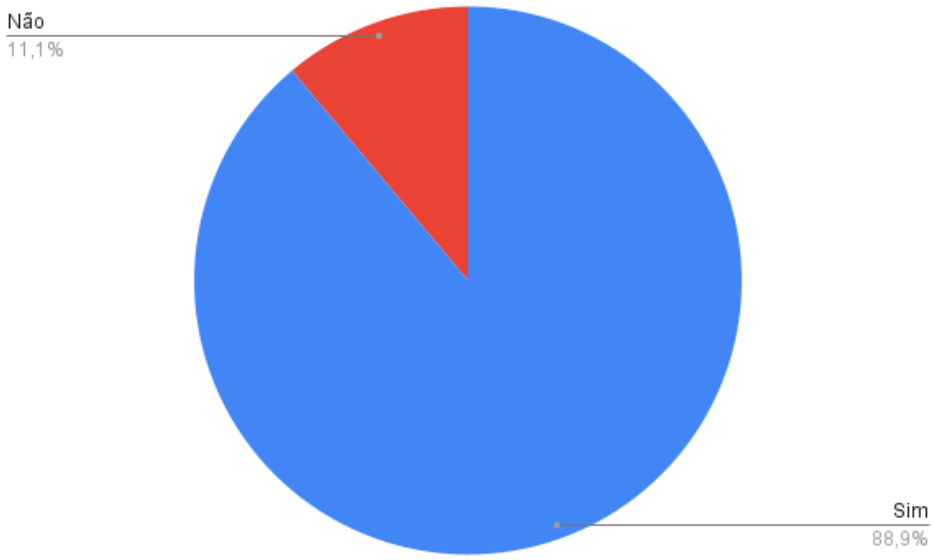
Fonte: Elaborada pelo autor.

Figura 31 – Necessidade de um pradrão para *logs*



Fonte: Elaborada pelo autor.

Figura 32 – Uso do SBIM em investigações



Fonte: Elaborada pelo autor.

Para finalizar a pesquisa, os participantes tiveram a oportunidade de submeter suas considerações a respeito do padrão de *logs* que faz parte do escopo desta dissertação de mestrado. Como se pode ver na Tabela 3, a proposta recebeu bons comentários devido a sua facilidade de entendimento em comparação às representações convencionais de *logs*.

Tabela 3 – Opiniões dos participantes

"Os arquivos .csv estão bem estruturados."
"Com os arquivos csv, as informações são identificadas mais rapidamente."
"Uma ótima representação sobre padrão de logs espero aprender mais sobre."
"Fica mais fácil de entender com uma representação de logs."
"Para análise de qualquer coisa, no geral é bem melhor ser padronizado."
"JSON e CSV ficam mais faceis do que o TXT."
"Os arquivos propostos em CSV foram mais fáceis de analisar, com menos ferramentas comparado aos demais."
"Achei bem interessante e sem dúvidas, os arquivos csv são mais fáceis e intuitivos de navegar que os outros."
"Os logs geralmente tem um padrão definidos para diferentes tecnologias, o uso de um padrão pode melhorar a construção das ferramentas de automação e criação de datasets para o uso de técnicas de detecção de invasão utilizando machine learning."
"A respeito do padrão de logs apresentado achei bem adequado, e disponibilizado de forma correta, gosto sempre de limitar minhas buscas por período de tempo, nível de log."
"Os csv facilitam a identificar as informações."
"A analise de Logs em arquivo Json se mostrou bem mais problemática em comparação com os logs já parseados".

Fonte: Elaborada pelo autor.

5.4 Conclusão

Neste capítulo foi explicado o processo de validação do SBIM no que diz respeito a seu aspecto de eficiência, processo do qual fez parte a coleta de respostas de participantes da pesquisa a partir de formulários. Desse ponto, os resultados foram apresentados e discutidos. Além das respostas puramente técnicas, os participantes também tiveram a oportunidade de prover suas opiniões sobre a proposta apresentada e considerar seu uso em projetos futuros.

O próximo capítulo foca na apresentação da conclusão deste projeto de Mestrado, indicando os trabalhos futuros e perspectivas para o uso do SBIM.

6 CONCLUSÕES

6.1 Caracterização e Contribuição da Pesquisa

Como um campo de pesquisa emergente, a demanda por dispositivos IoT está crescendo, assim como as oportunidades para explorar vulnerabilidades de segurança provenientes da adoção de novos protocolos e tecnologias. Nessa perspectiva, há uma escassez de padrões amplamente aceitos para investigar incidentes em TI no contexto de IoT e suas aplicações, como em Prédios Inteligentes. Um modelo para investigações forenses em IoT pode ajudar a estabelecer um entendimento comum no campo de pesquisa, facilitar o compartilhamento de conhecimento e contribuir para a evolução e consolidação deste campo.

Neste projeto de Mestrado, foi estabelecido o SBIM, um modelo eficiente de investigação forense para Prédios Inteligentes. O SBIM baseou-se no modelo do NIST, adicionando duas novas fases as quais possuem atividades que aumentam a eficiência do processo investigativo, sobretudo no que diz respeito à analise uniformizada de *logs* de eventos em TI. As principais contribuições deste projeto de Mestrado são as seguintes:

1. Consolidar o campo de Investigação Forense em Internet das Coisas, estabelecendo um modelo para investigação em Prédios Inteligentes;
2. Expandir do modelo de investigação forense digital criado pelo NIST;
3. Fornecer uma implementação capaz de padronizar o formato de arquivos de *logs* das tecnologias IP, LoRaWAN e BLE;
4. Aumentar a eficiência das equipes de investigação forense.

6.2 Dificuldades e Limitações

O maior desafio desta pesquisa de Dissertação de Mestrado foi a fase de validação do modelo proposto. Dentre as dificuldades, estavam encontrar participantes que tivessem condições de responder os formulários, captar o tempo de resposta de cada formulário e convencer cada participante a responder treze formulários longos de forma ininterrupta.

Primeiramente, tentou-se encontrar especialistas na área de perícia forense digital para responder os questionários. Todavia, após o envio dos formulários a esses profissionais, não foi obtida nenhuma resposta afirmativa sobre a aceitação de participação na pesquisa. Como segunda opção, outros participantes com conhecimentos na área de segurança da informação e

Tecnologia da Informação foram convidados a participar da pesquisa.

Nesse segundo momento, o desafio consistiu em convencer os participantes a responderem longos formulários de maneira ininterrupta, pois o tempo de resposta de cada formulário submetido seria utilizado como parâmetro importante na pesquisa. A partir daí, dezenas de respostas foram recebidas, o que viabilizou a validação da proposta deste trabalho.

Não obstante ter-se conseguido participantes para a pesquisa, o trabalho de validação prosseguiu mesmo algumas limitações, tais como o número de tecnologias de conectividade nos *logs*, as quais foram IP, LoRaWAN e BLE. Dois fatores foram importantes para a tomada de decisão em não aumentar o número de protocolos na fase de validação. O primeiro foi a consciência de que quanto maior o número de *logs* analisados na pesquisa maior seria o número de formulários, o que teria o potencial de aumentar o número de desistentes da pesquisa. O segundo foi a dificuldade de encontrar *logs* de outras tecnologias, como ocorreu com o LoRaWAN.

Finalmente, vale ressaltar outra limitação importante contida nos dados da pesquisa: o tempo de resposta captado, em alguns casos, pode não corresponder fielmente ao tempo gasto para responder os formulários. Dado que o tempo começa a ser contabilizado no momento em que o formulário é acessado via web, existe a possibilidade de o participante desviar o foco da pesquisa para outras atividades sem relação com o trabalho. Na análise dos dados (Capítulo 5), pôde-se perceber alguns tempos exagerados de respostas.

6.3 Trabalhos Futuros

Durante o desenvolvimento deste projeto de Mestrado, diferentes oportunidades puderam ser identificadas para dar continuidade a esta pesquisa.

Os próximos passos são com certeza continuar melhorando o modelo para cobrir diversas outras tecnologias de conectividade. Uma das abordagens possíveis para isso é propor uma avaliação na indústria, com profissionais experientes na área de investigação forense digital. Ao conduzir tal processo, as sugestões poderiam aprimorar o SBIM em diferentes direções.

Outra iniciativa interessante é usar o SBIM como material didático em cursos de Segurança da Informação e Investigação Forense. Como os alunos geralmente não estão familiarizados com os conceitos e práticas presentes em Investigações Forenses Digitais, um experimento pode ser conduzido para avaliar o conhecimento obtido com o uso do modelo. Por exemplo, um questionário pode ser usado para avaliar o conhecimento após a exploração do modelo durante o

curso.

Com a maior visibilidade do SBIM, ele também poderá ser usado para ajudar pesquisadores e profissionais na construção de novos modelos para diferentes aplicações da Internet das Coisas, orientando-os desde as fases de *Forensic Readiness* e Capacidade Forense até a fase de Apresentação.

Finalmente, o software da camada de Tradução pode tornar-se *open source* para que seja continuamente melhorado e orientado a *plugins*, além da implementação de uma interface gráfica para a ferramenta. Com cada *plugin* representando unicamente o formato de *log* de determinada tecnologia, abre-se a oportunidade de aumentar o número de protocolos suportados pela aplicação.

REFERÊNCIAS

AL-MASRI, E.; BAI, Y.; LI, J. A fog-based digital forensics investigation framework for iot systems. In: IEEE. **2018 IEEE international conference on smart cloud (SmartCloud)**. [*S. l.*], 2018. p. 196–201.

ALABDULSALAM, S.; SCHAEFER, K.; KECHADI, T.; LE-KHAC, N.-A. Internet of things forensics–challenges and a case study. In: SPRINGER. **IFIP International Conference on Digital Forensics**. [*S. l.*], 2018. p. 35–48.

ALAM, M. N.; KABIR, M. S. Forensics in the internet of things: Application specific investigation model, challenges and future directions. In: **2023 4th International Conference for Emerging Technology (INCET)**. [*S. l.*: *s. n.*], 2023. p. 1–6.

ALENEZI, A.; ZULKIPLI, N. H. N.; ATLAM, H. F.; WALTERS, R. J.; WILLS, G. B. The impact of cloud forensic readiness on security. In: **CLOSER**. [*S. l.*: *s. n.*], 2017. p. 511–517.

ALEX, M. E.; KISHORE, R. Forensics framework for cloud computing. **Computers and Electrical Engineering**, Elsevier, v. 60, p. 193–205, 2017.

ALLIANCE, L. **LoRaWAN**. 2015. Disponível em: https://lora-alliance.org/about-lorawan/. Acesso em: 18 dez. 2021.

ALLIANCE, L. **LoRaWAN Specification v1.1**. 2017. Disponível em: https://lora-alliance.org/wp-content/uploads/2020/11/lorawantm_specification_-v1.1.pdf. Acesso em: 18 dez. 2021.

ARSHAD, H.; JANTAN, A. B.; ABIODUN, O. I. Digital forensics: review of issues in scientific validation of digital evidence. **Journal of Information Processing Systems**, Korea Information Processing Society, v. 14, n. 2, p. 346–376, 2018.

ASSURANCE, N. T. A. for I. **Good Practice Forensics Readiness Guideline**. 2015. Disponível em: https://dokumen.tips/documents/good-practice-guide-forensic-readiness-ncsc-site-18aa-forensic-readiness.html. Acesso em: 30 out. 2021.

ATZORI, L.; IERA, A.; MORABITO, G. The internet of things: A survey. **Computer networks**, Elsevier, v. 54, n. 15, p. 2787–2805, 2010.

BABUN, L.; SIKDER, A. K.; ACAR, A.; ULUAGAC, A. S. Iotdots: A digital forensics framework for smart environments. **arXiv preprint arXiv:1809.00745**, 2018.

BAJER, M. Building an iot data hub with elasticsearch, logstash and kibana. In: IEEE. **2017 5th International Conference on Future Internet of Things and Cloud Workshops (FiCloudW)**. [*S. l.*], 2017. p. 63–68.

BAJRAMOVIC, E.; WAEDT, K.; CIRIELLO, A.; GUPTA, D. Forensic readiness of smart buildings: Preconditions for subsequent cybersecurity tests. In: IEEE. **2016 IEEE International Smart Cities Conference (ISC2)**. [*S. l.*], 2016. p. 1–6.

BATOV, E. I. The distinctive features of "smart" buildings. **Procedia Engineering**, Elsevier, v. 111, p. 103–107, 2015.

BV, E. **Elasticsearch reference**. 2018. Disponível em: https://www.elastic.co/guide/en/elasticsearch/reference/6.2/index.html. Acesso em: 16 dez. 2021.

BV, E. **Filebeat reference**. 2018. Disponível em: https://www.elastic.co/guide/en/beats/filebeat/6.2/index.html. Acesso em: 16 dez. 2021.

BV, E. **Logstash reference**. 2018. Disponível em: https://www.elastic.co/guide/en/logstash/6.2/index.html. Acesso em: 16 dez. 2021.

CAO, X.; SHILA, D. M.; CHENG, Y.; YANG, Z.; ZHOU, Y.; CHEN, J. Ghost-in-zigbee: Energy depletion attack on zigbee-based wireless networks. **IEEE Internet of Things Journal**, IEEE, v. 3, n. 5, p. 816–829, 2016.

CAVIGLIONE, L.; WENDZEL, S.; MAZURCZYK, W. The future of digital forensics: Challenges and the road ahead. **IEEE Security & Privacy**, IEEE, v. 15, n. 6, p. 12–17, 2017.

CHAO-YANG, Z. Dos attack analysis and study of new measures to prevent. In: IEEE. **2011 International Conference on Intelligence Science and Information Engineering**. [*S. l.*], 2011. p. 426–429.

CHEN, Y.-J.; CHIEN, H.-Y. Iot-based green house system with splunk data analysis. In: IEEE. **2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST)**. [*S. l.*], 2017. p. 260–263.

CHERNYSHEV, M.; ZEADALLY, S.; BAIG, Z.; WOODWARD, A. Internet of things forensics: The need, process models, and open issues. **IT professional**, IEEE, v. 20, n. 3, p. 40–49, 2018.

CONTI, M.; DEHGHANTANHA, A.; FRANKE, K.; WATSON, S. **Internet of Things security and forensics: Challenges and opportunities**. [*S. l.*]: Elsevier, 2018.

COPPOLINO, L.; D'ALESSANDRO, V.; D'ANTONIO, S.; LEVY, L.; ROMANO, L. My smart home is under attack. In: IEEE. **2015 IEEE 18th International Conference on Computational Science and Engineering**. [*S. l.*], 2015. p. 145–151.

COSIC, J.; COSIC, Z. Chain of custody and life cycle of digital evidence. **Computer technology and application**, David Publishing Company, Inc., v. 3, n. 2, 2012.

CYNTHIA, J.; SULTANA, H. P.; SAROJA, M.; SENTHIL, J. Security protocols for iot. In: **Ubiquitous computing and computing security of IoT**. [*S. l.*]: Springer, 2019. p. 1–28.

DATA, T. **Fluent Bit v1.8 Documentation**. 2018. Disponível em: https://docs.fluentbit.io/manual/. Acesso em: 19 dez. 2021.

DHARUR, S.; SWAMINATHAN, K. Efficient surveillance and monitoring using the elk stack for iot powered smart buildings. In: IEEE. **2018 2nd international conference on inventive systems and control (icisc)**. [*S. l.*], 2018. p. 700–705.

DORAI, G.; HOUSHMAND, S.; BAGGILI, I. I know what you did last summer: Your smart home internet of things and your iphone forensically ratting you out. In: **Proceedings of the 13th International Conference on Availability, Reliability and Security**. [*S. l.*: *s. n.*], 2018. p. 1–10.

ELECTRICAL, I. of; (IEEE), E. E. **Ethernet**. 1983. Disponível em: https://www.ieee802.org/3/. Acesso em: 18 dez. 2021.

ELECTRICAL, I. of; (IEEE), E. E. **WiFi**. 1997. Disponível em: https://www.ieee802.org/11/. Acesso em: 18 dez. 2021.

FOUNDATION, C. N. C. **Fluentd documentation**. 2018. Disponível em: https://docs.fluentd.org/. Acesso em: 19 dez. 2021.

FOUNDATION, L. T. documentation. **LibreOffice Documentation**. 2021. Disponível em: https://documentation.libreoffice.org/en/english-documentation/. Acesso em: 20 dez. 2021.

GERHARDS, R. **The syslog protocol**. 2009. Disponível em: https://datatracker.ietf.org/doc/html/rfc5424. Acesso em: 16 dez. 2021.

GERHARDS, R. **rsyslog 8.36.0 documentation**. 2017. Disponível em: https://www.rsyslog.com/doc/v8-stable/index.html. Acesso em: 16 dez. 2021.

GËRVALLA, M.; PRENIQI, N.; KOPACEK, P. It infrastructure library (itil) framework approach to it governance. **IFAC-PapersOnLine**, Elsevier, v. 51, n. 30, p. 181–185, 2018.

GNU. **GNU grep - GNU Grep: Print lines matching a pattern**. 2021. Disponível em: https://www.gnu.org/software/grep/manual/. Acesso em: 20 dez. 2021.

GNU. **The GNU nano homepage**. 2021. Disponível em: https://www.nano-editor.org/. Acesso em: 20 dez. 2021.

GSMA. **Long Term Evolution for Machines: LTE-M**. 2017. Disponível em: https://www.gsma.com/iot/long-term-evolution-machine-type-communication-lte-mtc-cat-m1/. Acesso em: 19 dez. 2021.

GSMA. **Narrowband Internet of Things**. 2017. Disponível em: https://www.gsma.com/iot/narrow-band-internet-of-things-nb-iot/. Acesso em: 19 dez. 2021.

HAMMOUDI, S.; ALIOUAT, Z.; HAROUS, S. Challenges and research directions for internet of things. **Telecommunication Systems**, Springer, v. 67, n. 2, p. 367–385, 2018.

HARBAWI, M.; VAROL, A. An improved digital evidence acquisition model for the internet of things forensic i: A theoretical framework. In: IEEE. **2017 5th International Symposium on Digital Forensic and Security (ISDFS)**. [*S. l.*], 2017. p. 1–6.

HARBI, Y.; ALIOUAT, Z.; HAROUS, S.; BENTALEB, A.; REFOUFI, A. A review of security in internet of things. **Wireless Personal Communications**, Springer, v. 108, n. 1, p. 325–344, 2019.

HARBI, Y.; ALIOUAT, Z.; REFOUFI, A.; HAROUS, S. Recent security trends in internet of things: A comprehensive survey. **IEEE Access**, IEEE, 2021.

HAXHIBEQIRI, J.; POORTER, E. D.; MOERMAN, I.; HOEBEKE, J. A survey of lorawan for iot: From technology to application. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 18, n. 11, p. 3995, 2018.

HE, P.; ZHU, J.; HE, S.; LI, J.; LYU, M. R. An evaluation study on log parsing and its use in log mining. In: **2016 46th Annual IEEE/IFIP International Conference on Dependable Systems and Networks (DSN)**. [*S. l.*: *s. n.*], 2016. p. 654–661.

HEGARTY, R.; LAMB, D. J.; ATTWOOD, A. *et al.* Digital evidence challenges in the internet of things. In: **INC**. [*S. l.*: *s. n.*], 2014. p. 163–172.

HOSSAIN, M. M.; HASAN, R.; ZAWOAD, S. Probe-iot: A public digital ledger based forensic investigation framework for iot. In: **INFOCOM workshops**. [*S. l.*: *s. n.*], 2018. p. 1–2.

HOSSAIN, M. M.; HASAN, R.; ZAWOAD, S. *et al.* Trust-iov: A trustworthy forensic investigation framework for the internet of vehicles (iov). In: **ICIOT**. [*S. l.*: *s. n.*], 2017. p. 25–32.

HUMMEN, R.; HILLER, J.; WIRTZ, H.; HENZE, M.; SHAFAGH, H.; WEHRLE, K. 6lowpan fragmentation attacks and mitigation mechanisms. In: **Proceedings of the sixth ACM conference on Security and privacy in wireless and mobile networks**. [*S. l.*: *s. n.*], 2013. p. 55–66.

IBRAHIM, N. M.; AL-NEMRAT, A.; JAHANKHANI, H.; BASHROUSH, R. Sufficiency of windows event log as evidence in digital forensics. In: **Global Security, Safety and Sustainability & e-Democracy**. [*S. l.*]: Springer, 2011. p. 253–262.

IEEE; GROUP, T. O. **The open group base specifications issue 7, 2018 edition – IEEE std 1003.1-2017 (revision of IEEE std 1003.1-2008) — crontab**. 2018. Disponível em: https://pubs.opengroup.org/onlinepubs/9699919799/utilities/crontab.html. Acesso em: 16 dez. 2021.

INC., S. **Sigfox story**. 2010. Disponível em: https://www.sigfox.com/en/sigfox-story. Acesso em: 19 dez. 2021.

INC, S. **Splunk documentation**. 2018. Disponível em: https://docs.splunk.com/Documentation. Acesso em: 19 dez. 2021.

INTERNATIONALORGANIZATIONFOR STANDARDIZATION. **ISO/IEC 27005:2018**: Information technology — security techniques — information security risk management. [*S. l.*], 2018.

ISO/IEC. **Guidelines for identification, collection, acquisition and preservation of digital evidence**. 2012. Disponível em: https://www.iso.org/standard/44381.html. Acesso em: 15 nov. 2021.

ISO/IEC. **Governance of digital forensic risk framework**. 2015. Disponível em: https://www.iso.org/standard/53241.html. Acesso em: 15 nov. 2021.

ISO/IEC. **Guidance on assuring suitability and adequacy of incident investigative method**. 2015. Disponível em: https://www.iso.org/standard/44405.html. Acesso em: 15 nov. 2021.

ISO/IEC. **Guidelines for the analysis and interpretation of digital evidence**. 2015. Disponível em: https://www.iso.org/standard/44406.html. Acesso em: 15 nov. 2021.

ISO/IEC. **SO/IEC-27043—Information Technology—Security Techniques—Incident Investigation Principles and Processes**. 2015. Disponível em: https://www.iso.org/standard/44407.html. Acesso em: 30 out. 2021.

ISO/IEC. **Information security incident management — Part 1: Principles of incident management**. 2016. Disponível em: https://www.iso.org/standard/60803.html. Acesso em: 15 nov. 2021.

JAYATHILAKE, D. Towards structured log analysis. In: IEEE. **2012 Ninth International Conference on Computer Science and Software Engineering (JCSSE)**. [*S. l.*], 2012. p. 259–264.

JIA, X.; FENG, Q.; FAN, T.; LEI, Q. Rfid technology and its applications in internet of things (iot). In: IEEE. **2012 2nd international conference on consumer electronics, communications and networks (CECNet)**. [*S. l.*], 2012. p. 1282–1285.

KAVIS, M. J. **Architecting the cloud: design decisions for cloud computing service models (SaaS, PaaS, and IaaS)**. [*S. l.*]: John Wiley & Sons, 2014.

KEBANDE, V. R.; KARIE, N. M.; MICHAEL, A.; MALAPANE, S.; KIGWANA, I.; VENTER, H.; WARIO, R. D. Towards an integrated digital forensic investigation framework for an iot-based ecosystem. In: IEEE. **2018 IEEE International Conference on Smart Internet of Things (SmartIoT)**. [*S. l.*], 2018. p. 93–98.

KEBANDE, V. R.; KARIE, N. M.; VENTER, H. Cloud-centric framework for isolating big data as forensic evidence from iot infrastructures. In: IEEE. **2017 1st International Conference on Next Generation Computing Applications (NextComp)**. [*S. l.*], 2017. p. 54–60.

KEBANDE, V. R.; KARIE, N. M.; VENTER, H. Adding digital forensic readiness as a security component to the iot domain. Indonesian Society for Knowledge and Human Development, 2018.

KEBANDE, V. R.; RAY, I. A generic digital forensic investigation framework for internet of things (iot). In: IEEE. **2016 IEEE 4th International Conference on Future Internet of Things and Cloud (FiCloud)**. [*S. l.*], 2016. p. 356–362.

KEBANDE, V. R.; VENTER, H. S. Novel digital forensic readiness technique in the cloud environment. **Australian Journal of Forensic Sciences**, Taylor & Francis, v. 50, n. 5, p. 552–591, 2018.

KHAUND, K. Cybersecurity in smart buildings inaction is not option any more. **A Frost & Sullivan Collaborative Industry Perspective**, 2015.

KIBIRIGE, G. W.; SANGA, C. A survey on detection of sinkhole attack in wireless sensor network. **arXiv preprint arXiv:1505.01941**, 2015.

KIM, J.; PARK, J.; LEE, S. An improved iot forensic model to identify interconnectivity between things. **Forensic Science International: Digital Investigation**, Elsevier, v. 44, p. 301499, 2023.

KOCAKULAK, M.; BUTUN, I. An overview of wireless sensor networks towards internet of things. In: IEEE. **2017 IEEE 7th annual computing and communication workshop and conference (CCWC)**. [*S. l.*], 2017. p. 1–6.

KUROSE, J. F.; ROSS, K. W. **Computer Networking: A Top-Down Approach**. 7. ed. Boston, MA: Pearson, 2016. ISBN 978-0-13-359414-0.

LABS, L. **Symphony Link**. 2020. Disponível em: https://www.link-labs.com/symphony. Acesso em: 19 dez. 2021.

LALLY, G.; SGANDURRA, D. Towards a framework for testing the security of iot devices consistently. In: SPRINGER. **International workshop on emerging technologies for authorization and authentication**. [*S. l.*], 2018. p. 88–102.

LEE, I.; LEE, K. The internet of things (iot): Applications, investments, and challenges for enterprises. **Business Horizons**, Elsevier, v. 58, n. 4, p. 431–440, 2015.

LEVINE, B. N.; SHIELDS, C.; MARGOLIN, N. B. A survey of solutions to the sybil attack. **University of Massachusetts Amherst, Amherst, MA**, v. 7, p. 224, 2006.

LI, S.; XU, L. D.; ZHAO, S. The internet of things: a survey. **Information Systems Frontiers**, Springer, v. 17, n. 2, p. 243–259, 2015.

LIN, J.; YU, W.; ZHANG, N.; YANG, X.; ZHANG, H.; ZHAO, W. A survey on internet of things: Architecture, enabling technologies, security and privacy, and applications. **IEEE internet of things journal**, IEEE, v. 4, n. 5, p. 1125–1142, 2017.

LINUX.DE. **sfdisk(8) - Linux man page**. 2018. Disponível em: https://linux.die.net/man/8/sfdisk. Acesso em: 19 dez. 2021.

LINUX.DE. **sgdisk(8) - Linux man page**. 2018. Disponível em: https://linux.die.net/man/8/sgdisk. Acesso em: 19 dez. 2021.

LLC, O. I. **syslog-ng open source edition 3.16 – adminis- tration guide**. 2018. Disponível em: https://www.syslog-ng.com/technical-documents/doc/syslog-ng-open-source-edition/3.16/administration-guide. Acesso em: 16 dez. 2021.

LONE, A. H.; MIR, R. N. Forensic-chain: Ethereum blockchain based digital forensics chain of custody. **Sci. Pract. Cyber Secur. J**, v. 1, p. 21–27, 2018.

LONE, A. H.; MIR, R. N. Forensic-chain: Blockchain based digital forensics chain of custody with poc in hyperledger composer. **Digital investigation**, Elsevier, v. 28, p. 44–55, 2019.

LOSAVIO, M. M.; CHOW, K.; KOLTAY, A.; JAMES, J. The internet of things and the smart city: Legal challenges with digital forensics, privacy, and security. **Security and Privacy**, Wiley Online Library, v. 1, n. 3, p. e23, 2018.

MANYIKA, J.; CHUI, M.; BISSON, P.; WOETZEL, J.; DOBBS, R.; BUGHIN, J.; AHARON, D. Unlocking the potential of the internet of things. **McKinsey Global Institute**, v. 1, 2015.

MARTY, R. Cloud application logging for forensics. In: **proceedings of the 2011 ACM Symposium on Applied Computing**. [*S. l.*: *s. n.*], 2011. p. 178–184.

MCKEMMISH, R. When is digital evidence forensically sound? In: SPRINGER. **IFIP international conference on digital forensics**. [*S. l.*], 2008. p. 3–15.

MEFFERT, C.; CLARK, D.; BAGGILI, I.; BREITINGER, F. Forensic state acquisition from internet of things (fsaiot) a general framework and practical approach for iot forensics through iot device state acquisition. In: **Proceedings of the 12th International Conference on Availability, Reliability and Security**. [*S. l.*: *s. n.*], 2017. p. 1–11.

MICROSOFT. **Microsoft Word**. 2021. Disponível em: https://www.microsoft.com/en-us/microsoft-365/word. Acesso em: 20 dez. 2021.

MINERVA, R.; BIRU, A.; ROTONDI, D. Towards a definition of the internet of things (iot). **IEEE Internet Initiative**, IEEE, v. 1, n. 1, p. 1–86, 2015.

MOOLENAAR, B. **VIM REFERENCE MANUAL**. 2021. Disponível em: https://vimhelp.org/intro.txt.html. Acesso em: 20 dez. 2021.

MORGNER, P.; MATTEJAT, S.; BENENSON, Z.; MÜLLER, C.; ARMKNECHT, F. Insecure to the touch: attacking zigbee 3.0 via touchlink commissioning. In: **Proceedings of the 10th ACM Conference on Security and Privacy in Wireless and Mobile Networks**. [*S. l.*: *s. n.*], 2017. p. 230–240.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **Guide to Integrating Forensic Techniques into Incident Response**: Special publication 800-86. [*S. l.*], 2006.

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY. **NIST Cloud Computing Forensic Science Challenges**: Draft nistir 8006. [*S. l.*], 2014.

NIETO, A.; RIOS, R.; LOPEZ, J. A methodology for privacy-aware iot-forensics. In: IEEE. **2017 IEEE Trustcom/BigDataSE/ICESS**. [*S. l.*], 2017. p. 626–633.

NIETO, A.; RIOS, R.; LOPEZ, J. Iot-forensics meets privacy: towards cooperative digital investigations. **Sensors**, Multidisciplinary Digital Publishing Institute, v. 18, n. 2, p. 492, 2018.

NIST. **NIST general information**. 2008. Disponível em: https://www.nist.gov/director/pao/nist-general-information. Acesso em: 8 nov. 2021.

NIST. **Computer Forensics Tool Testing Program (CFTT)**. 2017. Disponível em: https://www.nist.gov/itl/ssd/software-quality-group/computer-forensics-tool-testing-program-cftt/cftt-technical/disk. Acesso em: 18 dez. 2021.

OLINER, A.; GANAPATHI, A.; XU, W. Advances and challenges in log analysis. **Communications of the ACM**, ACM New York, NY, USA, v. 55, n. 2, p. 55–61, 2012.

ORIWOH, E.; JAZANI, D.; EPIPHANIOU, G.; SANT, P. Internet of things forensics: Challenges and approaches. In: IEEE. **9th IEEE International Conference on Collaborative computing: networking, Applications and Worksharing**. [*S. l.*], 2013. p. 608–615.

O'SHAUGHNESSY, S.; KEANE, A. Impact of cloud computing on digital forensic investigations. In: SPRINGER. **Ifip international conference on digital forensics**. [*S. l.*], 2013. p. 291–303.

PASCUCCI, M. Audit log security: How to monitor and protect audit logs. **TechTarget SearchSecurity**, 2013.

PERI, N. **Fluentd vs. Logstash: A comparison of log collectors**. 2015. Disponível em: https://logz.io/blog/fluentd-logstash/. Acesso em: 19 dez. 2021.

PERUMAL, S.; NORWAWI, N. M.; RAMAN, V. Internet of things (iot) digital forensic investigation model: Top-down forensic approach methodology. In: IEEE. **2015 Fifth International Conference on Digital Information Processing and Communications (ICDIPC)**. [*S. l.*], 2015. p. 19–23.

PMI (Ed.). **A Guide to the Project Management Body of Knowledge (PMBOK Guide)**. 5. ed. Newtown Square, PA: Project Management Institute, 2013. ISBN 978-1-935589-67-9.

PROJECT, G. **Network Manager**. 2018. Disponível em: https://networkmanager.dev/. Acesso em: 18 dez. 2021.

PROPRIETARY, B. S. **Bluetooth Core Specification**. 2021. Disponível em: https://www.bluetooth.com/specifications/specs/core-specification/. Acesso em: 19 dez. 2021.

PROPRIETARY, B. S. **Bluetooth overview**. 2021. Disponível em: https://www.bluetooth.com/learn-about-bluetooth/tech-overview/. Acesso em: 19 dez. 2021.

PWC. **Global Economic Crime and Fraud Survey 2018**. 2018. Disponível em: https://www.pwc.com/gx/en/forensics/global-economic-crime-and-fraud-survey-2018.pdf. Acesso em: 26 out. 2021.

QUICK, D.; CHOO, K.-K. R. Pervasive social networking forensics: Intelligence and evidence from mobile device extracts. **Journal of Network and Computer Applications**, Elsevier, v. 86, p. 24–33, 2017.

QUICK, D.; CHOO, K.-K. R. Iot device forensics and data reduction. **IEEE Access**, IEEE, v. 6, p. 47566–47574, 2018.

RAHMAN, K. S.; BISHOP, M.; HOLT, A. Internet of things mobility forensics. **de Proceedings of the 2016 Information Security Research and Education (INSuRE)**, 2016.

RAHMAN, R. A.; SHAH, B. Security analysis of iot protocols: A focus in coap. In: IEEE. **2016 3rd MEC international conference on big data and smart city (ICBDSC)**. [*S. l.*], 2016. p. 1–7.

RANA, N.; SANSANWAL, G.; KHATTER, K.; SINGH, S. Taxonomy of digital forensics: Investigation tools and challenges. **arXiv preprint arXiv:1709.06529**, 2017.

RGHIOUT, A.; KHANNOUS, A.; BOUHORMA, M. Denial-of-service attacks on 6lowpan-rpl networks: Issues and practical solutions. **Journal of Advanced Computer Science & Technology**, v. 3, n. 2, p. 143–153, 2014.

RINNE, T.; YLöNEN, T. **scp(1) – Linux man page**. 2013. Disponível em: https://linux.die.net/man/1/scp. Acesso em: 16 dez. 2021.

RONDEAU, C. M.; TEMPLE, M. A.; LOPEZ, J. Industrial iot cross-layer forensic investigation. **Wiley Interdisciplinary Reviews: Forensic Science**, Wiley Online Library, v. 1, n. 1, p. e1322, 2019.

SAADEH, M.; SLEIT, A.; QATAWNEH, M.; ALMOBAIDEEN, W. Authentication techniques for the internet of things: A survey. In: IEEE. **2016 cybersecurity and cyberforensics conference (CCC)**. [*S. l.*], 2016. p. 28–34.

SADINENI, L.; PILLI, E.; BATTULA, R. B. A holistic forensic model for the internet of things. In: SPRINGER. **IFIP International Conference on Digital Forensics**. [*S. l.*], 2019. p. 3–18.

SECURITY, H. **Industrial Control Systems Cybersecurity Emergency Response Team (ICS-CERT) Monitor**. 2015.

SERVIDA, F.; CASEY, E. Iot forensic challenges and opportunities for digital traces. **Digital Investigation**, Elsevier, v. 28, p. S22–S29, 2019.

SILVA, I.; LEANDRO, R.; MACEDO, D.; GUEDES, L. A. A dependability evaluation tool for the internet of things. **Computers & Electrical Engineering**, v. 39, n. 7, p. 2005–2018, 2013. ISSN 0045-7906. Disponível em: https://www.sciencedirect.com/science/article/pii/S0045790613001171.

SINGH, M.; RAJAN, M.; SHIVRAJ, V.; BALAMURALIDHAR, P. Secure mqtt for internet of things (iot). In: IEEE. **2015 fifth international conference on communication systems and network technologies**. [*S. l.*], 2015. p. 746–751.

SNOONIAN, D. Smart buildings. **IEEE spectrum**, IEEE, v. 40, n. 8, p. 18–23, 2003.

SOMMER, P. **Digital Evidence, Digital Investigations and E-Disclosure: A Guide to Forensic Readiness for Organizations, Security Advisers and Lawyers**. 2012.

SPECIFICATIONS, T. O. G. B. **DD documentation**. 2018. Disponível em: https://pubs.opengroup.org/onlinepubs/9699919799/utilities/dd.html. Acesso em: 19 dez. 2021.

STOYANOV, Y. An approach to use the web services and open source software to store and share user applications and data. In: **Proc. Annu. Univ. Sci. Conf. NVU**. [*S. l.*: *s. n.*], 2014. v. 9, p. 92–96.

STOYANOVA, M.; NIKOLOUDAKIS, Y.; PANAGIOTAKIS, S.; PALLIS, E.; MARKAKIS, E. K. A survey on the internet of things (iot) forensics: challenges, approaches, and open issues. **IEEE Communications Surveys & Tutorials**, IEEE, v. 22, n. 2, p. 1191–1221, 2020.

SYMANTEC. **Internet Security Threat Report (ISTR): Volume 23**. 2018. Disponível em: https://www.symantec.com/content/dam/symantec/docs/reports/istr-23-2018-en.pdf. Acesso em: 26 out. 2021.

Sá, A. **Eficiência, eficácia e efetividade – material teórico**. 2020. Disponível em: https://www.tecconcursos.com.br/blog/eficiencia-eficacia-e-efetividade-material-teorico/. Acesso em: 03 jan. 2022.

TRENWITH, P. M.; VENTER, H. S. Digital forensic readiness in the cloud. In: IEEE. **2013 Information Security for South Africa**. [*S. l.*], 2013. p. 1–5.

TRIDGELL, P. M. e. W. D. A. **rsync(1)**. 2018. Disponível em: https://download.samba.org/pub/rsync/rsync.html. Acesso em: 16 dez. 2021.

VAINIO, A. **Implementation of Centralized Logging and Log Analysis in Cloud Transition.** 2018. Dissertação (Master of Science in Technology) – School of science, Aalto University, Otaniemi, 2018.

VEGA, C.; ROQUERO, P.; LEIRA, R.; GONZALEZ, I.; ARACIL, J. Loginson: a transform and load system for very large-scale log analysis in large it infrastructures. **The Journal of Supercomputing**, Springer, v. 73, n. 9, p. 3879–3900, 2017.

WANG, K.; DU, M.; SUN, Y.; VINEL, A.; ZHANG, Y. Attack detection and distributed forensics in machine-to-machine networks. **IEEE Network**, IEEE, v. 30, n. 6, p. 49–55, 2016.

YAKUBU, O.; ADJEI, O.; NARENDRA, B. C. A review of prospects and challenges of internet of things. **International Journal of Computer Applications**, Foundation of Computer Science, v. 139, n. 10, p. 33–39, 2016.

YAQOOB, I.; HASHEM, I. A. T.; AHMED, A.; KAZMI, S. A.; HONG, C. S. Internet of things forensics: Recent advances, taxonomy, requirements, and open challenges. **Future Generation Computer Systems**, Elsevier, v. 92, p. 265–275, 2019.

ZAWOAD, S.; HASAN, R. Faiot: Towards building a forensics aware eco system for the internet of things. In: IEEE. **2015 IEEE International Conference on Services Computing**. [*S. l.*], 2015. p. 279–284.

ZIA, T.; LIU, P.; HAN, W. Application-specific digital forensics investigative model in internet of things (iot). In: **Proceedings of the 12th International Conference on Availability, Reliability and Security**. [*S. l.*: *s. n.*], 2017. p. 1–7.

ZULKIPLI, N. H. N.; ALENEZI, A.; WILLS, G. B. Iot forensic: bridging the challenges in digital forensic and the internet of things. In: SCITEPRESS. **International Conference on Internet of Things, Big Data and Security**. [*S. l.*], 2017. v. 2, p. 315–324.

APÊNDICE A – FORMULÁRIO 1

Questão 1. Que horas o dispositivo com IP 192.168.1.9 conectou-se pela última vez?

Questão 2. O dispositivo com IP 10.0.0.108 passou quanto tempo conectado na rede no dia 9 de Março?

Questão 3. Que horas o dispositivo com IP 192.168.1.3 conectou-se à rede?

Questão 4. Qual dispositivo conectou-se à rede ’iPhone’?

Questão 5. O IP 10.0.0.107 conectou-se primeiro à rede ’Oi 31EC’ do que o IP 10.0.0.108?

APÊNDICE B – FORMULÁRIO 2

Questão 1. Que horas o dispositivo com o endereço local ’48:2c:a0:ba:15:35’ conectou-se à rede pela primeira vez?

Questão 2. Quando o dispositivo ’30:c0:1b:7e:b1:a0’ desconectou-se pela primeira vez?

Questão 3. Quando o dispositivo ’48:2c:a0:ba:15:35’ desconectou-se pela segunda vez?

Questão 4. Quando o dispositivo ’30:c0:1b:7e: b1:a0” conectou-se pela penúltima vez?

Questão 5. A primeira desconexão entre os dispositivos ’48:2c:a0’ e ’30:c0:1b:7e’ ocorreu no dia 10 de dezembro.

APÊNDICE C – FORMULÁRIO 3

Questão 1. Que horas o dispositivo com o endereço local cee11d9f-81e6-a5ab1f1c conectou-se à rede?

Questão 2. Qual o endereço local do dispositivo com endereço global ’111435912f75ff55’?

Questão 3. Que dia, mês e ano o dispositivo ’dcba2ff14e8bd9e5’ conectou-se à rede?

Questão 4. O dispositivo ’bd9d862e4d6c51b1’ conectou-se alguma vez à rede?

Questão 5. O dispositivo ’4e19b3cfe1b39e36’ conectou-se primeiro à rede em comparação ao dispositivo ’16ac427a8ef9e2c3’?

APÊNDICE D – FORMULÁRIO 4

Questão 1. Que horas o dispositivo com IP 192.168.1.9 conectou-se pela última vez?

Questão 2. O dispositivo com IP 10.0.0.108 passou quanto tempo conectado na rede no dia 9 de Março?

Questão 3. Que horas o dispositivo com IP 192.168.1.3 conectou-se à rede?

Questão 4. O dispositivo IP 192.168.1.3 conectou-se alguma vez à rede 'Oi 31EC'?

Questão 5. O IP 10.0.0.107 conectou-se primeiro à rede 'Asgard' do que o IP 10.0.0.105?

APÊNDICE E – FORMULÁRIO 5

Questão 1. Que horas o dispositivo com o endereço local ’48:2c:a0:ba:15:35’ conectou-se à rede pela primeira vez?

Questão 2. Que horas o dispositivo ’30:c0:1b:7e:b1:a0’ desconectou-se pela primeira vez?

Questão 3. Que horas o dispositivo ’48:2c:a0:ba:15:35’ desconectou-se pela segunda vez?

Questão 4. Que horas o dispositivo ’30:c0:1b:7e: b1:a0” conectou-se pela penúltima vez?

Questão 5. A primeira desconexão entre os dispositivos ’48:2c:a0:ba:15:35’ e ’30:c0:1b:7e:b1’ ocorreu no dia 9 de dezembro.

APÊNDICE F – FORMULÁRIO 6

Questão 1. dispositivo com o endereço local ’cee11d9f-81e6’ conectou-se à rede?

Questão 2. Qual o endereço local do dispositivo com endereço global ’111435912f75ff55’?

Questão 3. Que dia, mês e ano o dispositivo ’dcba2ff14e8bd9e5’ conectou-se à rede?

Questão 4. O dispositivo ’bd9d862e4d6c55b1’ conectou-se alguma vez à rede?

Questão 5. O dispositivo ’4e19b3cfe1b39e36’ conectou-se primeiro à rede em comparação ao dispositivo ’16ac427a8ef9e2c3’?

APÊNDICE G – FORMULÁRIO 7

Questão 1. Que horas o dispositivo com IP 192.168.1.9 conectou-se pela última vez?

Questão 2. O dispositivo com IP 10.0.0.108 passou quanto tempo conectado na rede no dia 9 de Março?

Questão 3. Que horas o dispositivo com IP 192.168.1.127 conectou-se à rede?

Questão 4. Quais dispositivos conectaram-se à rede ’Oi 31EC’?

Questão 5. O IP 10.0.0.107 conectou-se primeiro à rede ’Asgard’ do que o IP 10.0.0.105.

APÊNDICE H – FORMULÁRIO 8

Questão 1. Que horas o dispositivo com o endereço local ’48:2c:a0:ba:15:35’ conectou-se à rede pela primeira vez?

Questão 2. Quando o dispositivo ’30:c0:1b:7e:b1:a0’ desconectou-se pela primeira vez?

Questão 3. Quando o dispositivo ’48:2c:a0:ba:15:35’ desconectou-se pela segunda vez?

Questão 4. Quando o dispositivo ’30:c0:1b:7e: b1:a0” conectou-se pela penúltima vez?

Questão 5. A última desconexão ocorreu às 15:40.

APÊNDICE I – FORMULÁRIO 9

Questão 1. Que horas o dispositivo com o endereço local ’cee11d9f-81e6-a5ab1f1c-51c1ca13777e’ conectou-se à rede? primeira vez?

Questão 2. Qual o endereço local do dispositivo com endereço global ’111435912f75ff55’?

Questão 3. Que dia, mês e ano o dispositivo ’dcba2ff14e8bd9e5’ conectou-se à rede?

Questão 4. O dispositivo ’ad51a647d8fce724’ conectou-se alguma vez à rede?

Questão 5. O dispositivo ’bd9d862e4d6c51b1’ conectou-se primeiro à rede em comparação ao dispositivo ’5386d42d-8ae5-633f-6f7c173c84c89db8’?

APÊNDICE J – FORMULÁRIO 10

Questão 1. Quando o dispositivo com IP 192.168.1.9 conectou-se pela última vez? primeira vez?

Questão 2. O dispositivo com IP 10.0.0.108 passou quanto tempo conectado na rede no dia 9 de Março?

Questão 3. Que horas o dispositivo com IP 192.168.1.3 conectou-se à rede?

Questão 4. O dispositivo IP 192.168.1.3 conectou-se alguma vez à rede Tenda0DDE70?

Questão 5. No dia 8 de janeiro, o dispositivo 192.168.1.127 conectou-se à rede Tenda0DDE70, pouco depois das 15h

APÊNDICE K – FORMULÁRIO 11

Questão 1. Que horas o dispositivo com o endereço local ’48:2c:a0:ba:15:35’ conectou-se à rede pela primeira vez?

Questão 2. Que horas o dispositivo ’30:c0:1b:7e:b1:a0’ desconectou-se pela primeira vez?

Questão 3. Que horas o dispositivo ’48:2c:a0:ba:15:35’ desconectou-se pela segunda vez?

Questão 4. Que horas o dispositivo ’30:c0:1b:7e: b1:a0” conectou-se pela penúltima vez?

Questão 5. Qual o dia e hora da primeira desconexão entre os dispositivos ’48:2c:a0:ba:15:35’ e ’30:c0:1b:7e:b1:a0’?

APÊNDICE L – FORMULÁRIO 12

Questão 1. Que horas o dispositivo com o endereço local ’cee11d9f-81e6-a5ab1f1c-51c1ca13777e’ conectou-se à rede

Questão 2. Qual o endereço local do dispositivo com endereço global ’111435912f75ff55’?

Questão 3. Que dia, mês e ano o dispositivo ’dcba2ff14e8bd9e5’ conectou-se à rede?

Questão 4. O dispositivo ’bd9d862e4d6c51b1’ conectou-se alguma vez à rede?

Questão 5. O dispositivo ’4e19b3cfe1b39e37’ conectou-se primeiro à rede em comparação ao dispositivo ’16ac427a8ef9e2c3’?

APÊNDICE M – FORMULÁRIO 13

Questão 1. As principais informações sobre o momento de conexão e desconexão dos dispositivos na rede foram claramente apresentadas pelos arquivos "bluetooth.csv", "ip.csv"e "lorawan.csv". (Por favor, se não concorda, marque "outro"e forneça os motivos)

(a) Não concordo fortemente
(b) Não concordo
(c) Não tenho certeza
(d) Concordo
(e) Concordo fortemente
(f) Outra

Questão 2. As informações contidas nos arquivos "bluetooth.csv", "ip.csv"e "lorawan.csv"são úteis para a investigação de incidentes em uma organização. (Por favor, se não concorda, marque "outro"e forneça os motivos)

(a) Não concordo fortemente
(b) Não concordo
(c) Não tenho certeza
(d) Concordo
(e) Concordo fortemente
(f) Outra

Questão 3. A estrutura sintática e semântica dos arquivos "bluetooth.csv", "ip.csv"e "lorawan.csv"é bem mais fácil de ser aprendida do que a dos arquivos "logip.txt", "lorawan.json", "bluetooth.cfa"e "bluetooth.csv". (Por favor, se não concorda, marque "outro"e forneça os motivos)

(a) Não concordo fortemente
(b) Não concordo
(c) Não tenho certeza
(d) Concordo
(e) Concordo fortemente
(f) Outra

Questão 4. Na sua opinião, um formato de log que siga algum padrão prédefinido é necessário?

(a) sim
(b) não

Questão 5. Você consideraria utilizar a representação padrão de logs aqui apresentada (nos arquivos bluetooth.csv, ip.csv e lorawan.csv) em futura investigação de incidentes?

(a) sim

(b) não

Questão 6. Por favor, use este espaço para comentários adicionais ou sugestões a respeito do padrão de logs aqui apresentado.

APÊNDICE N – FORMULÁRIO 14

Questão 1. Indique sua(s) área(s) de atuação/experiência:

(a) Infraestrutura

(b) Administração de Sistemas

(c) Investigação forense

(d) Outra

Questão 2. Anos de experiência na área:

(a) 0 a 1

(b) 1 a 3

(c) 3 a 5

(d) Mais que 5 anos

Questão 3. Você tem experiência com ferramentas de processamento de texto, tais como: grep, sed ou awk?

(a) Sim

(b) Não

Questão 4. Qual Sistema Operacional você mais utiliza?

(a) GNU/Linux

(b) Windows

(c) MacOS

(d) Outro