UMA NOVA ABORDAGEM DE DOIS NÍVEIS PARA OBTENÇÃO DE MÉTRICAS DE CONFIANÇA E AVALIAÇÃO DA SEGURANÇA NA COMUNICAÇÃO ENTRE DISPOSITIVOS DA INTERNET DAS COISAS

Evandro Luiz Cardoso Macedo

Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia de Sistemas e Computação, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia de Sistemas e Computação.

Orientador: Luís Felipe Magalhães de Moraes

Rio de Janeiro
Março de 2022

UMA NOVA ABORDAGEM DE DOIS NÍVEIS PARA OBTENÇÃO DE MÉTRICAS DE CONFIANÇA E AVALIAÇÃO DA SEGURANÇA NA COMUNICAÇÃO ENTRE DISPOSITIVOS DA INTERNET DAS COISAS

Evandro Luiz Cardoso Macedo

TESE SUBMETIDA AO CORPO DOCENTE DO INSTITUTO ALBERTO LUIZ COIMBRA DE PÓS-GRADUAÇÃO E PESQUISA DE ENGENHARIA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE DOUTOR EM CIÊNCIAS EM ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Orientador: Luís Felipe Magalhães de Moraes

Aprovada por: Prof. Luís Felipe Magalhães de Moraes
Prof. Claudio Luis de Amorim
Prof. Felipe Maia Galvão França
Prof. Flávia Coimbra Delicato
Prof. Nilton Alves Júnior
Prof. Reginaldo Palazzo Júnior

RIO DE JANEIRO, RJ – BRASIL
MARÇO DE 2022

Macedo, Evandro Luiz Cardoso

Uma Nova Abordagem de Dois Níveis para Obtenção de Métricas de Confiança e Avaliação da Segurança na Comunicação entre Dispositivos da Internet das Coisas/Evandro Luiz Cardoso Macedo. – Rio de Janeiro: UFRJ/COPPE, 2022.

XIV, 98 p.: il.; 29, 7cm.

Orientador: Luís Felipe Magalhães de Moraes

Tese (doutorado) – UFRJ/COPPE/Programa de Engenharia de Sistemas e Computação, 2022.

Referências Bibliográficas: p. 56 – 68.

1. Internet das Coisas. 2. segurança. 3. confiança. 4. entropia. 5. cadeia de blocos. I. Moraes, Luís Felipe Magalhães de. II. Universidade Federal do Rio de Janeiro, COPPE, Programa de Engenharia de Sistemas e Computação. III. Título.

"Feliz aquele que transfere o que
sabe e aprende o que ensina."
Cora Coralina

Agradecimentos

Não há outra forma de começar a agradecer senão por Aquele que sempre faz o impossível por mim, o meu Melhor Amigo, meu Deus, e seu filho Jesus Cristo, mesmo eu não sendo digno por todas as minhas falhas. A Ele toda a glória.

Obrigado a minha amada esposa Tais, pela paciência nos momentos de ausência, pelo apoio, amor e cuidado, e também à minha enteada, Paula, que "veio no pacote" para iluminar minha vida e trazer alegrias. Aos meus pais, Duilio e Maria Bernadete, exemplos para mim e me formaram com valiosos princípios. Se eu cheguei até aqui foi por vocês. E a minha irmã Carla, pela amizade e pelo carinho. Amo todos vocês.

Agradeço imensamente ao meu orientador Prof. Luís Felipe, por todas as orientações, ideias e discussões que contribuíram para que este trabalho se tornasse realidade. Obrigado, não só por me fazer evoluir como pesquisador no meio acadêmico, mas também por me ajudar a evoluir como pessoa. Obrigado por me permitir fazer parte da família do Laboratório Ravel, pelas oportunidades que construiu e continua construindo para me ajudar e pela amizade.

Aos membros da banca, professores Claudio Amorim, Felipe França, Flávia Delicato, Nilton Alves e Reginaldo Palazzo, meus agradecimentos por aceitarem o convite para avaliar nosso trabalho e pelas contribuições significativas para melhorá-lo.

Agradeço em especial à Profa. Flávia, por toda ajuda e contribuição ao longo do doutorado, pelas parcerias e colaborações nos projetos e produções de artigos. A senhora tem a minha admiração e gratidão. Que possamos colaborar sempre.

A very special thanks to Prof. Giancarlo Fortino for receiving me at SensysCal Lab, and all the support and guidance during my stay at the University of Calabria in Italy. I hope to see you again and keep on collaborating. Un abbraccio!

Aos amigos de jornada de doutorado e do Ravel, obrigado pelo apoio, pelas ideias, conversas e momentos de descontração. Ganhei amigos e irmãos.

Ao PESC/COPPE/UFRJ e seus funcionários por me receber todos esses anos e batalhar para que os alunos tenham infraestrutura e ensino de excelência.

À equipe da Rede-Rio/FAPERJ, pelo apoio e por também me acolher e permitir que eu contribua para o projeto. É um privilégio poder colaborar com essa equipe.

Agradeço à CAPES, ao CNPq, à FAPERJ, à DELL EMC e à UFF, pelo apoio financeiro essencial.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Doutor em Ciências (D.Sc.)

UMA NOVA ABORDAGEM DE DOIS NÍVEIS PARA OBTENÇÃO DE MÉTRICAS DE CONFIANÇA E AVALIAÇÃO DA SEGURANÇA NA COMUNICAÇÃO ENTRE DISPOSITIVOS DA INTERNET DAS COISAS

Evandro Luiz Cardoso Macedo

Março/2022

Orientador: Luís Felipe Magalhães de Moraes

Programa: Engenharia de Sistemas e Computação

A Internet das Coisas (*Internet of Things – IoT*) traz uma nova onda de evolução da Internet, habilitando a comunicação de novos dispositivos, tornando-os inteligentes. IoT possibilita materializar conceitos como Sistemas Ciber-Físicos (*Cyber-Physical Systems – CPS*) e permite o surgimento de novos sistemas e aplicações. Entretanto, a disseminação massiva e onipresente de dispositivos IoT interconectados expõe cada vez mais as vulnerabilidades dos dados e das aplicações relacionadas. Se a segurança de qualquer componente de tais sistemas for comprometida, um vazamento de dados associado pode causar sérias ameaças à privacidade, perdas materiais e até mesmo colocar a vida das pessoas em risco. Desta forma, estudos sobre os aspectos de segurança da IoT têm se tornado cada vez mais importantes. Esta tese apresenta uma proposta de atribuição e obtenção de métricas de confiança na comunicação entre dispositivos IoT, de modo a atender às questões de pesquisa relacionadas aos aspectos de segurança neste contexto. A ideia principal consiste em uma abordagem de dois níveis que considera características de aplicação e de rede. Em particular, a confiança em um dispositivo é modelada combinando uma medida de entropia relativa da taxa de dados deste, bem como sua respectiva reputação a partir de um registro distribuído, considerando também o uso de Criptografia Baseada em Identidades (*Identity-Based Encryption – IBE*) para identificação dos dispositivos. Através de resultados numéricos, mostramos a eficácia da nossa abordagem em isolar dispositivos anômalos/não confiáveis com base na métrica de confiança proposta.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Doctor of Science (D.Sc.)

A NOVEL TWO-LEVEL APPROACH TO OBTAINING TRUST METRICS AND ASSESSING SECURITY IN COMMUNICATION BETWEEN INTERNET OF THINGS DEVICES

Evandro Luiz Cardoso Macedo

March/2022

Advisor: Luís Felipe Magalhães de Moraes

Department: Systems Engineering and Computer Science

The Internet of Things (IoT) brings a new wave of Internet evolution, enabling the communication of new devices, making them smart. IoT allows materializing concepts such as Cyber-Physical Systems (CPS) and permits the emergence of new systems and applications. However, the massive and ubiquitous spread of interconnected IoT devices increasingly exposes the vulnerability of data and related applications. If the security of any component in such systems gets compromised, an associated data leak may cause serious threats to privacy, material loss, and even put people's lives at risk. Therefore, studies on IoT security aspects have become increasingly important. This thesis presents a proposal to defining and obtaining trust metrics in communication between IoT devices, in order to address research questions related to security aspects in this context. The key idea consists of a two-level approach that considers application and network characteristics. In particular, trust is modeled by combining a relative entropy measure of a device data rate and the respective reputation from a distributed-ledger, considering also the use of Identity-Based Encryption (IBE) for devices identification. Through numerical results, we show the effectiveness of our approach in isolating anomalous/untrusted devices based on the proposed trust metric.

Sumário

Lista de Figuras

Lista de Tabelas

Lista de Siglas

CAGR	Compound Annual Growth Rate, p. 2
DDoS	Distributed Denial of Service, p. 3
DNS	Domain Name System, p. 3
DoS	Denial of Service, p. 47
IBE	Identity-Based Encryption, p. 14
IDS	Intrusion Detection System, p. 12
IP	Internet Protocol, p. 3
IoT	Internet of Things, p. 1
PKG	Private-Key Generator, p. 7
RFID	Radio-Frequency Identification, p. 2
RPL	Routing Protocol for Low Power and Lossy Networks, p. 14
RSA	Rivest-Shamir-Adleman, p. 73
SDN	Software-Defined Networks, p. 12
SIoT	Social IoT, p. 2
SMTP	Simple Mail Transfer Protocol, p. 14
TLS	Transport Layer Security, p. 72
WPAN	Wireless Personal Area Networks, p. 2

Capítulo 1

Introdução

Este capítulo apresenta a contextualização e motivação para o trabalho desenvolvido nesta tese. Além disso, apresentamos também a definição do problema de confiança abordado, o estado da arte das pesquisas envolvendo o tema de Internet das Coisas, as questões de pesquisa ainda em aberto, bem como as inovações e contribuições trazidas pela tese e a organização da mesma.

1.1 Contexto e Motivação

O uso massivo da Internet ao longo dos últimos anos vem pervadindo diversas áreas do cotidiano humano, seja pelas interações nas redes sociais, compras *online*, consumo de conteúdos de entretenimento, entre outros serviços. As evoluções consistentes da Internet trazem novas perspectivas e aplicações que outrora não seriam sequer imaginadas.

Iniciando em 1969, a ARPANET [1] deu início à Internet através de um projeto do Departamento de Defesa dos Estados Unidos, viabilizado por pesquisas científicas de grandes nomes da literatura, como Leonard Kleinrock [2], pioneiro no desenvolvimento da teoria matemática para as redes de pacotes de dados, conhecido como o "pai da comutação por pacotes e da Internet"; Claude Shannon [3], que formulou e quantificou o conceito fundamental de informação, conhecido como o "pai da Teoria da Informação", dentre outros pesquisadores renomados. Conforme as redes foram ganhando escala e se interconectando cada vez mais, surge a rede de redes, a Internet, permitindo a troca de informações entre computadores geograficamente distribuídos pelo mundo.

Como uma nova onda de evolução da Internet, a Internet das Coisas (*Internet of Things – IoT*) [4] estabelece um novo paradigma que vislumbra conectividade, não apenas entre computadores, mas entre quaisquer objetos físicos, potencialmente instrumentados com sensores e atuadores, conectados majoritariamente sem fio [5]. A IoT permite que sistemas heterogêneos sejam capazes de se comunicar e intero-

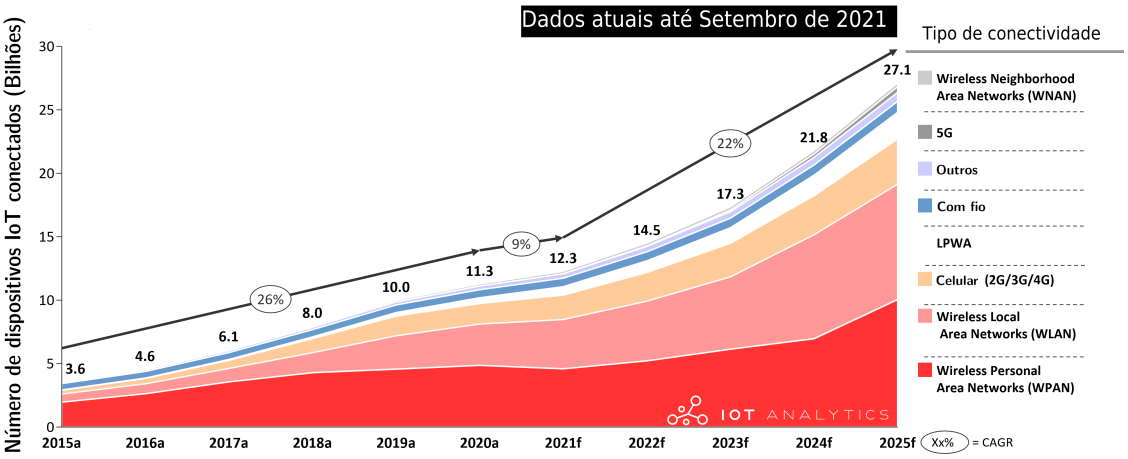


Figura 1.1: Crescimento estimado do número de dispositivos conectados [6]

perar de forma transparente, ubíqua e ininterrupta. O termo IoT surgiu em 1999, cunhado por Kelvin Ashton durante pesquisas envolvendo o uso de etiquetas eletrônicas através da tecnologia *Radio-Frequency Identification (RFID)* em produtos da cadeia de fornecimento, ficando o seu uso marcado para a história.

A IoT também apresenta outras especializações, ou ainda sub-paradigmas, que vêm se consolidando ao longo dos últimos anos, tais como, a IoT Médica, IoT Industrial, Internet de Todas as Coisas, IoT Social (SIoT) , entre outras. Em particular, no contexto da SIoT, o aspecto de confiança é extremamente relevante, visto que tal especialização traz um novo paradigma que habilita os dispositivos IoT poderem interagir entre si de maneira autonômica, sem intervenção humana. Através de SIoT os dispositivos podem estabelecer relações sociais, como, por exemplo, relações de amizade ou de negócios, entre outras. Abordamos mais detalhes sobre tal paradigma no Apêndice C.

Com uma estimativa de cerca de 27 bilhões de dispositivos conectados até 2025 [6], impulsionada pelo advento da quinta geração de redes móveis (5G) [7], a disseminação de IoT abre caminho para uma miríade de aplicações que podem impactar significativamente o modo de vida da sociedade atual, tais como, por exemplo, cidades inteligentes, aplicações de saúde, sistemas de transporte inteligente, IoT industrial, entre outros [8]. A Figura 1.1 ilustra a projeção do crescimento do número de dispositivos IoT conectados. Podemos perceber que, atualmente, já temos uma quantidade significativa de dispositivos somando cerca de 12 bilhões de conexões, sendo em maior proporção as conexões de dispositivos sem fio pessoais (*Wireless Personal Area Networks – WPAN*).

Além disso, IoT permite uma aquisição de dados sem precedentes, o que pode proporcionar melhorias para diversos processos de tomada de decisão. No entanto, tais benefícios também implicam na tarefa de fornecer segurança a cada um dos dispositivos envolvidos. Se a segurança de algum componente dos sistemas citados

for comprometida, um vazamento de dados pode ocorrer, causando sérias ameaças à privacidade, perdas materiais, ou até mesmo colocar vidas em risco. Por exemplo, um sensor adulterado pode expor dados privados ou fornecer medidas erradas sobre a frequência cardíaca de um paciente em aplicações de saúde, levando a erros de prescrição. Uma casa inteligente equipada com controle de dispositivos (a aplicação IoT mais comum atualmente [9]), pode ser comprometida e, em seguida, começar a ligar e desligar os dispositivos intermitentemente, o que pode queimar componentes individuais ou até causar curto-circuito, levando a consequências desastrosas. Outro exemplo seria um controle adulterado de um carro autônomo em um sistema de transporte inteligente, que pode causar sérios acidentes de trânsito.

Conforme o número de dispositivos IoT vem aumentando massivamente, a responsabilidade e os desafios em termos dos aspectos de segurança também aumentam. Por exemplo, o ataque de negação de serviço distribuído (*Distributed Denial of Service – DDoS*) relatado em [10], causou problemas a um grande provedor de serviços de DNS e afetou a disponibilidade de vários *sites* relacionados a diversas aplicações, como Twitter, Reddit, Spotify, PayPal, entre outros. Um aspecto de destaque é que os atacantes se valeram de dispositivos IoT para realizar tal ataque através do uso da *botnet* Mirai, composta por dispositivos como câmeras IP e roteadores, que apresentavam algum tipo de vulnerabilidade (*e.g.* senhas padrão, implementação de código inseguro, entre outros). Assim, apesar de promover diversos benefícios, IoT também expande a superfície dos ataques cibernéticos, aumentando o potencial e a eficácia de tais ataques. Neste contexto, não resta dúvida de que fornecer segurança aos sistemas IoT é de grande importância. Novos requisitos e desafios precisam ser considerados nos projetos e desenvolvimentos de sistemas e aplicações IoT [11], especialmente no que tange à segurança [12–14].

Em particular, a atribuição de valores de confiança (*trust*) na comunicação entre dispositivos IoT é de suma importância, visto que, nos sistemas IoT, a comunicação se baseia consideravelmente entre dispositivos (*machine-to-machine*), sendo extremamente importante e desafiador definir métricas para que as máquinas estabeleçam confiança entre si e possam operar de maneira autônoma. Como exemplo para ilustrar a importância de se ter uma métrica de confiança, podemos considerar uma aplicação de SIoT na qual *drones* inteligentes interagem entre si através de relações colaborativas (por exemplo, amizade ou parceria). Neste caso, uma métrica de confiança é essencial para o sucesso da aplicação, visto que a falta de confiança inviabilizaria tal aplicação. Além disso, o gerenciamento dedicado de cada dispositivo, aspecto relevante para segurança em redes, é impraticável, dada a quantidade e diversidade de tais elementos. Assim, este ainda é considerado um problema em aberto pela comunidade científica [13, 15–25].

Segundo Yan *et al.* [17] e Arabsorkhi *et al.* [26], estabelecer confiança entre

dispositivos IoT engloba aspectos fundamentais como crença, integridade, confiabilidade, disponibilidade, entre outros. Em particular, existem dificuldades que ainda prevalecem em aplicações de IoT, como a falta de informações sobre gerenciamento dos dispositivos, sobre como dados pessoais são armazenados, de que modo a privacidade é mantida e como os sistemas podem identificar vulnerabilidades de segurança em seus componentes [9, 27, 28]. Todos esses problemas dificultam a adoção de IoT e impactam em sua confiança. Especificamente, há uma variedade de definições para o conceito de confiança (conforme descrito em [17, 26, 29, 30]), herdadas do conceito de confiança do contexto humano para o contexto dos dispositivos IoT. De fato, confiança é um conceito complexo, ainda sem consenso na comunidade acadêmica. Assim, para termos uma definição da confiança, com base nos trabalhos encontrados na literatura [17, 26], consideramos que tal conceito deve compreender ao menos quatro aspectos no contexto dos dispositivos IoT: a garantia da identidade, o comportamento de rede, a integridade de dados e proteção dos dispositivos contra ataques. Como exemplo, podemos imaginar o cenário anteriormente citado do ataque ao provedor de DNS, se os dispositivos utilizassem um mecanismo de controle de confiança, tal ataque poderia ser evitado, visto que os dispositivos maliciosos perderiam a confiança dos outros elementos da rede por conta do comportamento de rede inadequado (um dos aspectos do conceito de confiança). Desta maneira, os próprios membros da rede a protegeriam ao isolar os dispositivos mal-intencionados (outro aspecto do conceito de confiança considerado).

Nesta tese, apresentamos e modelamos uma proposta de atribuição de confiança na comunicação entre dispositivos IoT para provisionar segurança aos sistemas de IoT. Nossa proposta é baseada em uma abordagem de dois níveis, que reúne dados das camadas de rede e de aplicação, respectivamente para o Nível Baixo (*Low Level*) e o Nível Alto (*High Level*) de nossa proposta. Nossa principal contribuição consiste em combinar as características do Nível Baixo (perspectiva de rede) usando a entropia relativa do fluxo de entrada de dados de um dispositivo IoT; e do Nível Alto (perspectiva de aplicação) com uma abordagem baseada em livro-razão distribuído (cadeia de blocos) para fornecer a reputação das identidades dos dispositivos. Desta maneira, temos o objetivo de quantificar a confiança e compor uma métrica abrangente, capaz de capturar mudanças no comportamento do tráfego dos dispositivos e isolar aqueles que apresentam comportamento inesperado. Até onde sabemos, nossa proposta é a primeira a prover uma métrica de confiança que considera características de rede e de aplicação em conjunto.

1.2 Definição do Problema de Confiança

Lidar com os aspectos de segurança é um grande desafio em IoT [4, 12–15, 31–36], principalmente devido à heterogeneidade entre os vários componentes e plataformas de interconexão, as restrições de recursos dos dispositivos e o número de dispositivos conectados que aumenta continuamente. Além disso, as tecnologias de comunicação sem fio, sendo comumente encontradas em soluções IoT (*e.g.* WiFi e LoRa), também desempenham um papel neste desafio, dado que são inerentemente mais vulneráveis a intrusões, interferências ou ouvintes não autorizados, visto que suas transmissões não são fisicamente limitadas como no caso das tecnologias com fio.

Para o melhor funcionamento das redes é importante que cada dispositivo IoT possa ter uma forma de calcular e atribuir confiança nos outros dispositivos de modo a estabelecer uma comunicação segura, mesmo considerando o contexto caótico e inseguro que tais elementos estão inseridos. Assim, o objetivo deste trabalho é investigar soluções de construção de confiança entre dispositivos IoT e propor abordagens que garantam a integridade das identidades dos dispositivos ao longo de sua comunicação [13].

É importante diferenciar os conceitos de confiança e de reputação para termos um entendimento comum. Por um lado, quando um elemento infere sobre o quanto pode confiar em outro com base nas recomendações de uma comunidade, temos o conceito de reputação. Por outro lado, quando um elemento infere sobre o quanto confia em outro com base em suas próprias observações, temos o conceito de confiança [37]. Entendemos confiança como uma medida local com a qual um elemento atribui um valor que representa o quanto ele confia em outro. Cada elemento terá uma percepção diferente sobre a confiança nos outros, portanto, a confiança não é um atributo global com um mesmo valor para todos os elementos, mas sim um atributo individual. Assim, cada elemento deve calcular o valor de confiança no outro usando informações específicas diretamente relacionadas a tal elemento. A reputação pode ser considerada um atributo global, visto que a reputação de um elemento para a comunidade será a mesma para todos os elementos. A reputação está embutida na confiança.

De maneira geral, em uma interação entre seres humanos, o conceito de confiança está relacionado aos comportamentos ao interagir e vivenciar com outras pessoas (no presente ou no passado). Também está relacionado com os comportamentos e atitudes da própria pessoa que está construindo confiança nos outros. Em outras palavras, os aspectos culturais e o ambiente em que as pessoas estão inseridas influenciam seus comportamentos, consequentemente influenciando a confiança depositada nos outros.

Diferente do conceito centrado no ser humano, no contexto da IoT, construir

confiança de um dispositivo para outro não deve considerar o comportamento e as ações do próprio dispositivo que está construindo confiança, mas apenas informações dos dispositivos com os quais se comunica. Por exemplo, um dispositivo comprometido com *software* malicioso não considerará seu próprio histórico para computar a confiança em outro dispositivo. Na verdade, ele sempre atribuirá a maior confiança possível nos outros indiscriminadamente, visto que ele quer transmitir para qualquer dispositivo que puder de modo a infectá-los. Em contrapartida, um dispositivo lícito não deve considerar o fato dele ser um dispositivo autêntico para atribuir maior confiança em outro dispositivo com o qual se comunica, pois poderá inferir uma confiança relevante para um dispositivo que é malicioso.

Portanto, a confiança para a IoT deve ser construída com base em informações vindas dos dispositivos com os quais se estabelece comunicação, e não nas informações provenientes do próprio dispositivo que está construindo confiança. Dessa forma, a métrica de confiança controla se um dispositivo aceitará ou não a conexão de outros dispositivos. A métrica de confiança é assimétrica, dependente do contexto, dinâmica e não necessariamente transitiva, portanto, estabelecer confiança entre dispositivos IoT é desafiador. Especificamente, a métrica de confiança na verdade é uma *quasimétrica*, visto que o axioma de simetria não é satisfeito. Consideramos assim a definição de confiança baseada em [17, 26], a qual deve compreender ao menos quatro aspectos no contexto dos dispositivos IoT: a garantia da identidade, o comportamento de rede, a integridade de dados e a proteção dos dispositivos contra ataques.

1.3 Questões de Pesquisa em Aberto

Os desafios de pesquisa apresentados anteriormente e corroborados com o estado da arte dão substrato para as seguintes questões de pesquisa:

- Como fornecer uma métrica (ou especificamente, uma quasimétrica) de confiança na comunicação entre dispositivos IoT de modo a proteger os dispositivos lícitos de uma rede e isolar outros dispositivos que de alguma maneira tiveram sua segurança comprometida, sendo potencialmente maliciosos?
- Como quantificar e modelar matematicamente o conceito de confiança?

1.4 Inovações e Contribuições da Tese

Os aspectos essenciais para o estabelecimento da confiança são a garantia da unicidade, veracidade e autenticidade da identidade dos dispositivos. Um dispositivo IoT deve saber se o outro com o qual pretende se comunicar é de fato o dispositivo

desejado. Tal problema levanta a necessidade de um mecanismo de identificação de dispositivos IoT que forneça garantia sobre a unicidade das identidades, bem como proteção contra adulteração, duplicação, falsificação e personificação.

Uma possível estratégia para atingir este objetivo é usar o esquema de criptografia baseada em identidades (*Identity-Based Encryption – IBE*). IBE, como o próprio nome indica, é baseado nas identidades únicas das partes comunicantes para gerar chaves privadas e públicas. Resumidamente, uma das vantagens do IBE é que, ao enviar uma mensagem, os remetentes não precisam trocar chaves com os destinatários, uma vez que as chaves públicas dos destinatários podem ser geradas a partir de informações já conhecidas pelos remetentes, como o endereço IP ou e-mail, por exemplo. Mais detalhes sobre IBE podem ser encontrados no Apêndice A.

A eficiência energética é um requisito fundamental nos sistemas IoT, visto que os dispositivos são geralmente alimentados por bateria. Desta forma, reduzir o número de mensagens trocadas entre os dispositivos durante a autenticação contribui diretamente na redução do consumo de recursos computacionais e de rede dos dispositivos, consequentemente também contribui para a economia de energia. Não obstante, o próprio IBE requer uma entidade central chamada Gerador de Chave Privada (*Private-Key Generator – PKG*), que gera as chaves privadas dos dispositivos. Isso é necessário ao usar IBE porque, neste esquema, uma entidade pode gerar as chaves privadas de todas as outras entidades, visto que tem acesso a todas as informações públicas necessárias para tal. Assim, o PKG adiciona um elemento aleatório a cada chave privada e as distribui para os respectivos dispositivos.

Na arquitetura IBE, o PKG é notavelmente um ponto de vulnerabilidade, dada sua capacidade de abrir qualquer mensagem ou personificar a identidade de um dispositivo. Para resolver esse problema, propomos o uso de uma abordagem baseada em livro-razão distribuído (*distributed-ledger*) através de uma cadeia de blocos (*blockchain*) [14, 38–41] para descentralizar o PKG. Em uma cadeia de blocos, uma rede de nós computacionais, também conhecidos como *full nodes*, trabalha de modo distribuído para validar as transações armazenadas no livro-razão (*ledger*). Assim, a cadeia de blocos fornece a reputação dos dispositivos de modo transparente, como parte dos serviços oferecidos pela própria rede (*by design*). Neste ponto, é importante notar que apenas os nós que fazem parte de um domínio administrativo podem ser eleitos como *full nodes*, uma vez que eles serão imprescindíveis para a operação correta e confiável da abordagem.

Para que a reputação dos dispositivos esteja disponível publicamente para consulta por qualquer dispositivo IoT interessado, bem como restrita a *full nodes* permitidos, consideramos uma cadeia de blocos permissionada e pública. Os *full nodes* não devem ser dispositivos IoT, pois geralmente tais dispositivos têm recursos limitados. Em vez disso, em um sistema IoT de três camadas típico, como descrito em

[42] e ilustrado na Figura 1.2, os *full nodes* podem ser elementos computacionais localizados na camada de Borda, com mais recursos que os dispositivos na camada de Percepção. Mais detalhes sobre cadeia de blocos e, mais geralmente, livro-razão distribuído, podem ser encontrados no Apêndice D.

Quando um dispositivo deseja se comunicar com outro e precisa verificar se o outro é de fato quem afirma ser, a identidade desse dispositivo é verificada em relação a sua reputação na cadeia de blocos. Assim, o PKG pode ser descentralizado, com os nós de borda oferecendo o serviço de autorização/autenticação. Com isso, a segurança, em termos da confiança entre os dispositivos, será fornecida pela própria rede, o que traz vantagens como negar o acesso a um dispositivo malicioso, não permitindo sequer a entrada deste na rede, por não possuir uma identidade válida e aceita pela maioria dos *full nodes*.

Utilizamos a cadeia de blocos pelas funcionalidades que tal tecnologia oferece, como o encadeamento de transações, que permite o rastreamento e auditoria das informações registradas. Além disso, é muito difícil alterar qualquer informação armazenada, visto que seria necessário poder computacional para alterar não só as informações desejadas, mas também todo o histórico de transações modificadas em decorrência de tais alterações. Conforme comentado em [43], os sistemas de gerenciamento de identidade e acesso baseados em cadeia de blocos são abordagens promissoras para melhorar a segurança de aplicações IoT. Assim, o Nível Alto de nossa abordagem considera um componente baseado em cadeia de blocos para melhorar a segurança e oferecer uma confiança inicial (reputação) para permitir a comunicação entre os dispositivos.

Por outro lado, para estabelecer um consenso em uma cadeia de blocos é fun-

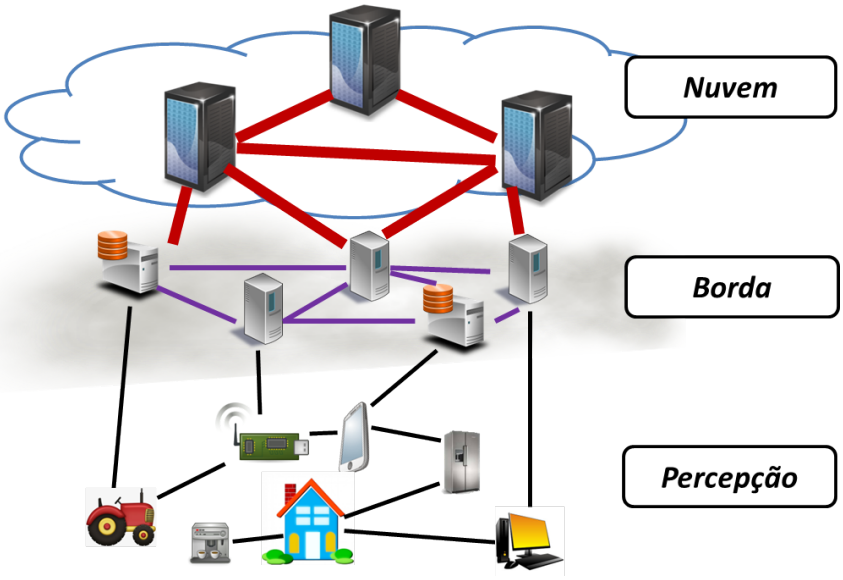


Figura 1.2: Arquitetura IoT de três camadas

damental que haja um número significativo de nós mineradores (*full nodes*) [16]. Além disso, a tecnologia de cadeia de blocos ainda apresenta limitações, como o tempo de validação da transação e a capacidade computacional necessária para formar *full nodes*. Tais aspectos podem incorrer em uma sobrecarga na comunicação entre os dispositivos IoT. Para lidar com esse problema, propomos a construção de uma infraestrutura de cadeia de blocos que possa lidar com vários requisitos e níveis de aplicação, desde aplicações locais associadas a uma cadeia de blocos local, até aplicações globais com uma cadeia de blocos em grande escala. Em particular, propomos o uso de um protocolo de consenso baseado em consórcio com prova de posse (*proof of stake*), que tem o potencial de se ajustar melhor a este contexto e pode ser implementado usando, por exemplo, o Tezos Blockchain [44].

Nossa abordagem é independente do modelo de comunicação, seja entre dispositivos, entre um dispositivo e um *gateway*, ou entre um dispositivo e a nuvem (considerando uma arquitetura de três camadas como em [42]).

Em relação aos quatro aspectos que assumimos para o conceito de confiança IoT na Seção 1.2, em nossa proposta incorporamos a garantia de identidade com o uso de IBE para identificação dos dispositivos; o comportamento da rede é abordado usando o componente Nível Baixo; a integridade dos dados é incorporada assumindo as comunicações bem sucedidas relatadas pelas confirmações das transações na cadeia de blocos; e a respetiva proteção dos dispositivos através de comunicações que apenas são permitidas para dispositivos lícitos. Nossa abordagem é projetada principalmente para aplicações IoT em que os dispositivos dependem principalmente da comunicação dispositivo a dispositivo.

A abordagem formulada para o problema de atribuição de métrica de confiança para dispositivos IoT apresentada nesta tese colabora para o avanço da linha de pesquisa e responde às questões de pesquisa colocadas anteriormente, sendo uma proposta inovadora com as seguintes contribuições:

- A abordagem de dois níveis para fornecer uma métrica de confiança baseada em informações provenientes tanto de características de rede, quanto de aplicação;
- O uso de uma estratégia baseada em livro-razão distribuído, mais especificamente cadeia de blocos, para fornecer um valor de confiança inicial e informações sobre características de aplicação, usado no *Nível Alto*;
- O uso de Teoria da Informação através do conceito de entropia relativa para fornecer confiança considerando a dinâmica das características do tráfego de rede dos dispositivos IoT, utilizado no *Nível Baixo*;
- A caracterização do tráfego de um dispositivo IoT e ajuste de distribuição;
- A quantificação e a modelagem matemática para o conceito de confiança.

1.5 Organização da Tese

Esta tese está organizada em 7 capítulos, contando ainda com 5 apêndices para fornecer mais detalhes sobre os temas abordados no trabalho.

O Capítulo 2 apresenta o estado da arte das pesquisas envolvendo o tema de segurança para Internet das Coisas, além de uma comparação qualitativa dos trabalhos relacionados com a proposta desta tese.

A proposta de modelagem do conceito de confiança baseada em dois níveis, defendida nesta tese, é apresentada no Capítulo 3.

No Capítulo 4 apresentamos experimentos realizados com o intuito de verificar a validade do modelo de confiança proposto.

No Capítulo 5, os valores de confiança obtidos a partir do modelo proposto são apresentados e discutidos, assim como uma caracterização do tráfego de um dos dispositivos IoT utilizados nos experimentos.

O Capítulo 6 define a métrica de tempo de contato entre dois dispositivos quaisquer e explora o comportamento do modelo de confiança proposto em termos desta métrica.

Por fim, o Capítulo 7 apresenta as conclusões da tese resumindo as principais contribuições e apresenta oportunidades de trabalhos, a fim de dar continuidade no desenvolvimento de novas soluções para os temas abordados.

Os Apêndices A, B, C, D e E apresentam, respectivamente, noções de criptografia, noções de teoria da informação, de Internet das Coisas, de livro-razão distribuído e um experimento sobre a propriedade de regularidade estatística, visando detalhar características específicas de cada tecnologia utilizada durante o desenvolvimento da tese.

Capítulo 2

Estado da Arte de Pesquisas em Segurança para Internet das Coisas

Os aspectos de segurança são reconhecidamente um grande desafio em IoT [4, 11–15, 31–35] por conta de diversos fatores, tais como, por exemplo, a heterogeneidade existente entre os diferentes dispositivos, componentes e plataformas de interconexão. A capacidade limitada de recursos dos dispositivos utilizados em IoT, assim como as diferentes tecnologias de comunicação sem fio inerentemente mais vulneráveis, fazem com que a atribuição de segurança em IoT seja um desafio. Em particular, o problema de atribuir métricas de confiança a dispositivos IoT é de suma importância, ainda sendo considerado como um problema em aberto [15, 16, 36, 45, 46].

Ling *et al.* [47] identificam funcionalidades básicas relacionadas a aspectos de segurança e privacidade em IoT, envolvendo desde os próprios dispositivos até os elementos de rede utilizados para conectá-los às diversas aplicações finais e dados relacionados. Ling *et al.* destacam a atualização de *firmware* dos dispositivos que, se não for realizada em tempo hábil, permite que vulnerabilidades de *software* sejam exploradas. O período de emparelhamento e associação também são aspectos relevantes, visto que um dispositivo conectando-se à rede pode se associar de maneira errada a algum elemento malicioso que esteja, por exemplo, atuando como uma infraestrutura de rede falsa, transmitindo dados através de um meio comprometido.

Uma análise sobre o nível de segurança de IoT é feita por Prokofiev *et al.* [33] apresentando estatísticas de países que possuem dispositivos vulneráveis, como, por exemplo, serem acessíveis através do protocolo *telnet*, conhecidamente inseguro e largamente usado para usurpar *logins* de usuários. No estudo, os três países com mais dispositivos acessíveis via *telnet* foram China, Estados Unidos e Brasil. Entre os dispositivos vulneráveis estão gravadores de vídeo, receptores de TV via satélite, roteadores, câmeras IP, entre outros. Muitos desses dispositivos acabam fazendo parte de *botnets*, como a Mirai, sendo transformados em poderosos vetores de ataque quando usados em conjunto, por exemplo, em um DDoS. O trabalho faz recomenda-

ções para alterar as configurações padrão dos dispositivos, incluindo a desativação de funções não utilizadas e protocolos conhecidamente inseguros, como o *telnet*. Esses e outros trabalhos ilustram como é factível usar dispositivos para expandir a superfície de ataque na Internet e como a segurança é um aspecto extremamente relevante em IoT, exigindo um esforço para fornecer aplicações minimamente seguras.

Sato *et al.* [45] propõem uma arquitetura IoT para confiança baseada na troca de garantias sobre a identidade dos dispositivos. O trabalho sugere construir confiança em uma determinada região espacial, que terá seu próprio controle sobre quais dispositivos serão cobertos por tal área.

Em [40], Banerjee *et al.* revisam a literatura em busca de artigos que propõem soluções de segurança para IoT. Os trabalhos encontrados discutem soluções baseadas em sistemas de detecção de intrusão (*Intrusion Detection System – IDS*) colaborativos, Redes Definidas por *Software* (*Software-Defined Networks – SDN*), modelos baseados em Cadeias de Markov, entre outros. Em seguida, os autores propõem o uso de cadeia de blocos para a criação de um banco de dados compartilhado usado como repositório de aplicações que usem tais dados em análises de métricas em IoT, detecção de *firmwares* comprometidos e possibilidade de autocorreção de dispositivos.

Os autores em [48] propõem uma arquitetura baseada em cadeia de blocos para fornecer um mecanismo de autenticação. Ao disparar contratos inteligentes, *tokens* são gerados para autorizar os dispositivos. O trabalho apresenta uma avaliação dos tempos de resposta para encriptação, decriptação, assinatura e verificação, sendo os dois últimos os mais longos.

Dai *et al.* [49] apresentam uma pesquisa aprofundada sobre a integração de cadeia de blocos e IoT, discutem os desafios deste novo paradigma e exploram uma arquitetura integrada chamada Cadeia de Bloco das Coisas. Os autores investigam as aplicações da arquitetura proposta em diversos domínios, por exemplo, cadeia de suprimentos, saúde, Internet de Veículos, entre outros.

Como o aspecto social também está relacionado aos dispositivos IoT e a maneira como podem interagir, algumas abordagens apresentam soluções baseadas em estratégias de redes sociais. A *Social IoT* (SIoT) é uma linha de pesquisa que vem ganhando relevância na comunidade científica. Fortino *et al.* [41] projetam um *framework* em que cada dispositivo IoT está associado a um agente de *software* capaz de explorar suas atitudes sociais para cooperar, bem como para formar estruturas sociais complexas de agentes. Os autores consideram o aspecto da reputação ao usar uma implementação de cadeia de blocos e os dispositivos podem usar serviços de rede de acordo com sua reputação fornecida pela cadeia de blocos.

Bernabe *et al.* [50] usam lógica *fuzzy* para fornecer uma solução de segurança abrangente através de um mecanismo de autorização leve e um novo modelo de

confiança desenvolvido especialmente para ambientes IoT. A abordagem dos autores considera quatro dimensões: qualidade de serviço, reputação, aspectos de segurança e relações sociais, para calcular os valores de confiança sobre os dispositivos IoT.

Os autores em [51] se concentram no nível de computação em névoa (*fog computing*) para oferecer um sistema de gerenciamento de confiança bidirecional, permitindo que um solicitante de serviço determine a confiabilidade de um provedor de serviço, e vice-versa, antes de iniciar uma conexão. Os autores usam lógica *fuzzy* para agregar a confiança obtida usando a qualidade do serviço, a qualidade da segurança, as relações sociais e as métricas de reputação passadas.

O trabalho de Filippi *et al.* [52] baseia-se na extensa discussão acadêmica sobre os conceitos de confiança e segurança para argumentar que a tecnologia de cadeia de blocos não é uma tecnologia que elimina a necessidade de confiança ("*trustless technology*"), mas é uma máquina de geração de confiança ("*confidence machine*"). Eles apontam que a tecnologia de cadeia de blocos depende de regras criptográficas e matemática para aumentar a confiança nas operações de um sistema computacional. No entanto, esse aumento na confiança depende da operação e governança adequadas da rede subjacente baseada em cadeia de blocos, o que requer a confiança de uma variedade de atores. Ou seja, apenas o fato de usar cadeia de blocos não implica em um sistema totalmente coberto de confiança.

Tang *et al.* [53] usam uma combinação de contratos inteligentes, em que cada interação entre os dispositivos é assinada pelos participantes e gravada numa cadeia de blocos. Os autores destacam que ainda há a necessidade de desenvolvimento de uma estrutura de consenso para estabelecer confiança em IoT.

A proposta descrita por Chen *et al.* [54] apresenta um modelo usando médias ponderadas de confiança, considerando também confiança relativa regional e de dados históricos. Para construir tal modelo, os autores consideram o histórico dos dados gerados por dispositivos em relação ao comportamento destes em situações de informações distorcidas, informações injetadas, mudança de frequência de transmissão e varredura. Através de simulações, os autores mostram que sua abordagem é ciente do consumo de energia com redução significativa do uso de recursos e com maior taxa de detecção em comparação com outras abordagens. A eficiência energética é de suma importância em IoT, dado que os dispositivos geralmente usam fontes de energia limitada.

Hongjun *et al.* [55] usam a Teoria da Informação para construir confiança entre os dispositivos. Eles representam os relacionamentos com um grafo direcionado e calculam a entropia da capacidade de um dispositivo em realizar uma ação e detectar dispositivos maliciosos na rede.

Khan *et al.* [56] propõem uma abordagem baseada em confiança para gerenciar a reputação de cada dispositivo de uma rede IoT com base no protocolo de rotea-

mento para redes de baixa potência e com perdas (*Routing Protocol for Low Power and Lossy Networks – RPL*). Os resultados alcançados com a avaliação de desempenho da proposta mostram valores de taxa de entrega média maiores do que outras propostas. A abordagem dos autores mostra a capacidade de detectar e também isolar nós maliciosos, resultando em resiliência da rede, bem como menor número de dispositivos com comportamento incorreto (dispositivos comprometidos) e menor número de caminhos que incluem um dispositivo comprometido.

O trabalho de Caminha *et al.* [57] apresenta um método de gerenciamento de confiança inteligente baseado em aprendizado de máquina que avalia automaticamente a confiança da IoT analisando os atributos do provedor de serviços. Eles também usam um recurso de janela deslizante elástica que ajuda a diferenciar os dispositivos quebrados ou com defeito, dos dispositivos com comportamento inadequado (potencialmente maliciosos).

Quanto ao uso de curvas elípticas para aplicar criptografia baseada em identidades (*Identity-Based Encryption – IBE*), o trabalho de Barbosa [58] propõe uma extensão do protocolo *Simple Mail Transfer Protocol* (SMTP) para oferecer troca segura de mensagens considerando o uso de mecanismos de criptografia baseados em curvas elípticas e identidade. A proposta considera a robustez dos mecanismos da curva elíptica, comentada no Apêndice A, a respeito da dificuldade computacional de quebrar tais criptografias.

Os autores em [59] propõem uma arquitetura de segurança que combina IBE com a infraestrutura de chave pública tradicional. O trabalho divide os elementos da solução proposta em três camadas: aplicação, transporte e sensoriamento. Já em [60], os autores consideram uma implementação conjunta usando IBE e cadeia de blocos. Os autores distribuem as identidades dos dispositivos na cadeia para concluir a autenticação do usuário e prover proteção à chave privada.

Junior e Kamienski [21] propõem uma estrutura para fornecer confiança no nível de dados para sistemas IoT baseados em computação em névoa. Sua estrutura procura garantir uma operação contínua e ininterrupta do fluxo de dados de IoT. Os autores também discutem os desafios e os compromissos relacionados aos mecanismos de confiança dos dados na IoT e apresentam propostas de fluxo de dados para os diferentes níveis de comunicação (*Things*, *Fog* e *Cloud*).

A maioria dos trabalhos considera características de nível de aplicação e apenas alguns (incluindo nossa abordagem) consideram características de rede. A Tabela 2.1 mostra uma comparação entre os trabalhos discutidos anteriormente e nossa abordagem em termos dos seguintes aspectos: (i) qual técnica é usada, (ii) se um conjunto de dados é considerado e, em caso afirmativo, que tipo de conjunto de dados, (iii) se a abordagem considera o nível de rede e/ou características de nível de aplicação, (iv) que tipo de arquitetura é usada, (v) se a abordagem é ciente de restrição de

recursos e/ou ciente de dinamismo, e (vi) se a abordagem oferece valores de confiança. Os atributos da tabela foram extraídos com base nas características comuns encontradas nos trabalhos relacionados ao tema. Tais trabalhos foram em parte identificados em [15], no qual apresentamos uma revisão sistemática da literatura seguindo um protocolo de pesquisa rigoroso. Nesta revisão, pudemos identificar o tema sobre confiança entre dispositivos IoT como um problema relevante e ainda em aberto.

Os trabalhos mencionados enfatizam a importância e relevância da construção de abordagens com foco em soluções para o problema de confiança entre dispositivos IoT. Nesta tese, além de apresentar um modelo de confiança que combina técnicas de cadeias de blocos e Teoria da Informação, a principal contribuição de nosso trabalho é a dupla perspectiva, tanto no nível de aplicação, quanto no nível de rede. Assim, fornecemos uma métrica de confiança mais abrangente que pode lidar com as particularidades dos dispositivos IoT.

Tabela 2.1: Comparação entre os trabalhos relacionados e a proposta desta tese

Referência	Técnica	Conjunto de Dados	Nível de Rede	Nível de Aplicação	Arquitetura	Ciente de Restrição de Recursos	Ciente de Dinamismo	Oferece Métrica de Confiança
[37]	Redes Bayesianas	Nenhum	Não	Sim	Distribuído	Não	Sim	Não
[41]	Baseado em Agentes + Cadeia de blocos	Nenhum	Não	Sim	Distribuído	Sim	Sim	Não
[50]	Lógica Fuzzy	Real	Não	Sim	Distribuído	Sim	Não	Sim
[53]	Cadeia de blocos	Nenhum	Não	Sim	Distribuído	Não	Não	Não
[54]	Médias Ponderadas + Fusão de Dados	Sintético	Não	Sim	Descentralizado	Sim	Sim	Sim
[55]	Entropia	Sintético	Não	Sim	Distribuído	Não	Não	Sim
[56]	RPL + Lógica Subjetiva	Nenhum	Sim	Não	Distribuído	Sim	Não	Não
[57]	Machine Learning	Real + Sintético	Não	Sim	Centralizado	Não	Não	Não
[60]	IBE + Cadeia de blocos	Nenhum	Não	Sim	Distribuído	Não	Não	Não
[61]	Cadeia de blocos	Real	Não	Sim	Distribuído	Sim	Não	Não
[62]	Baseado em Agentes + Microsserviços	Nenhum	Não	Sim	Distribuído	Sim	Sim	Não
Esta proposta	Entropia Relativa + Cadeia de Blocos	Real + Sintético	Sim	Sim	Distribuído	Sim	Sim	Sim

Capítulo 3

Modelagem do Conceito de Confiança

Este capítulo apresenta uma análise aprofundada dos componentes essenciais que definem o conceito de confiança no contexto de IoT. Além disso, apresentamos os detalhes da arquitetura proposta, bem como a modelagem matemática do conceito de confiança.

3.1 Identificação dos Componentes de Confiança

Para modelar e definir uma métrica de confiança no contexto de IoT, é preciso saber quais informações são necessárias para compor tal métrica. Considerando um exemplo de cenário de comunicação em que dois dispositivos IoT não possuem histórico de comunicação entre si, podemos identificar três etapas essenciais.

1. Primeiramente, os dispositivos precisam construir uma confiança inicial de modo que a comunicação possa ser estabelecida. Neste caso, identificamos a necessidade de um elemento externo a estes dispositivos que possa fornecer informações sobre os mesmos a respeito da reputação que estes têm perante a comunidade de dispositivos conectados. Com base neste valor inicial, os dispositivos podem decidir se prosseguirão com a comunicação ou não. Dado que os dispositivos tiveram sucesso em recuperar as respectivas reputações, adquirindo assim uma confiança inicial suficiente, a comunicação é então iniciada.

2. Em segundo lugar, visto que um dispositivo considerado, a princípio, como lícito, pode ser invadido, violado ou adulterado tornando-se potencialmente malicioso, surge a necessidade de cada dispositivo ajustar dinamicamente sua confiança no outro ao longo da comunicação. Tal ajuste contínuo nos valores de confiança pode ser feito com base no comportamento de rede observado a

partir da taxa de dados que um dispositivo recebe de outro. Assim, a confiança pode ser incrementada quando o comportamento é segundo o esperado, ou penalizada caso contrário. Os dispositivos devem ainda estipular um limiar de confiança mínima aceita, de acordo com a aplicação em questão, de modo a controlar suas comunicações. Caso os valores de confiança fiquem abaixo do limiar predefinido, então a comunicação deve ser encerrada, pois o dispositivo não é mais confiável.

3. Finalmente, após o término da comunicação, os valores de confiança adquiridos devem continuar sendo ajustados com o passar do tempo devido ao alto dinamismo característico de IoT. Especificamente, os valores de confiança devem diminuir ou expirar ao longo do tempo, enquanto não houver novamente comunicação entre os dispositivos, visto que estes podem ser adulterados no intervalo desde o último contato, tendo sua segurança comprometida. Assim, identificamos a necessidade de um mecanismo que diminua os valores de confiança periodicamente após o fim de uma comunicação. A confiança adquirida deve ser válida somente por um período de tempo após o término da comunicação e, caso os dispositivos queiram retomar a comunicação, então a confiança deve ser reconstruída.

Com base no cenário apresentado, identificamos três características relevantes para a construção da métrica de confiança. Precisamos definir:

- Uma confiança inicial para estabelecimento do primeiro contato;
- Ajustar a confiança de acordo com o comportamento do tráfego observado conforme as interações ocorrem;
- Decrementar a confiança com o passar do tempo após o fim de uma conexão.

Como mencionado no Capítulo 1, propomos uma estratégia baseada em dois níveis para a modelagem do conceito de confiança, considerando tanto as características de aplicação, quanto as características de rede, no que em nossa abordagem chamamos respectivamente de Nível Alto e Nível Baixo [63]. No Nível Alto, adotamos uma abordagem baseada em cadeia de blocos para armazenar as identidades dos dispositivos e utilizá-las em um esquema de criptografia baseado em identidades (IBE). Já o Nível Baixo é responsável por ajustar dinamicamente os valores de confiança dos dispositivos IoT durante a comunicação entre estes.

O Nível Baixo é composto por dois componentes: (i) componente de análise do comportamento do tráfego do dispositivo; (ii) componente de decaimento temporal da confiança. No componente de análise de tráfego usamos o conceito de entropia relativa de Teoria da Informação [3], também conhecido como divergência de

Kullback-Leibler [64], para modelar a vazão de entrada de um dispositivo de modo a capturar mudanças na distribuição do comportamento do tráfego de tal dispositivo. O componente de decaimento temporal da confiança, também pertencente ao Nível Baixo, lida com o dinamismo e os aspectos oportunistas da IoT, diminuindo os valores de confiança assim que os dispositivos param de se comunicar. Considerar os aspectos dinâmicos em IoT é essencial devido às frequentes mudanças na topologia, seja por conta do esgotamento da bateria dos dispositivos, por dispositivos que entram e saem da rede, interrupções ou perdas de conexão, mobilidade, desligamentos programados, entre outros.

3.2 Arquitetura da Proposta

O paradigma emergente de Computação na Borda [65, 66] visa mover recursos de computação, processamento e armazenamento para a borda da rede, em vez de centralizá-los em *data centers* remotos na nuvem. Tal abordagem cria uma infraestrutura que oferece menor latência para as aplicações em comparação com a nuvem. Os elementos de borda podem variar de *switches*, roteadores, estações base, até *gateways* inteligentes ou micro *data centers*, geralmente tendo recursos mais limitados do que os elementos de nuvem, mas com mais recursos do que os dispositivos de IoT. Impulsionado por este paradigma, consideramos em nossa abordagem uma arquitetura típica composta por três camadas, conforme descrito em [42], consistindo em uma Camada de Dispositivos, uma Camada de Borda e uma Camada de Nuvem.

Além disso, consideramos também o ciclo de vida do processo de gerenciamento da confiança composto por cinco etapas, conforme proposto em [67]. Tais etapas contemplam:

1. **Observação**, um período no qual dados dos dispositivos são capturados como insumos para geração dos valores de confiança;
2. **Pontuação**, quando os valores de confiança são computados e classificados;
3. **Seleção**, onde as decisões com base nos valores de confiança obtidos são tomadas;
4. **Transação**, etapa em que as comunicações entre os dispositivos ocorrem dentre aqueles dispositivos selecionados na etapa anterior que alcançaram o mínimo de confiança estipulado por suas respectivas aplicações;
5. **Recompensa ou Punição**, que ajusta os valores de confiança com base nos resultados das comunicações entre os dispositivos.

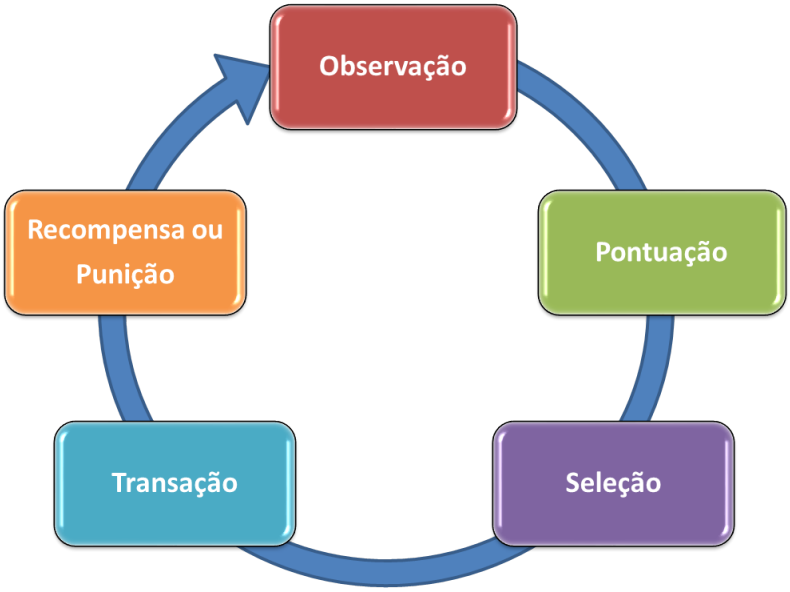


Figura 3.1: Ciclo de vida do processo de gerenciamento de confiança. Adaptado de [67]

A Figura 3.1 ilustra o ciclo de vida que consideramos para o gerenciamento da confiança. No exemplo a seguir veremos que nossa abordagem contempla todas as etapas deste ciclo de vida.

Vamos considerar um cenário em que dois dispositivos na Camada de Dispositivos desejam se comunicar e não têm histórico de comunicação previamente estabelecido. Eles precisam inferir um valor de confiança inicial, pois têm pouca ou nenhuma informação sobre o outro. Na Figura 3.2, são ilustrados os possíveis casos de utilização do Nível Alto da nossa abordagem. Cada dispositivo IoT consulta a reputação (confiança inicial) da identidade do outro dispositivo através do Nível Alto de nossa proposta (etapa de Observação do ciclo de vida). O Nível Alto também é utilizado quando o valor de confiança é reduzido a um patamar abaixo de um limiar mínimo de confiança predefinido e o dispositivo precisa adquirir novamente o valor

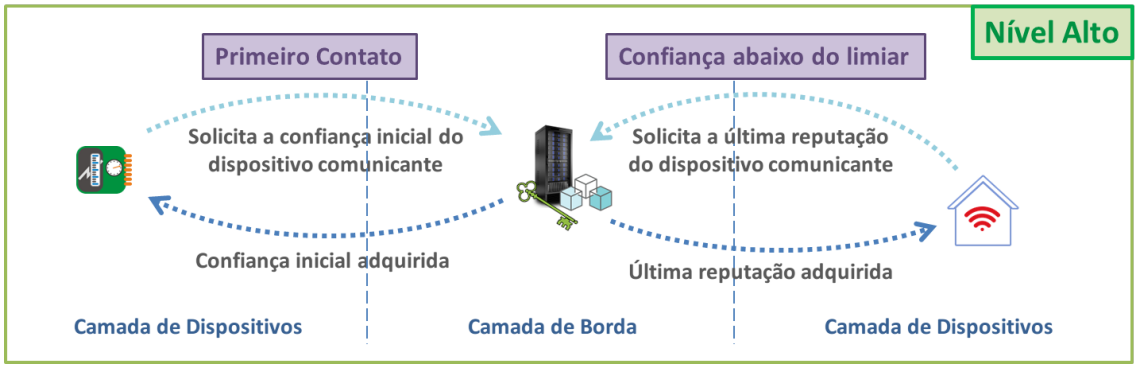


Figura 3.2: Uso do Nível Alto: (i) quando o primeiro contato é estabelecido e um valor de confiança inicial precisa ser adquirido; (ii) quando o valor de confiança é reduzido a um patamar abaixo de um limiar mínimo de confiança predefinido

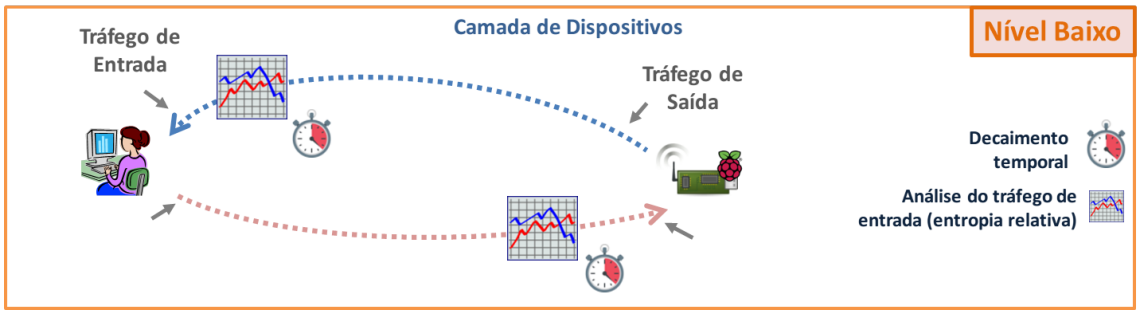


Figura 3.3: Exemplo de uso do Nível Baixo: cada dispositivo constrói sua confiança conforme a entropia relativa do tráfego recebido do dispositivo comunicante. O decaimento temporal da confiança se dá após um período de ociosidade ou desconexão dos dispositivos em termos de tráfego

de reputação mais atual a partir do Nível Alto.

Já na Figura 3.3, o Nível Baixo é utilizado durante toda a comunicação após a obtenção da confiança mínima. Neste momento, a taxa de dados recebida por um dispositivo[1] é analisada de modo a identificar mudanças no padrão de tráfego (etapa de Observação do ciclo de vida). Em resumo, as etapas operacionais da proposta são definidas da seguinte forma:

- **Etapa 1:** Cada dispositivo IoT consulta a identidade do outro (baseado no esquema de IBE) em uma infraestrutura de cadeia de blocos que armazena as identidades dos membros da rede. Para lidar com os altos custos computacionais comumente encontrados em cadeias de blocos, sugerimos a implementação dos *full nodes* na Camada de Borda, e não na Camada de Dispositivos, uma vez que os dispositivos IoT são conhecidamente restritos de recursos;

- **Etapa 2:** Uma vez que uma confiança inicial mínima é estabelecida, a comunicação inicia, podendo acontecer: (i) diretamente entre elementos da Camada de Dispositivos; (ii) entre um elemento da Camada de Dispositivos e um da Camada de Borda, caso as solicitações sejam tratadas por esta; ou (iii) entre as Camadas de Dispositivos e de Nuvem, se as solicitações só puderem ser tratadas por nós computacionais da nuvem. Enquanto ocorre a comunicação entre os dispositivos, o Nível Baixo passa a operar. Cada dispositivo calcula a entropia relativa do tráfego do outro dispositivo e usa tal informação para ajustar os respectivos valores de confiança (etapa de Pontuação do ciclo de vida). Se um dispositivo começa a se comportar anormalmente, então a confiança atribuída será penalizada, potencialmente encerrando a comunicação, caso os valores de confiança sejam reduzidos para um valor abaixo de um limiar mínimo previamente estabelecido, contemplando a etapa de Seleção do ciclo de

[1]Podemos considerar a taxa de dados recebida por um dispositivo como o tráfego de entrada formado pelos fluxos provenientes do dispositivo comunicante.

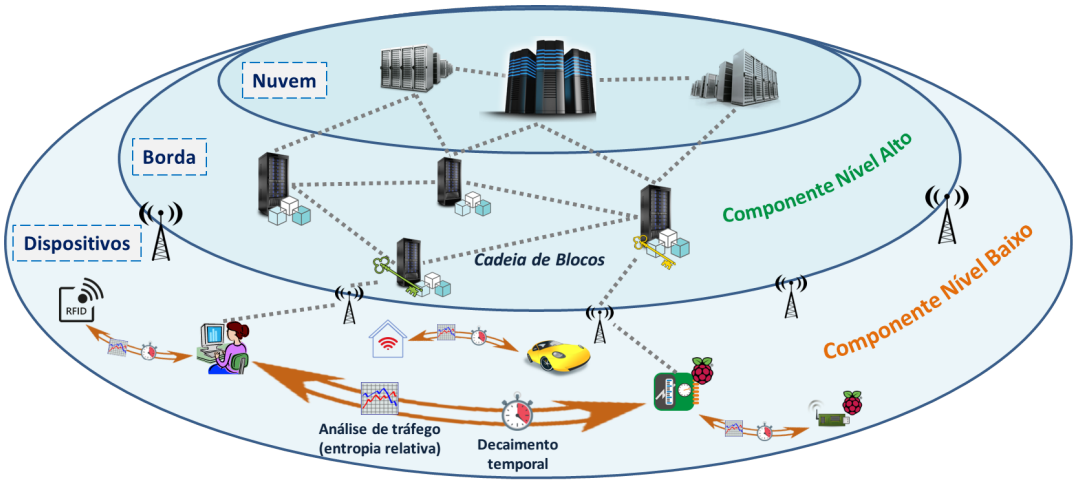


Figura 3.4: Cenário de abordagem de confiança de dois níveis, considerando uma arquitetura de camada de árvore e a colocação de nós da cadeia de blocos na borda

vida. No nosso caso, um comportamento anormal significa que a distribuição do padrão de tráfego observado divergiu da distribuição de tráfego esperada. Caso o valor de confiança seja suficiente, então a comunicação é permitida, remetendo à etapa de Transação do ciclo de vida.

- **Etapa 3:** Quando a comunicação termina, um componente temporal passa a reduzir o valor de confiança, atuando como um prazo de validade (*timeout*) para a confiança adquirida. Caso a confiança reduza até alcançar o limiar mínimo, então os dispositivos precisam consultar a cadeia de blocos novamente e reiniciar todo o processo para readquirir a confiança necessária para estabelecer comunicação. Neste caso, o *timeout* atua como parte da última etapa do ciclo de vida da confiança.

Deta maneira, propomos uma abordagem de atribuição de confiança baseada em características de aplicação e de rede que podem ser observadas por dispositivos IoT. A Figura 3.4 ilustra o esquema de funcionamento da nossa abordagem. Podemos perceber os dispositivos comunicando entre si na Camada de Dispositivos, onde o componente Nível Baixo atua. Da mesma forma, podemos observar as conectividades dos dispositivos com a Camada de Borda, onde o componente Nível Alto atua, com a cadeia de blocos formada pelos nós de borda.

3.3 Modelagem Matemática

A partir do cenário exposto, começamos nossa modelagem matemática do conceito de confiança em IoT. Inicialmente, consideramos a variável aleatória X_{ji} para representar o comportamento da taxa de dados em *Bytes* por segundo ($Bytes/s$) que

um dispositivo j deve receber de um dispositivo i, para qualquer par de dispositivos em uma rede IoT. X_{ji} é uma variável aleatória inteira, não-negativa, que assume valores no intervalo $\mathcal{S}_{ji} = [0, \Delta, 2\Delta, 3\Delta, ..., R_{ji}^{max}]$, onde Δ é um inteiro positivo e R_{ji}^{max} é a taxa máxima recebida. Assim, após observar amostras da variável aleatória X_{ji}, podemos obter a respectiva distribuição amostral,

$$p_{X_{ji}}(x) \triangleq Pr[X_{ji} = x], \quad x \in \mathcal{S}_{ji}.$$

Neste ponto, vale notar que $p_{X_{ji}}(x)$ é a distribuição amostral obtida a partir da frequência relativa das amostras observadas. Através da propriedade de regularidade estatística [68], sabemos que cada valor $p_{X_{ji}}(x)$ com $x \in \mathcal{S}_{ji}$ é obtido a partir da convergência de longas sequências de rodadas de um experimento aleatório. Alguns experimentos sobre a propriedade de regularidade estatística podem ser vistos no Apêndice E. Assim, a partir da distribuição previamente estimada da variável aleatória X_{ji}, podemos obter métricas de informação com base nas fórmulas da Teoria da Informação de Shannon [3], como discutiremos a seguir.

Seja TR_{ji} a confiança do dispositivo j no dispositivo i. Os valores de confiança variam no intervalo entre 0.0 (zero), indicando o menor valor de confiança possível (ou simplesmente sem confiança), até 1.0, que representa o valor máximo de confiança. TR_{ji} é definida por três componentes:

1. TR_{ji} é inicialmente calculado com base na confiança da identidade de i, obtida a partir de sua reputação armazenada em uma cadeia de blocos pública e permissionada, expressa por C_1 (Equação 3.1). Os valores iniciais de reputação representados pelas transações na cadeia de blocos são obtidos através de comunicações anteriores estabelecidas com sucesso, as quais os dispositivos registram no final de uma interação. O número de confirmações que uma transação possui na cadeia de blocos fornece o valor de reputação dado pelo componente C_1.

$$C_1 = \text{número de confirmações que um dispositivo } i \text{ apresenta} \tag{3.1}$$

   Para fins de modelagem, este componente segue uma distribuição Gaussiana com parâmetros $\mu = 1$ e $\sigma^2 = 1$ para ter amostras do número de confirmações normalmente distribuídas com valores próximos ao recomendado [38]. Quanto mais confirmações uma transação tiver, mais difícil será adulterá-la. Portanto, com base no número de confirmações, podemos considerar que a transação é fortemente aceita pelos membros da cadeia de blocos (*full nodes*) e a probabilidade de violação pode ser considerada insignificante;

2. TR_{ji} também é influenciado pelo cálculo da entropia relativa da taxa de dados

recebida por um dispositivo. O valor de TR_{ji} muda quando o comportamento da taxa de dados atual do dispositivo se desvia do comportamento estimado devido a qualquer tipo de condição anômala. A distribuição estimada da taxa de dados recebida X_{ji} pode ser obtida através de observações iniciais do comportamento da taxa de dados do dispositivo comunicante, pelas quais o dispositivo receptor aprende a distribuição "verdadeira". Vejamos a seguir alguns conceitos necessários para o desenvolvimento do nosso modelo de confiança.

A auto-informação de um evento $\{X_{ji} = x\}$ é definida como [3, 69–71]:

$$I(x) \ = \ -\log p_{X_{ji}}(x).$$

A média da auto-informação é a entropia da variável aleatória X_{ji}, conforme a Equação 3.2.

$$H(X_{ji}) \ = -\sum_{x \in \mathcal{S}_{ji}} p_{X_{ji}}(x) \log p_{X_{ji}}(x) \tag{3.2}$$

Da mesma forma, definimos a variável aleatória inteira, não-negativa, Y_{ji}, que representa a taxa de dados observada que flui para um dispositivo j gerado por um dispositivo i. Y_{ji} também assume valores no intervalo $\mathcal{S}_{ji}$, conforme a distribuição

$$q_{Y_{ji}}(y) \triangleq Pr[Y_{ji} = y], \quad y \in \mathcal{S}_{ji}.$$

Usando as definições anteriores, podemos calcular a entropia relativa descrita na Equação 3.3, que representa a divergência de Kullback-Leibler [64, 70, 71], um tipo de "distância" entre duas distribuições. Essa medida não é considerada efetivamente uma distância, pois não é simétrica. Por outro lado, essa é uma característica desejada ao se tratar de confiança, visto que esta também não é simétrica (a confiança do dispositivo i no dispositivo j não é necessariamente a mesma confiança que j tem em i).

$$D_{KL}(p||q) = \sum_{x \in \mathcal{S}_{ji}} p_{X_{ji}}(x) \log \frac{p_{X_{ji}}(x)}{q_{Y_{ji}}(x)} \tag{3.3}$$

Assim, $p_{X_{ji}}(x)$ é a distribuição **estimada** da vazão (ou tráfego) do remetente i para o destinatário j, e $q_{Y_{ji}}(x)$ é a distribuição **observada** da respectiva vazão. À medida que $q_{Y_{ji}}(x)$ se aproxima de $p_{X_{ji}}(x)$ na Equação 3.3, a entropia relativa ("distância") $D(p||q)$ diminui. Assim, modelamos o comportamento da taxa de dados quando a distribuição observada difere da distribuição verdadeira (estimada) e ajustamos a confiança de um determinado dispositivo.

De modo a capturar a essência do conceito de divergência e trazê-lo para o contexto de cálculo de confiança, definimos o componente C_2 com a seguinte estratégia:

- Se o valor da divergência $D(p||q)$ obtido for menor que 1.0, então o valor de confiança calculado tem como base a fórmula:

$$C_2 = 1.0 - D(p||q)$$

- Para valores de divergência $D(p||q)$ maiores que 1, então o valor de confiança calculado tem como base a fórmula:

$$C_2 = -0.5 + \left(\frac{1}{D(p||q)}\right)$$

O racional por trás de tal estratégia está em atribuir maior confiança para dispositivos que apresentem valores de divergência abaixo de 1.0 e penalizar os que excedem 1.0. Valores abaixo de 1.0 indicam uma "proximidade" das distribuições (pouca divergência), ou ainda, pouco ganho de informação (informação mútua próxima de zero, o que significa pouca redução da incerteza sobre uma variável aleatória quando se observa outra) [69, 70]. Outro ponto é em relação aos valores escolhidos para C_2, cujo valor 1.0 é relativo a confiança máxima sendo decrementada pelo fator de divergência na primeira fórmula; e o valor -0.5 para representar uma penalização inicial amenizada segundo o inverso da divergência calculada. Tais valores são parametrizáveis no modelo.

3. Para o terceiro componente, consideramos um decaimento temporal que funciona como um *timeout*, diminuindo o valor de confiança quando os dispositivos param de se comunicar. IoT é um ambiente altamente dinâmico e oportunista. Os dispositivos se movem constantemente, às vezes por longas distâncias. Como um dispositivo não é capaz de saber para onde o outro dispositivo comunicante se moveu, ou em quais redes ele se associou, ou quais pessoas tiveram acesso a ele, manter um valor de confiança inalterado ao longo do tempo não representará efetivamente o quanto o dispositivo é confiável. Ou seja, uma vez que o contexto no qual as interações ocorrem está sujeito a mudanças, é necessário que haja uma expiração do valor de confiança a partir do momento em que a comunicação é encerrada. Quando a confiança alcança um valor abaixo de um limiar predefinido, os dispositivos precisam retornar ao primeiro caso de estabelecimento de confiança, ou seja, os dispositivos precisam obter novamente uma confiança mínima a partir da cadeia de blocos. Em nosso modelo, consideramos um decaimento temporal proporcional à confiança

adquirida, conforme descrito na Equação 3.4,

$$C_3 = TR_{ji} \times d \tag{3.4}$$

no qual d é um fator de decaimento.

O cálculo da confiança é atualizado conforme as amostras de tráfego dos dispositivos vão sendo coletadas e a comunicação entre estes evolui com o tempo. Cada dispositivo recalcula sua confiança no outro com o qual tem uma comunicação estabelecida com base na confiança atual calculada. Para isso, um dispositivo considera os seguintes casos:

1. Caso o valor de confiança esteja abaixo do limiar estipulado como aceitável, então o novo valor de confiança é calculado utilizando o componente Nível Alto de acordo com a fórmula abaixo:

$$TR_{ji}^{\text{atualizado}} = TR_{ji}^{\text{atual}} + C_1 \tag{3.5}$$

2. Caso o valor de confiança atual esteja acima do limiar, então o valor de confiança atualizado é calculado segundo o componente Nível Baixo conforme a fórmula abaixo:

$$TR_{ji}^{\text{atualizado}} = TR_{ji}^{\text{atual}} + C_2 \tag{3.6}$$

3. No caso em que não há comunicação, o componente de decaimento temporal passa a valer, conforme a fórmula abaixo:

$$TR_{ji}^{\text{atualizado}} = TR_{ji}^{\text{atual}} - C_3 \tag{3.7}$$

Assim, conforme mencionado na Seção 1.4, nossa proposta incorpora (i) a *garantia da identidade* com o uso de IBE para identificação de dispositivos, (ii) o *comportamento de rede* é abordado ao utilizar o componente Nível Baixo, (iii) a *integridade dos dados* é incorporada assumindo as comunicações bem sucedidas relatadas pelas confirmações das transações na cadeia de blocos, e (iv) a respectiva *proteção* dos dispositivos através de comunicações permitidas apenas para dispositivos lícitos, ou seja, aqueles que alcançaram ao menos o mínimo de confiança estipulado. Desta maneira, cobrimos os quatro aspectos do conceito de confiança definidos na Seção 1.2.

Capítulo 4

Verificação do Modelo Proposto

Neste capítulo, desenvolvemos uma avaliação experimental para analisar o comportamento do modelo de confiança proposto durante a comunicação entre dois dispositivos IoT quaisquer e verificar se o modelo suporta a dinamicidade das taxas de dados geradas pelos dispositivos.

4.1 Análise do Comportamento do Modelo Usando Dados Sintéticos

Nesta avaliação consideramos os resultados publicados no artigo [72]. Supomos que um dispositivo i envia dados para um dispositivo j de acordo com alguma distribuição de probabilidade, por exemplo, Poisson. Isso significa que, em cada intervalo de tempo τ (por exemplo, um segundo), um valor de *Bytes* por τ é gerado de acordo com essa distribuição. Neste caso, definimos como tráfego de i para j a soma das taxas de dados de todos os fluxos de i para j.

Supomos ainda que, após um determinado tempo da comunicação ter sido estabelecida e a troca de dados entre i e j ter iniciado, o dispositivo i passa a gerar tráfego para o dispositivo j de acordo com uma distribuição diferente da qual j espera receber. Com isso, queremos demonstrar o dinamismo da métrica de confiança conforme as mudanças do comportamento estatístico do tráfego de i para j ocorrem.

A cada instante de tempo, os valores de tráfego gerados de acordo com uma distribuição escolhida são considerados e os componentes das Equações 3.5, 3.6 e 3.7 são calculados. Consideramos também o uso de uma janela deslizante no tempo que contém os valores das taxas de dados recebidas dentro de um determinado período, em vez de utilizar o intervalo de dados completo. Tal estratégia visa contemplar as restrições de capacidade e de recursos, muito comuns dos dispositivos IoT, os quais podem não ser capazes de processar grandes quantidades de dados. Desta forma, fazemos testes com diferentes configurações de tamanhos de janela e vemos

o impacto de tais mudanças no comportamento da métrica de confiança proposta.

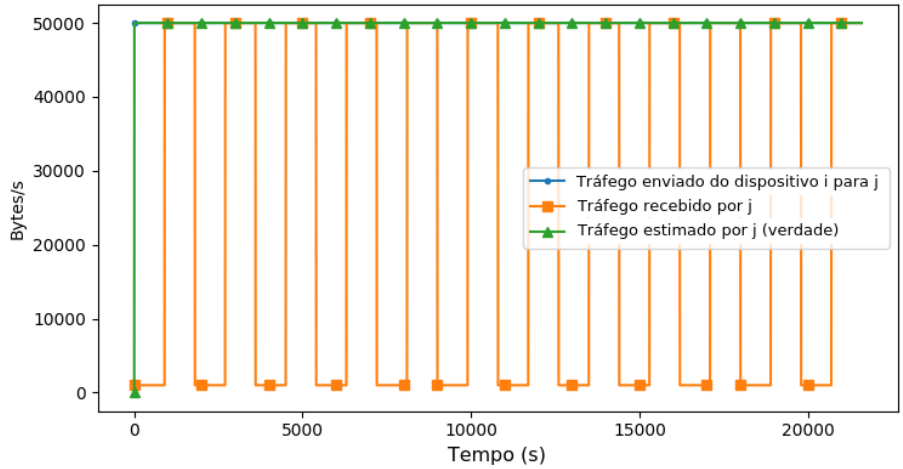


Figura 4.1: Padrão de tráfego sintético com variações periódicas

A Figura 4.1 mostra um padrão de tráfego sintético constante (curva verde com símbolos triangulares), representando o tráfego de um dispositivo IoT de uma aplicação de vídeo, que tem essa característica de envio contínuo; e outro com variações periódicas determinísticas a cada 1000 segundos (curva laranja com símbolos quadrados), que pode representar uma anomalia ou um dispositivo configurado para outra aplicação IoT, representando, por exemplo, o tráfego de sensores e atuadores que coletam informações ou recebem comandos com determinada frequência. Destacamos a diferença entre o tráfego efetivamente recebido pelo dispositivo j e o que este esperava receber, caracterizada por uma potencial modificação maliciosa em comparação com o tráfego enviado por i (curva azul por baixo da curva verde, a qual apenas observamos o ponto inicial no canto superior esquerdo do gráfico).

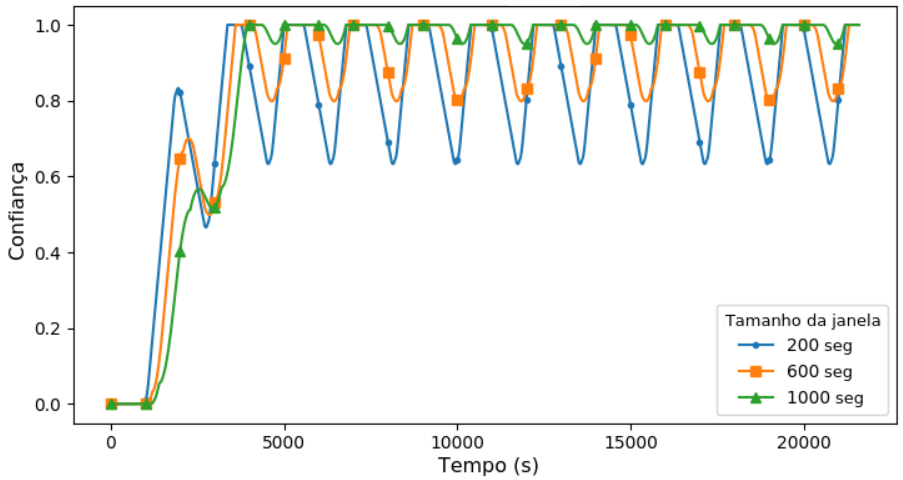


Figura 4.2: Valores de confiança usando um padrão de tráfego sintético com variações periódicas

Com este tipo de padrão de tráfego, na Figura 4.2 podemos perceber que o tamanho da janela deslizante tem impacto relevante nos resultados, o que é explicado pela sua relação com o tamanho do período de mudanças neste padrão de tráfego. Ainda nesta figura, podemos ver também que, quando o tráfego recebido corresponde ao esperado, o valor de confiança aumenta. Em contrapartida, quando o tráfego recebido diverge do esperado, a confiança é penalizada. Considerando esses resultados é possível ver que o componente Nível Baixo (C_2) é capaz de capturar as mudanças no padrão de tráfego dos nós, o que ajuda a compor a métrica de confiança e garante sua eficácia.

Percebemos também que se o tamanho da janela é menor que o intervalo em que as anomalias de tráfego ocorrem, então a métrica fica mais sensível às mudanças de padrão de tráfego, penalizando mais os valores de confiança, ao passo que janelas maiores, próximas do intervalo em que as mudanças ocorrem, as variações dos valores de confiança são suavizadas. Mais a frente na Seção 4.2 veremos outros impactos do tamanho da janela usando dados reais de tráfego.

Analisando a métrica de confiança com outro padrão de tráfego, desta vez probabilístico, a Figura 4.3 ilustra a geração de tráfego de dois dispositivos de acordo com uma distribuição de Poisson com parâmetro $\lambda = 10$ *Bytes/s*. Durante o intervalo entre os instantes 400 e 600 do experimento, o tráfego do dispositivo j muda seu padrão para outra distribuição, a saber, uma distribuição de Pareto com parâmetro $\alpha = 10$ *Bytes/s*. É importante ressaltar que o objetivo deste experimento não é investigar o comportamento específico do tráfego de acordo com as distribuições escolhidas, mas sim o comportamento da métrica de confiança em condições de variação dos perfis de tráfego, definidos aqui segundo as referidas distribuições. Ou seja, a escolha das distribuições foi arbitrária apenas para o exemplo em questão. Mais a frente no Capítulo 5 veremos como tais distribuições poderiam ser obtidas de maneira a representar de fato um determinado tipo de aplicação de IoT.

A métrica de confiança ao longo do tempo é mostrada na Figura 4.4. Para este experimento, foi utilizada uma janela deslizante de tamanho 200, visto que esta se mostrou mais sensível às variações de comportamento do tráfego. De acordo com a Figura 4.4, os valores de confiança variam ao longo do tempo, reduzindo quando o padrão de tráfego muda (entre os instantes 400 e 600) e se recupera quando o padrão volta ao comportamento esperado. As linhas verde e vermelha representam os limites definidos por cada dispositivo para decidir qual deve ser o valor mínimo de confiança aceito para estabelecer uma comunicação. Os valores de limite (limiares) são essenciais para nossa abordagem e devem ser configurados corretamente de acordo com as aplicações as quais os dispositivos estão associados e possivelmente ajustados ao longo do tempo para cobrir as mudanças na aplicação. Desta forma, fica a cargo do Nível Alto de nossa abordagem disponibilizar informações de nível de

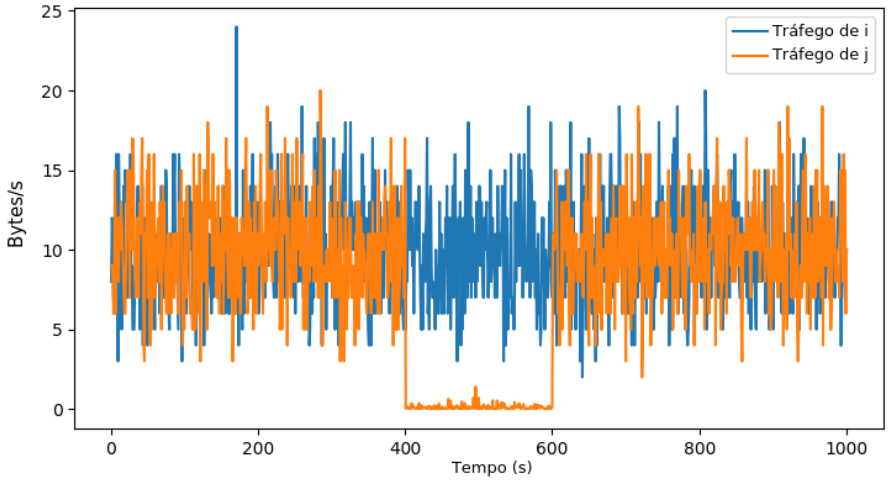


Figura 4.3: Tráfego gerado ao longo do tempo

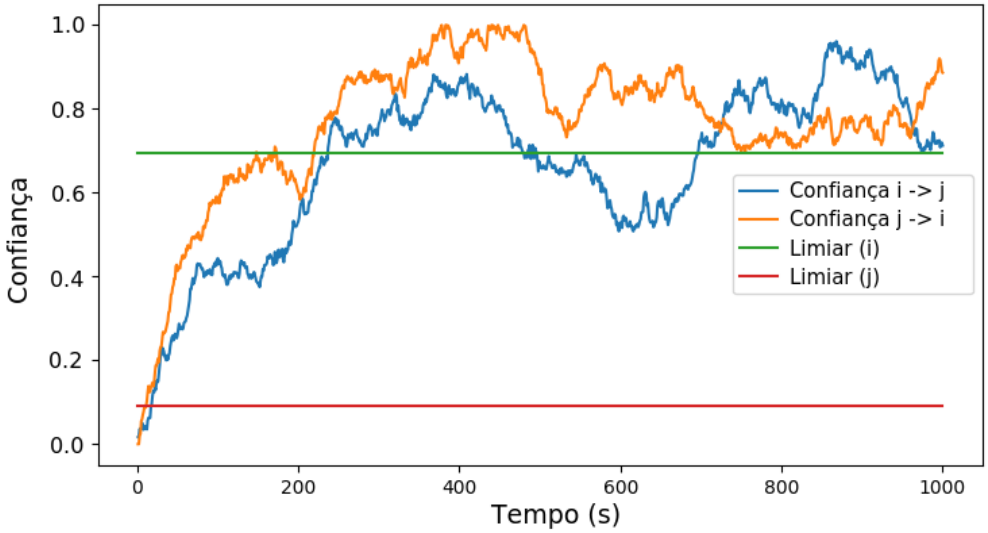


Figura 4.4: Confiança calculada ao longo do tempo considerando um tráfego sintético

aplicação que incluem o limite selecionado definido pela aplicação em questão, que pode aceitar valores de confiança mais baixos (menos restritos) ou apenas valores de confiança mais altos (mais restritos). Neste exemplo, o dispositivo i para de confiar no dispositivo j assim que começa a perceber o comportamento anômalo de j. Por outro lado, o dispositivo j continua confiando em i, uma vez que seu limite está definido como baixo $(0, 1)$, o que ilustra um dispositivo potencialmente malicioso, que confia facilmente em qualquer dispositivo para poder se comunicar com qualquer um.

A Figura 4.5 mostra como o componente de entropia relativa se comporta ao longo do tempo. Ele reconhece as mudanças de comportamento no padrão de tráfego para cada dispositivo, especialmente para o i (curva azul) durante o início do aprendizado da distribuição e o intervalo de anormalidade (do instante 400 a 600),

em que o valor do componente C_2 é diminuído. Vale perceber também que o dispositivo j adota valores para o componente C_2 de modo a manter sua confiança no dispositivo i estabilizada a ponto de poder continuar com a comunicação.

Durante o funcionamento da rede, é necessário que existam mecanismos que permitam o ajuste dos valores de confiança dinamicamente. Tal ajuste é feito por meio do componente C_2 baseado na entropia relativa, que ajusta a confiança em função no comportamento do tráfego, juntamente com o componente C_3 que reduz os valores de confiança obtidos ao longo do tempo após o término de uma comunicação. Assim, comportamentos que se desviam do esperado contribuem para a variação da entropia, o que consequentemente impacta o valor de confiança.

4.2 Análise do Comportamento do Modelo Usando Dados Reais

Para este experimento, consideramos o conjunto de dados de tráfego entre dispositivos IoT disponibilizados por Sivanathan *et al.* [73] da Universidade de Sidney. O conjunto de dados original é formado por diversos pacotes de dados sem modificações (*raw packets*) coletados durante um período de 3 semanas. Esses pacotes incluem informações de fluxo que formam *traces* para cada par de dispositivos (origem e destino).

Os dados foram coletados em um ambiente de *campus* universitário inteligente com mais de 20 dispositivos IoT, incluindo câmeras, luzes inteligentes, sensores de

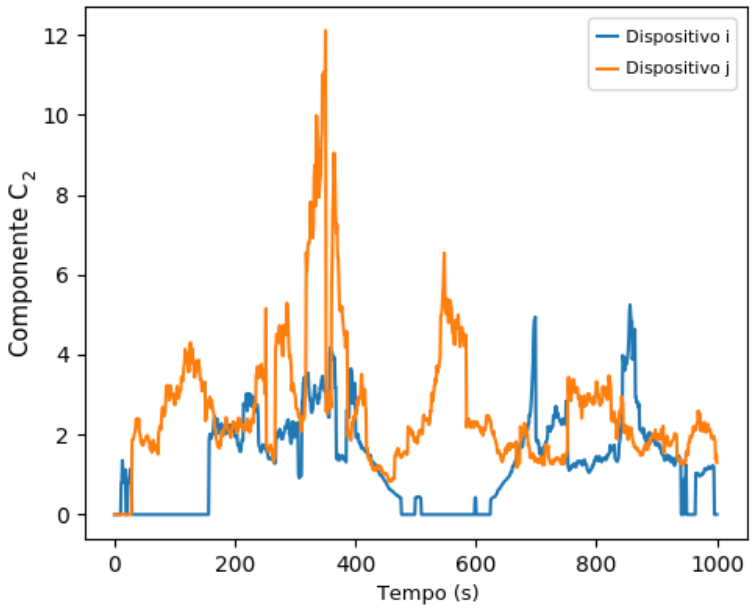


Figura 4.5: Componente C_2 baseado em entropia relativa ao longo do tempo

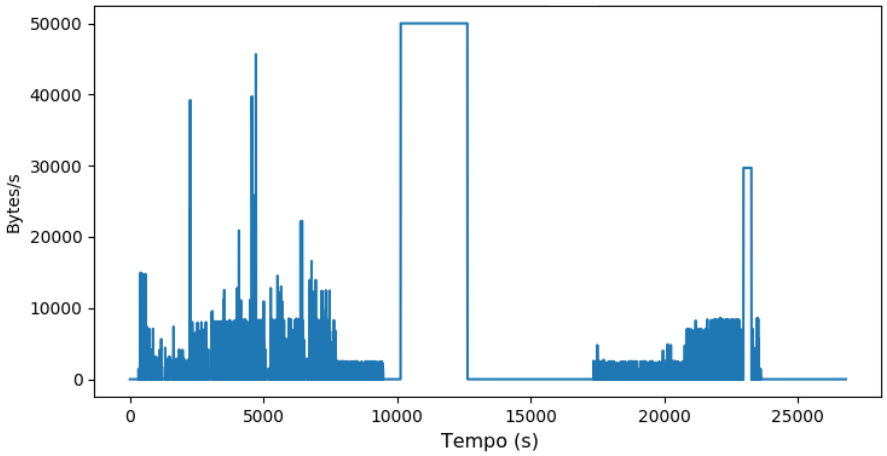


Figura 4.6: Tráfego produzido por um nó IoT usando o dia 28/09/2016 do conjunto de dados de [73] com uma modificação ligeiramente manual (picos nos 10000s e 23000s)

atividade e monitores de saúde. Em particular, em nossos experimentos, consideramos o período de duração de um dia do conjunto de dados e calculamos o tráfego em *Bytes/s* a partir dos fluxos de cada par de dispositivos, de acordo com a tupla (IP de origem, IP de destino), somando a quantidade de *Bytes* transmitidos em um segundo. Desta maneira, temos uma relação de uma amostra por segundo.

Foi gerada uma versão modificada das amostras de tráfego originais através da inserção de valores de vazão mais altos (50 KBytes/s) do que os encontrados no *trace* durante um determinado período de tempo, de modo a observar como a abordagem proposta se comporta sob as condições de inserção de tráfego anômalo. Na Figura 4.6 podemos observar o histórico de tráfego ao longo do tempo que um dispositivo IoT enviou para outro com a inserção de alguns picos de variações nos instantes 10000 e 23000.

A Figura 4.7 mostra os resultados dos valores de confiança, que aumentam assim que o padrão de tráfego começa a se estabilizar. Então, quando o primeiro pico é atingido, o valor de confiança cai, com sensibilidades diferentes dependendo do tamanho da janela deslizante. Então, à medida que o pico se torna o novo padrão, a confiança começa a aumentar novamente. Essas mudanças são devidas ao componente Nível Baixo, produzido pelas variações na distribuição do tráfego ao longo do tempo, capturado pela entropia relativa.

Neste caso, vale perceber os efeitos dos diferentes tamanhos de janela. Além das considerações feitas na Seção 4.1, desta vez quanto maior o tamanho das janelas, maior a sensibilidade da métrica para os intervalos de anomalias. Isso porque, neste caso, os tamanhos de janela investigados não foram próximos do maior intervalo anômalo introduzido. Isto reforça haver um compromisso (*tradeoff*) para a escolha do tamanho da janela. Se o tamanho da janela for muito menor do que o intervalo

de anomalia, então a métrica de confiança também perde sensibilidade, visto que em alguns momentos toda a janela deslizante estará dentro do intervalo anômalo, o que será traduzido como um novo comportamento de tráfego. Isso pode ser utilizado inclusive do ponto de vista de tráfego lícito, quando, por exemplo, um dispositivo de fato muda o comportamento de tráfego por conta de uma atualização vinda da aplicação associada. O mesmo ocorre quando o tamanho de janela se aproxima do tamanho do intervalo anômalo, como visto na Seção 4.1.

4.2.1 Comparativo entre Divergência de Kullback-Leibler e Informação Mútua

A informação mútua é outro conceito relacionado à divergência de Kullback-Leibler. Na verdade, como Príncipe [70] bem define, a informação mútua é um caso especial da entropia relativa $D_{KL}(p||q)$, com p sendo a distribuição conjunta das variáveis aleatórias e q o produto das distribuições marginais, conforme a Equação 4.1.

$$D_{KL}(p(X,Y)||p(X)q(Y)) = I(X,Y) \tag{4.1}$$

Alguns experimentos foram realizados para comparar as métricas de informação mútua e entropia relativa, de modo a estudar o comportamento de tais métricas quando utilizadas para o cálculo da confiança. Para este comparativo foram escolhidos quatro perfis de tráfego de diferentes dispositivos IoT a partir do conjunto de dados disponibilizado em [73]. Para cada perfil, foram consideradas duas possíveis abordagens para o cálculo da confiança pelo componente Nível Baixo: (i) Informa-

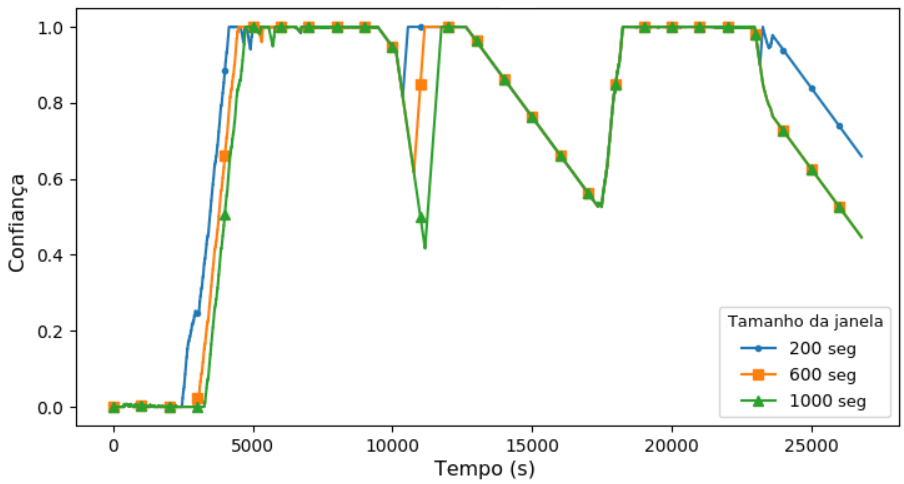
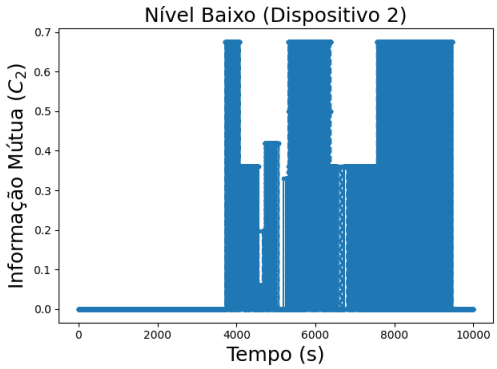
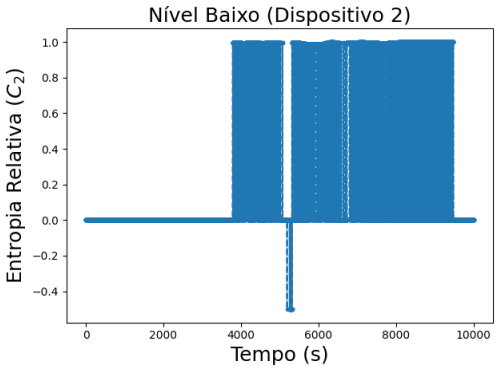


Figura 4.7: Valores de confiança obtidos com nosso modelo de confiança usando o tráfego do conjunto de dados [73]

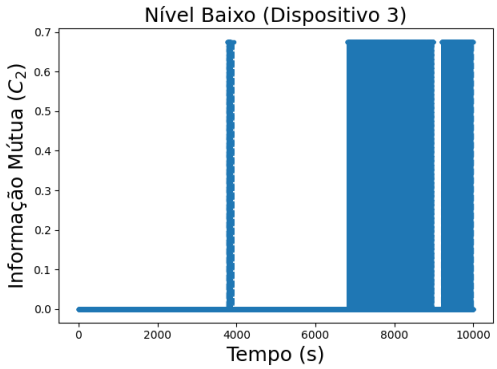


(a) Componente C_2 calculado com base na informação mútua

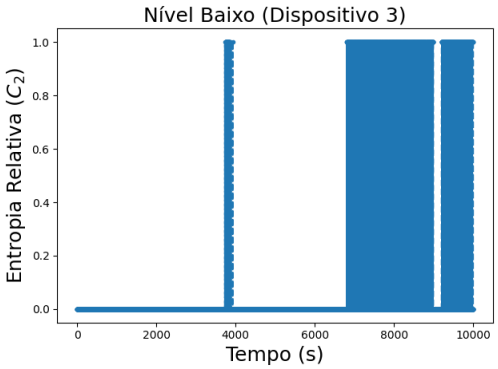


(b) Componente C_2 calculado com base na entropia relativa

Figura 4.8: Comparação do cálculo do componente C_2 para o dispositivo 2

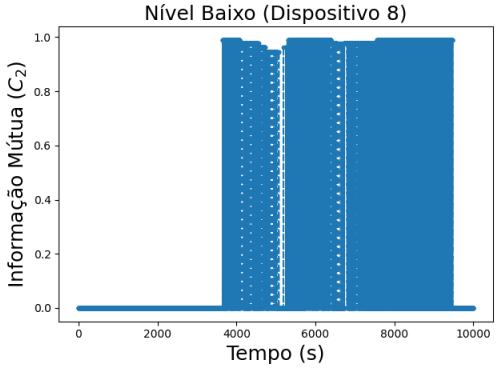


(a) Componente C_2 calculado com base na informação mútua para o tráfego do dispositivo 3

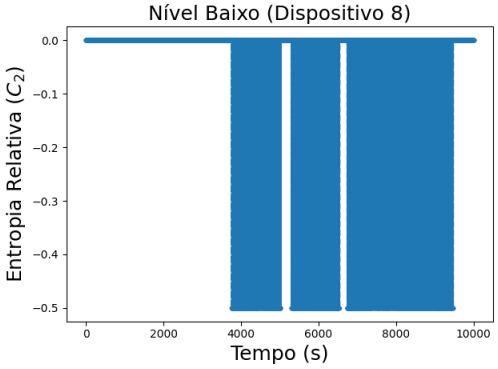


(b) Componente C_2 calculado com base na entropia relativa para o tráfego do dispositivo 3

Figura 4.9: Comparação do cálculo do componente C_2 para o dispositivo 3

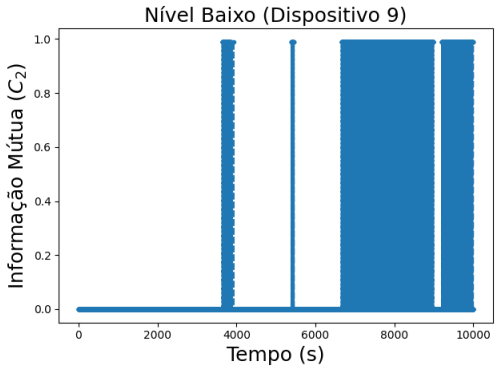


(a) Componente C_2 calculado com base na informação mútua para o tráfego do dispositivo 8

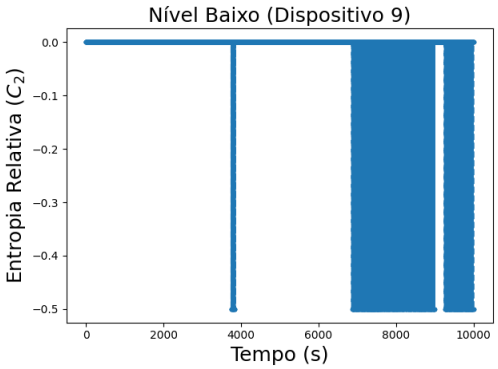


(b) Componente C_2 calculado com base na entropia relativa para o tráfego do dispositivo 8

Figura 4.10: Comparação do cálculo do componente C_2 para o dispositivo 8



(a) Componente C_2 calculado com base na informação mútua para o tráfego do dispositivo 9



(b) Componente C_2 calculado com base na entropia relativa para o tráfego do dispositivo 9

Figura 4.11: Comparação do cálculo do componente C_2 para o dispositivo 9

ção mútua entre as distribuições amostrais (estimada e observada); e (ii) Entropia relativa entre as distribuições amostrais (estimada e observada).

A métrica de confiança calculada no Nível Baixo a partir das abordagens (i) e (ii) utilizou as Equações 3.5, 3.6 e 3.7 do nosso modelo proposto no Capítulo 3. Através dos experimentos é possível observar a semelhança nos resultados das duas abordagens, conforme as Figuras 4.8, 4.9, 4.10 e 4.11, o que corrobora com as colocações de Príncipe [70], quando diz que a informação mútua é um caso especial da entropia relativa. Em particular, pode-se observar que a informação mútua não forneceu valores maiores que 1, enquanto a entropia relativa ofereceu valores abaixo e acima de 1, o que permitiu obter uma maior gama de valores de confiança. Este comportamento pode ser observado nos casos dos dispositivos 2, 8 e 9, onde a confiança é sempre incrementada na abordagem 1, enquanto sofre algumas penalizações usando a abordagem 2.

Capítulo 5

Caracterização de Tráfego e Análise dos Valores de Confiança

Neste capítulo apresentamos uma caracterização do tráfego de dispositivos IoT e observamos o impacto do uso de características de rede para calcular a métrica de confiança proposta. Para os resultados a seguir, também consideramos o conjunto de dados de tráfego entre dispositivos IoT disponibilizados por Sivanathan *et al.* [73], conforme descrito na Seção 4.2.

Analisamos a distribuição das amostras através de histogramas, a serem discutidos mais adiante, nos quais o eixo das abscissas corresponde à taxa de dados recebida por um dispositivo IoT (ou ainda, o tráfego de entrada) em *Bytes/s*, e o eixo das ordenadas corresponde à frequência relativa observada segundo o número de amostras. O eixo das abscissas varia entre os valores mínimo e máximo encontrados na faixa amostral considerada em cada caso.

Foram gerados gráficos considerando 10 *bins* e 100 *bins*, *i. e.*, 10 intervalos de mesmo comprimento que seccionam o eixo das abscissas em 10 partes iguais e 100 intervalos de mesmo comprimento que seccionam o eixo das abscissas em 100 partes iguais, respectivamente.

Obtivemos 21600 amostras de tráfego a partir de um dia do conjunto de dados. Tal quantidade corresponde a um período de coleta de cerca de 6 horas. A Figura 5.1 ilustra o histórico de tráfego original de um dispositivo IoT enviando dados para outro dispositivo ao longo do tempo durante o dia 28/09/2016 do conjunto de dados. Em particular, os dados escolhidos são provenientes de uma câmera de vídeo que apresenta cinco faixas de taxa de dados bem definidas por conta de diferentes configurações de gravação, como pode ser visto na Figura 5.1. Também consideramos uma versão modificada da mesma amostra de tráfego com a inserção de 50 KBytes/s do instante de 13000 seg a 17300 seg (Figura 5.2). Usamos esta versão modificada do tráfego para analisar o comportamento de nossa abordagem em tais circunstâncias anômalas, que podem caracterizar um comportamento de tráfego malicioso.

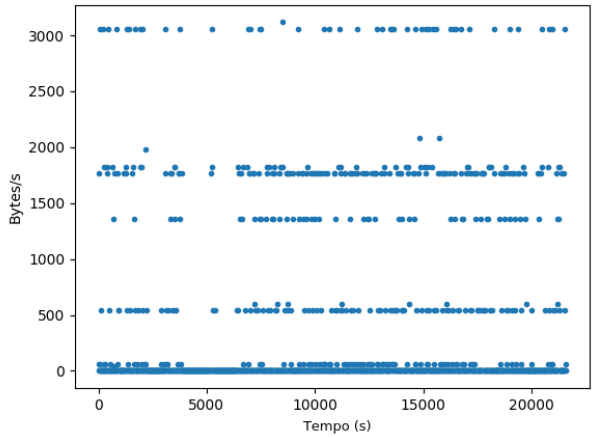


Figura 5.1: Tráfego enviado por um dispositivo usando o dia 28/09/2016 do conjunto de dados [73]

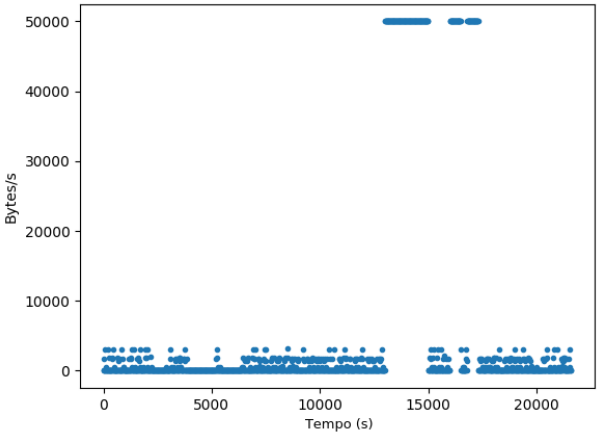


Figura 5.2: Tráfego modificado enviado por um dispositivo malicioso usando o dia 28/09/2016 do conjunto de dados [73]

Nos gráficos das Figuras 5.3 e 5.4, todas as amostras disponíveis foram usadas para calcular as frequências relativas. A maioria dos valores de tráfego apresenta relativamente taxa baixa ou zero, embora também tenhamos valores com taxas mais significativas, bastante característico de tráfego de dispositivos IoT. Comparando a Figura 5.3 com a Figura 5.4, podemos observar a inserção de valores anômalos através do aumento da frequência relativa próximo ao intervalo de 50 KBytes/s.

Para fornecer uma análise mais próxima da perspectiva de um dispositivo IoT, considerando suas restrições de recursos e a capacidade que este teria para realizar as computações necessárias, consideramos também intervalos de amostras através de uma janela deslizante ao invés de todo o conjunto de dados de uma só vez. Assim, usamos, por exemplo, o intervalo da amostra 100 a amostra 200, ou da amostra 500 a amostra 1000, sendo o tamanho de cada intervalo especificado junto à análise. Com

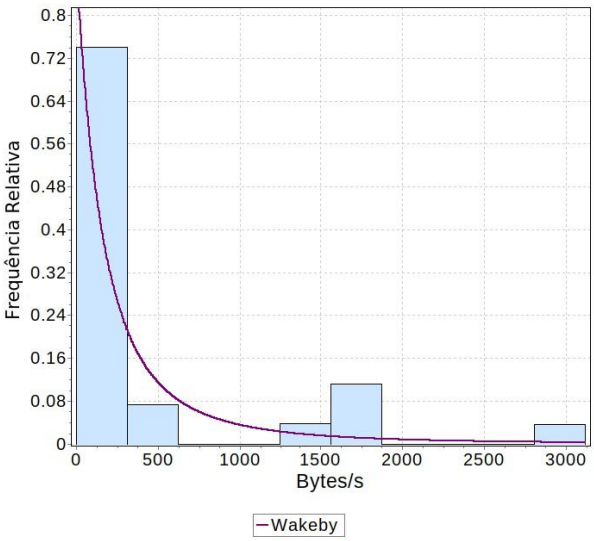


Figura 5.3: Histograma de todas as amostras do *trace* original

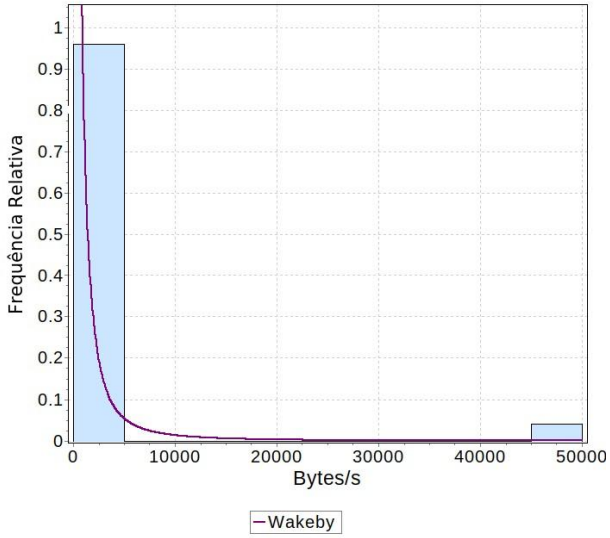


Figura 5.4: Histograma de todas as amostras do *trace* modificado

isso, analisando um intervalo menor de amostras dentro do *trace*, podemos observar que o comportamento anômalo não aparece até o momento em que o evento de anormalidade ocorra dentro da janela de observação. Na Figura 5.5, o intervalo considerado contempla as primeiras 1000 amostras e a anomalia não aparece. Por outro lado, considerando o intervalo entre 15500 e 16500 das amostras (Figura 5.6), o histograma muda significativamente, ressaltando as maiores frequências para o tráfego próximo de zero e de 50 KBytes/s.

Fazemos também um estudo sobre qual distribuição de probabilidade melhor se ajusta aos dados coletados. Considerando uma comparação entre diversas distribuições e configurações de parâmetros através do cálculo da estatística de Kolmogorov-Smirnov (K–S), podemos dizer que a distribuição Wakeby se ajustou à distribuição

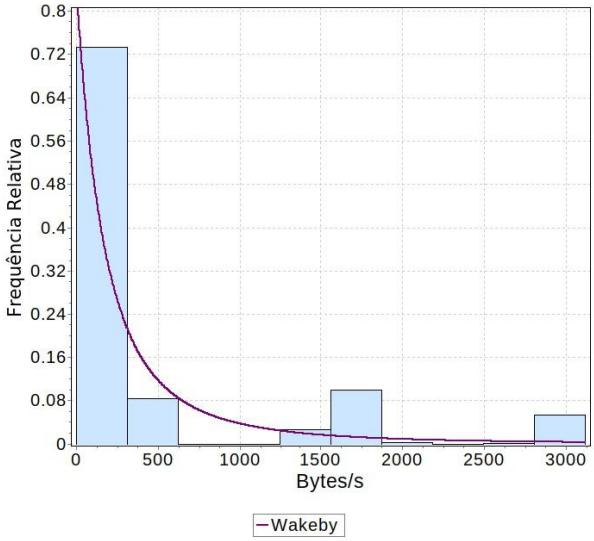


Figura 5.5: Histograma do intervalo entre 1 e 1000 das amostras

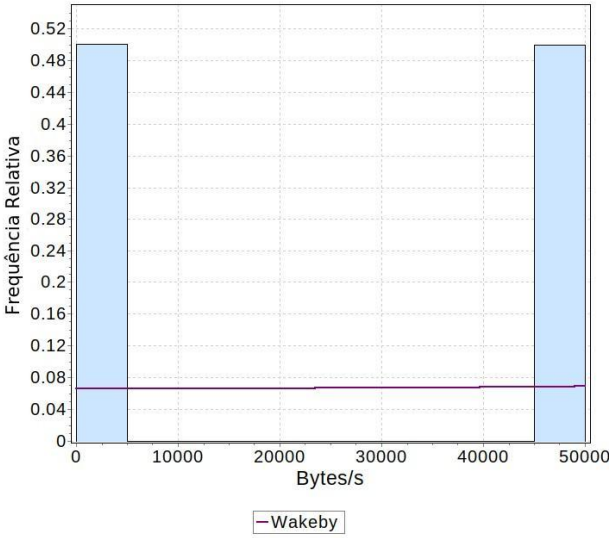


Figura 5.6: Histograma de um intervalo entre 15500 e 16500 das amostras

amostral do tráfego com a estatística K–S de $0,413$ para o *trace* original e de $0,362$ para o *trace* modificado. Para o intervalo menor em que ocorre o comportamento anômalo do tráfego, podemos perceber que a distribuição Wakeby não se ajusta bem às frequências relativas (Figura 5.6). Com isso, o dispositivo que recebe o tráfego pode calcular a divergência entre a distribuição amostral dos dados observados e a distribuição estimada (cujo tráfego, por hipótese, deve seguir uma Wakeby no exemplo). Tal divergência é dada justamente pelo cálculo da entropia relativa (ou divergência de Kullback-Leibler), que discutimos no Capítulo 3.

A Figura 5.7 mostra os resultados dos valores de confiança calculados de acordo com as Equações 3.5, 3.6 e 3.7 para o *trace* original. A confiança aumenta assim que o padrão de tráfego começa a se estabilizar e varia segundo o comportamento

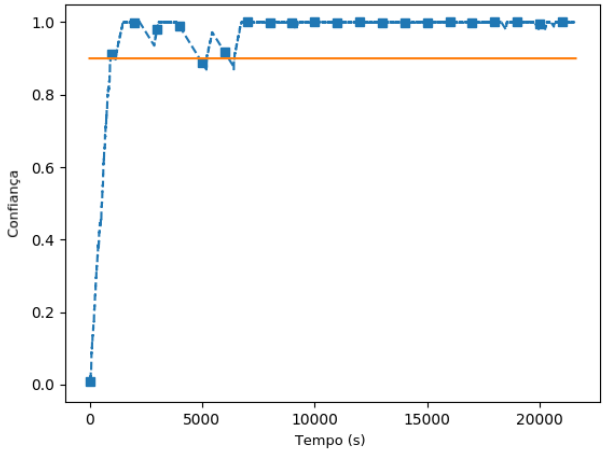


Figura 5.7: Valores de confiança obtidos com nosso modelo de confiança usando o tráfego do conjunto de dados [73]

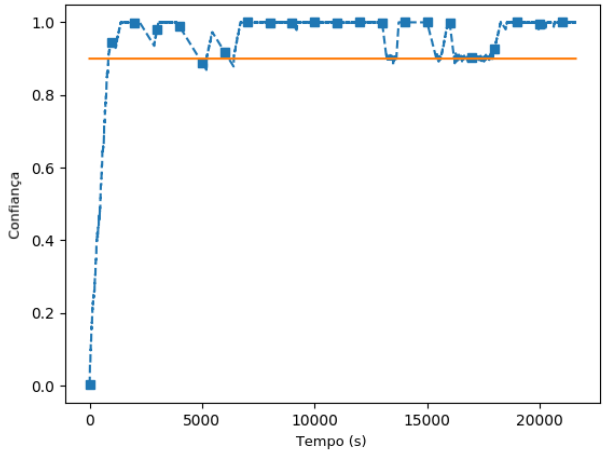


Figura 5.8: Valores de confiança obtidos com nosso modelo de confiança usando a versão modificada do tráfego do conjunto de dados [73]

do tráfego. Em seguida, os valores de confiança estabilizam a partir do instante 7000 até as últimas amostras. Porém, na Figura 5.8, quando o instante em que o comportamento anômalo inserido é alcançado (13000 seg), os valores de confiança passam a apresentar um comportamento diferente em relação ao da Figura 5.7.

Durante o intervalo de anormalidade de tráfego, os valores de confiança são penalizados (Figura 5.8), ficando abaixo do limite estabelecido ($0,9$ no exemplo), o que indica ao dispositivo receptor que o emissor não se mostra confiável. Desta maneira, a comunicação entre tais dispositivos passa a não ser mais permitida. O dispositivo receptor passa então a computar a confiança no outro através da reputação fornecida pelo Nível Alto (estimado segundo a Equação 3.1). Então, conforme o período anômalo passa, os valores de confiança começam a aumentar novamente. Essas mu-

danças são produzidas pelas variações na distribuição do tráfego ao longo do tempo, capturadas pela entropia relativa e componentes temporais (o Nível Baixo de nossa proposta), que não poderiam ser percebidas de uma perspectiva de Nível Alto.

Diante desses resultados, nossa abordagem se mostra capaz de capturar mudanças nos padrões de tráfego dos dispositivos através da análise das respectivas frequências relativas, o que permite compor uma métrica de confiança. Desta maneira, nossa abordagem impede efetivamente que dispositivos maliciosos se comuniquem com outros dispositivos na rede, o que potencialmente inviabiliza ataques que se utilizem de dispositivos vulneráveis, como ataques de negação de serviço distribuído, por exemplo.

Capítulo 6

Avaliação da Abordagem Integrada Segundo o Tempo de Contato entre Dispositivos

Neste capítulo, avaliamos o potencial de nossa abordagem em compor uma métrica de confiança, considerando o Nível Baixo baseado nas variações de comportamento do tráfego dos dispositivos IoT e o Nível Alto com as informações de reputação, com base em uma de nossas contribuições de artigo científico a ser publicado [74].

6.1 Definição de Tempo de Contato

Para observar o comportamento da abordagem, definimos a métrica de tempo de contato de acordo com as seguintes equações. Seja NS o número de amostras no experimento. Seja NW (Equação 6.1) uma função binária que assume 1 (um) quando um dispositivo i deseja transmitir para o dispositivo j, ou seja, o tráfego observado é positivo; ou 0 (zero) caso contrário.

$$NW = \begin{cases} 1, & \text{if } y > 0, \quad y \in S \quad \text{(cf. Capítulo 3)} \\ 0, & \text{caso contrário} \end{cases} \tag{6.1}$$

Seja também NT (Equação 6.2) outra função binária que assume 1 (um) quando o valor de confiança calculado TR_{ji} é maior ou igual ao limite estabelecido, caso em que a comunicação é permitida; ou 0 (zero) caso contrário.

$$NT = \begin{cases} 1, & \text{if } TR_{ji} \geqslant \text{limiar} \\ 0, & \text{caso contrário} \end{cases} \tag{6.2}$$

Com isso, definimos $TempoContato$ como a soma descrita na Equação 6.3. Assumimos que pelo menos uma amostra apresentará intenção de transmitir, ou seja, $y > 0$, portanto, o denominador na Equação 6.3 é sempre maior que zero.

$$TempoContato = \frac{\sum_{k=1}^{NS} NT_k}{\sum_{k=1}^{NS} NW_k}, \qquad \sum_{k=1}^{NS} NW_k > 0 \tag{6.3}$$

Assim, o tempo de contato é a fração de tempo durante a qual a confiança em um dispositivo permanece acima de um certo limite durante o experimento. Durante esse tempo, os dispositivos podem se comunicar (estabelecer contato). Caso contrário, quando os valores de confiança caem abaixo do limite, não há mais contato entre os dispositivos. O limite depende da aplicação em questão, que pode aceitar valores de confiança mais baixos (menos restritos) ou apenas valores de confiança mais altos (mais restritos).

6.2 Experimentos

Para executar os experimentos elaborados nós usamos *traces* reais obtidos a partir do conjunto de dados encontrado em [73], a fim de validar a proposta com dados provenientes de uma aplicação IoT real. Vale ressaltar que não estamos simulando valores para a métrica de confiança, mas sim de fato calculando tal métrica conforme o modelo proposto no Capítulo 3, sob um cenário usando dados reais. Supomos que todo o dinamismo típico do contexto de IoT seja refletido nas amostras do conjunto de dados. Por exemplo, a interrupção da conectividade devido à mobilidade faz com que a taxa de dados recebida por um dispositivo seja zero. Neste caso, a distribuição da taxa de dados inclui tal comportamento.

Conforme enunciamos na Seção 4.2, o conjunto de dados em [73] é composto por *traces* de tráfego obtidos em um ambiente de *campus* universitário inteligente com mais de 20 dispositivos IoT, incluindo câmeras, luzes inteligentes, sensores de atividade e monitores de saúde. Esses *traces* são armazenados em arquivos de fluxos do tipo *pcap* com um período de coleta de três semanas. Consideramos o período de um dia do conjunto de dados nos experimentos e extraímos o tráfego em *Bytes/s* dos fluxos para cada par de dispositivos de acordo com a tupla (IP de origem, IP de destino) somando a quantidade de *Bytes* transmitidos em um segundo. Esperamos que nossa abordagem possa detectar possíveis mudanças no padrão de tráfego e seja capaz de ajustar os valores da métrica de confiança dos respectivos dispositivos IoT, ora aumentando a confiança se o padrão de tráfego observado for conforme o estimado, ora penalizando-a caso contrário.

Nos experimentos, consideramos os fluxos entre quaisquer dois dispositivos do conjunto de dados. Ao longo do texto, usamos dispositivo i e dispositivo j para nos referirmos a tais dispositivos. Os experimentos consistem em reproduzir os valores de tráfego obtidos do conjunto de dados e calcular a métrica de confiança segundo as Equações 3.5, 3.6 e 3.7. Consideramos assim as seguintes premissas de implementação dos experimentos:

- Os resultados foram obtidos considerando o tamanho da janela deslizante com 600 segundos, que contém os valores de tráfego usados na distribuição de tráfego estimada;
- Para obter a distribuição amostral das taxas de dados, calculamos as frequências relativas considerando 10 intervalos de tamanho $\Delta = 10000$ e definimos o valor máximo de taxa de dados de maneira simplificada com mesmo valor para todos os dispositivos, com $R_{ji}^{max} = 100\ KBps$. Desta forma, o espaço amostra fica definido como $\mathcal{S}_{ji} = [0, 10KBps, 20KBps, ..., 100KBps]$;
- O valor do tráfego esperado usado para comparar com o valor do tráfego realmente recebido é calculado a partir de um Filtro de Kalman com $média = 0$ e $covariância = 1$, visto que não requer muitos recursos;
- Os valores de confiança iniciais do Nível Alto são obtidos sinteticamente a partir de uma distribuição Gaussiana com os parâmetros $\mu = 1$ e $\sigma^2 = 1$ para ter o número de confirmações normalmente distribuídas.

Os valores de confiança iniciais são fornecidos pelo Nível Alto sempre que um par de dispositivos deseja se comunicar e não possuem histórico de interações entre eles, ou seja, sempre que não há informações suficientes para inferir a confiança de um determinado dispositivo, a reputação deste respectivo dispositivo é obtida a partir do Nível Alto. Particularmente, as consultas à cadeia de blocos do Nível Alto são feitas em duas situações: (i) no início de uma nova comunicação; ou (ii) quando o valor de confiança fica abaixo do limite estabelecido (vide Figura 3.2).

Variamos o número de pares de dispositivos até 4 pares (total de 8 dispositivos), cada par com um padrão de tráfego diferente, como pode ser visto na Figura 6.1. Os dispositivos 1 e 2 são rotulados como dispositivos lícitos apresentando padrões de tráfego conforme os estimados, enquanto os dispositivos 3 e 4 são rotulados como dispositivos maliciosos com distribuições de tráfego diferente das estimadas.

A Figura 6.2 mostra os respectivos valores de confiança calculados para cada padrão de tráfego considerado. Para os dispositivos 1 e 2, os valores de confiança nos primeiros segundos sobem gradativamente conforme os valores obtidos no Nível Alto (Figura 6.3) de 0 (zero) para $0,8$, sendo esta a confiança mínima predefinida

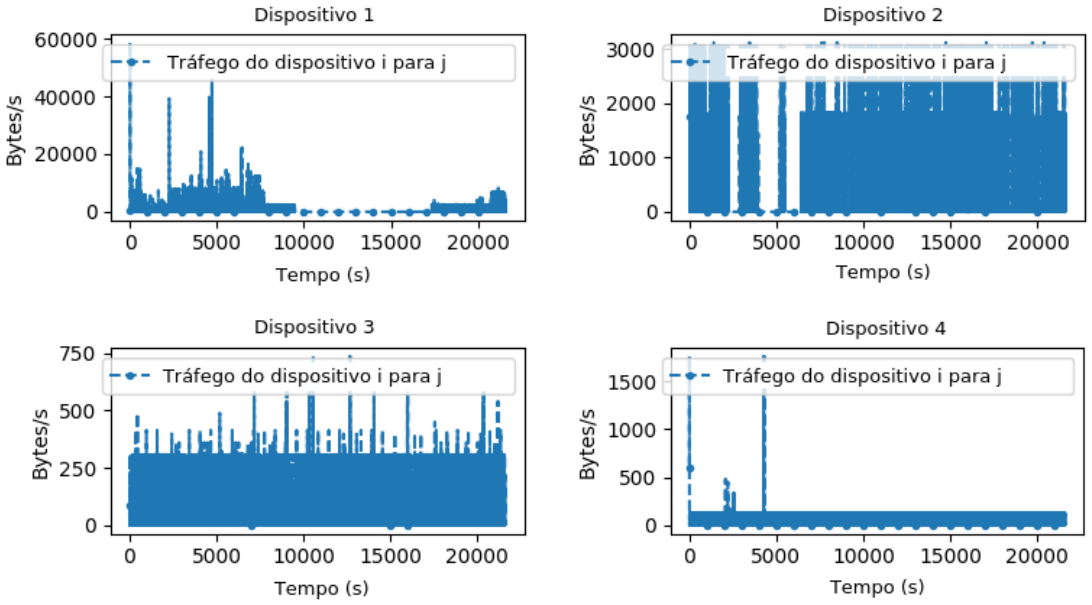


Figura 6.1: *Traces* de tráfego do conjunto de dados [73] para cada dispositivo remetente i. Os dois primeiros gráficos (tráfego dos dispositivos 1 e 2) são *traces* de dispositivos lícitos, enquanto os dois últimos gráficos (tráfego dos dispositivos 3 e 4) vêm de dispositivos maliciosos. O tráfego é dado em Bytes/s durante um período de 21.600 segundos (6 horas)

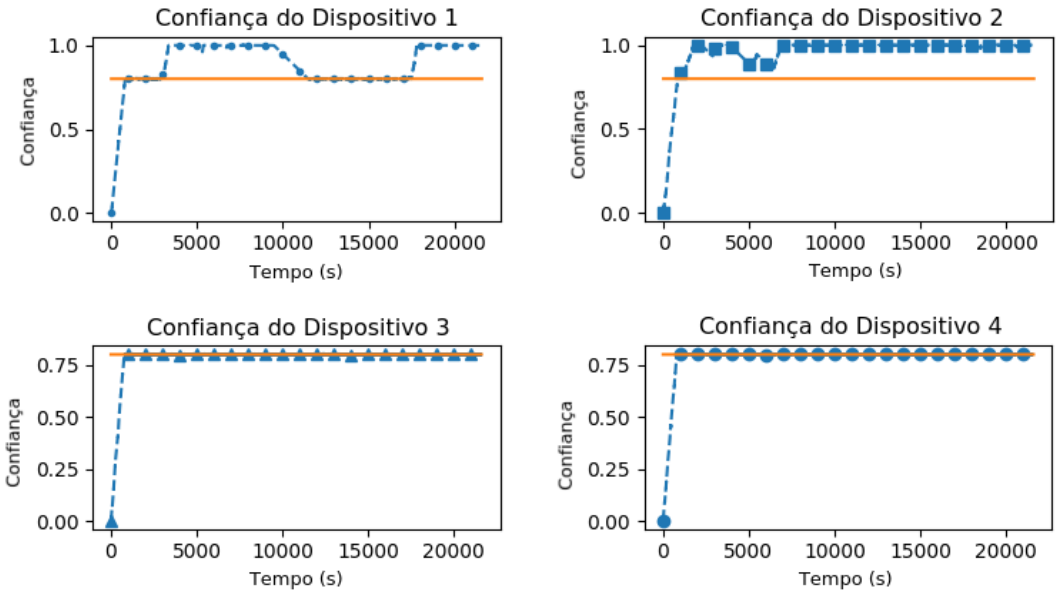


Figura 6.2: Valores de confiança calculados ao longo do tempo para cada dispositivo de acordo com os respectivos padrões de tráfego mostrados na Figura 6.1

para os dispositivos do experimento. Como os dispositivos alcançaram um valor de confiança acima do limite mínimo estabelecido, a comunicação entre os dispositivos 1 e 2 é permitida e o Nível Baixo começa a atuar.

Já os dispositivos 3 e 4 não conseguiram alcançar valores de confiança suficien-

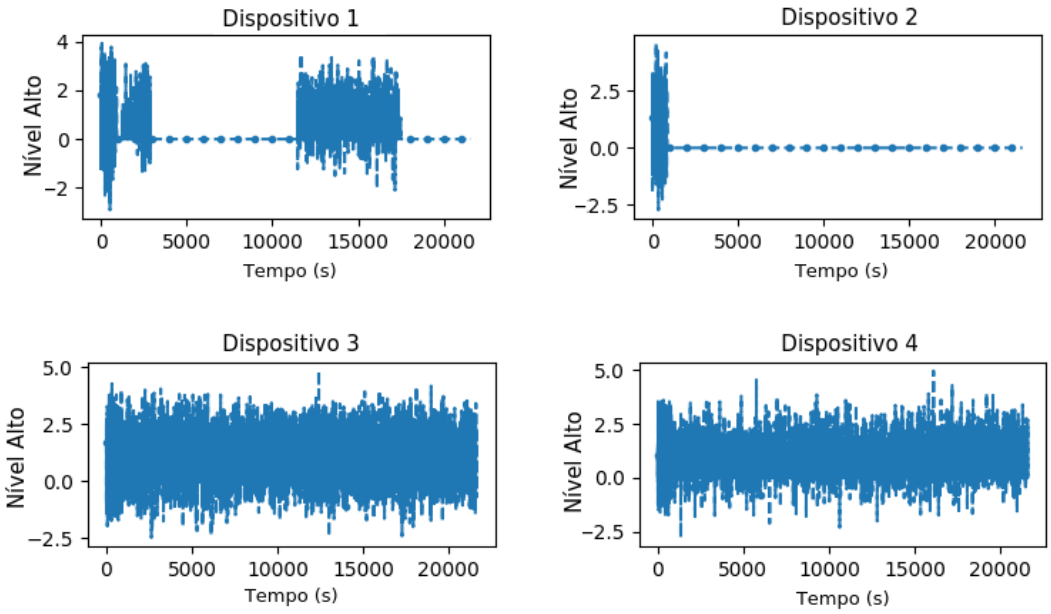


Figura 6.3: Pontuação da cadeia de blocos do Nível Alto quando questionado por cada dispositivo receptor. Observe que a pontuação de C_1 só é consultada nos casos em que os valores de confiança do dispositivo remetente permanecem abaixo do limite

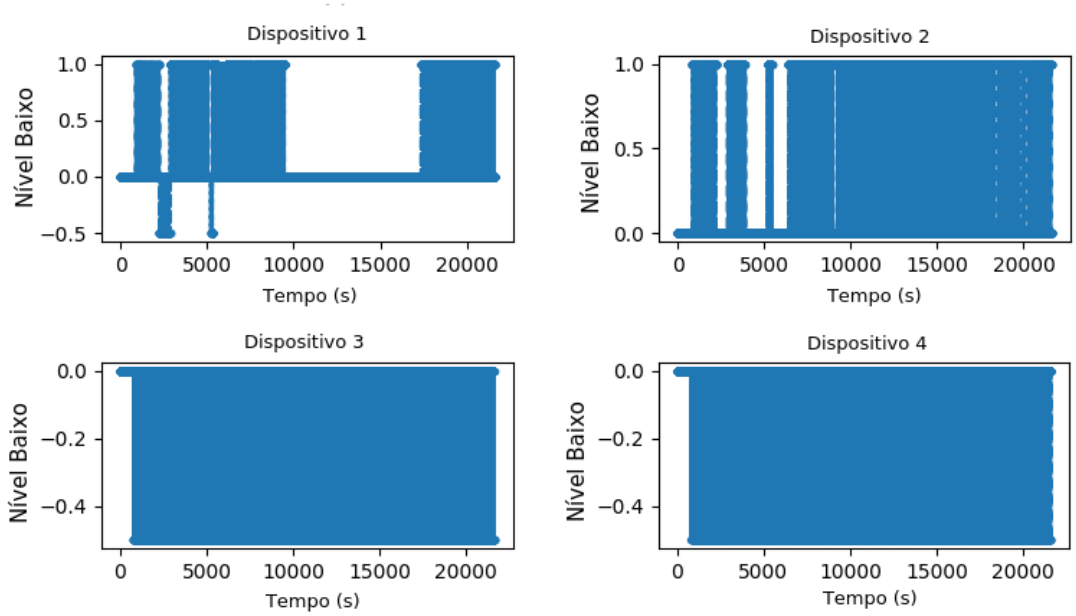


Figura 6.4: Pontuação de entropia relativa no Nível Baixo calculada por cada dispositivo receptor. Observe que a pontuação de C_2 é calculada apenas nos casos em que os valores de confiança permanecem acima do limite

tes para ultrapassar o limite predefinido, o que não permite que tais dispositivos se comuniquem com seus respectivos pares. Desta forma, os outros dispositivos lícitos conectados à rede são de fato protegidos dos dispositivos maliciosos, uma vez que

Tabela 6.1: Parâmetros dos experimentos

Parâmetro	**Valor**
Número de amostras de tráfego	21600
Número de dispositivos IoT	8 (4 pares); 60 (30 pares)
Tamanho da janela deslizante	600 s
Fração de dispositivos maliciosos	10%, 50%, 90%
Componente *Nível Alto*	$N(1, 1)$
Limiar de confiança	0.8
Componente de decaimento temporal	-0.1 confiança/s
Tráfego estimado	Filtro de Kalman ($\mu = 0$, $cov = 1$)
Tamanho do intervalo (Δ)	10
Taxa de dados máxima (R_{ji}^{max})	100 KBps
Número de intervalos	10

estes não conseguem estabelecer comunicação, sendo isolados da rede, o que potencialmente evita, por exemplo, uma possível propagação de *software* malicioso, ou inviabiliza um ataque de negação de serviço (*Denial of Service – DoS*) . A Tabela 6.1 resume os parâmetros usados no experimento.

Os comportamentos de Nível Alto e Nível Baixo são mostrados nas Figuras 6.3 e 6.4 de acordo com cada par de dispositivos no experimento. Na Figura 6.3, observamos as consultas ao Nível Alto para obter os valores de confiança iniciais quando estes estão abaixo do limite. Novamente, isso significa que o Nível Alto é consultado apenas quando o Nível Baixo não pode ser aplicado. O oposto ocorre para o Nível Baixo, como pode ser observado na Figura 6.4. A pontuação de entropia relativa é calculada pelo dispositivo receptor quando os valores de confiança estão acima do limite, momento em que os dispositivos estão efetivamente se comunicando, o que significa haver tráfego para calcular a entropia relativa. Por exemplo, para o dispositivo 2, o Nível Alto apresenta valores apenas no início quando os dispositivos começam a estabelecer a conectividade, então o Nível Baixo passa a valer e se mantém ativo até o final do experimento, dado que os valores de confiança permaneceram acima do limite. Por outro lado, para o dispositivo 4 (rotulado como malicioso), o Nível Alto é constantemente consultado e o Nível Baixo retorna apenas valores negativos (penaliza o valor de confiança), uma vez que a distribuição do tráfego enviado pelo dispositivo malicioso difere da distribuição estimada pelo dispositivo receptor.

Na Figura 6.5, mostramos o tempo médio de contato para um par de até quatro pares de dispositivos. Durante os primeiros dois conjuntos de experimentos (um e dois pares), apenas os dispositivos lícitos são considerados, portanto, o tempo de contato permanece alto. Para os conjuntos a seguir, com 3 e 4 pares, vemos que o tempo médio de contato é reduzido e não aumenta, o que confirma que a introdução de dispositivos maliciosos teve um impacto no tempo de contato, e também revela a eficácia da abordagem de confiança em conter dispositivos maliciosos na rede.

Também analisamos nossa abordagem considerando uma configuração com um grande número de dispositivos de acordo com a Tabela 6.1. Variamos o número de pares de 1 a 30 e, com 30 pares de dispositivos, ou seja, 60 dispositivos no total, obtivemos um tempo de contato de $0,84$ para uma configuração de rede com 10% de dispositivos maliciosos. Com 50% de dispositivos maliciosos, obtivemos $0,73$ de tempo de contato e $0,58$ com 90% de dispositivos maliciosos. Quanto maior a taxa de dispositivos maliciosos, menor é o tempo de contato, o que significa que tais dispositivos maliciosos não conseguem se comunicar. Isso confirma que nossa abordagem evita que possíveis ataques à segurança sejam bem-sucedidos. A Figura 6.6 ilustra o tempo de contato de acordo com o número de pares e a taxa de dispositivos maliciosos.

Considerando a matriz de confusão descrita na Tabela 6.2, temos os quatro casos de classificação possíveis, tomando como verdadeiro positivo (VP) quando um dispositivo malicioso tenta estabelecer comunicação, mas é impedido por não ter alcançado o mínimo de confiança estabelecido pelo limiar. O falso positivo (FP) acontece quando um dispositivo lícito tenta estabelecer comunicação com outro dispositivo, mas é incorretamente impedido de fazê-lo por falta de confiança. Já o falso negativo (FN) ocorre quando um dispositivo malicioso consegue indevidamente estabelecer comunicação, dado que adquiriu confiança necessária para isto. Por fim, o verdadeiro negativo (VN) acontece quando um acertadamente um dispositivo lícito consegue transmitir, visto que adquiriu confiança suficiente. Com isso, calculamos as métricas de acurácia, precisão (também chamada de *positive prediction value – PPV*) e cobertura (também chamada de *recall* ou *true positive rate – TPR*), conforme as Equações 6.4 e 6.5, respectivamente.

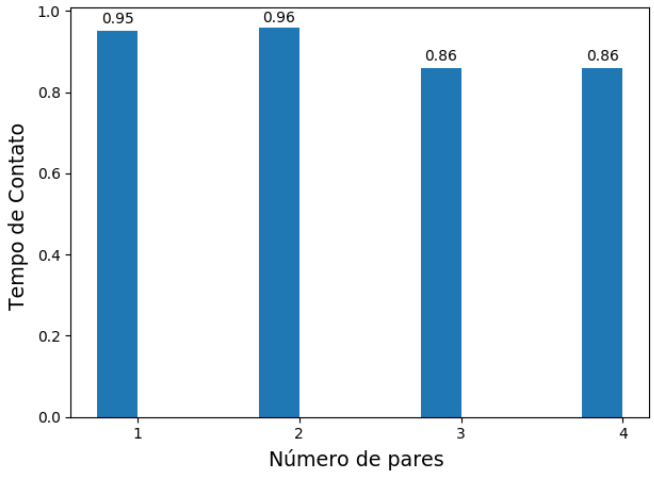


Figura 6.5: Histograma do tempo de contato conforme o número de pares de dispositivos aumenta

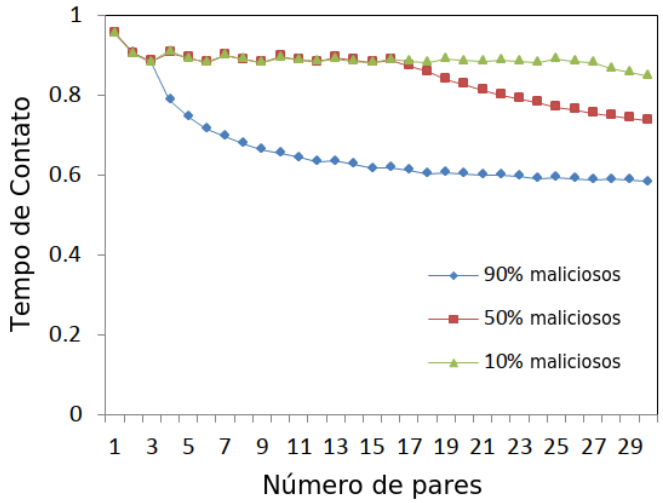


Figura 6.6: Variação do tempo de contato de acordo com o número de pares de dispositivos e a taxa de dispositivos maliciosos

$$Acurácia = \frac{VP + VN}{VP + VN + FP + FN} \tag{6.4}$$

$$Precisão = \frac{VP}{VP + FP}; \quad Cobertura = \frac{VP}{VP + FN} \tag{6.5}$$

Para o caso de 10% de dispositivos maliciosos na rede e obtivemos $0,84$ de precisão, com precisão de $0,94$ e $0,88$ de *recall*. Com 50% de dispositivos maliciosos, obtivemos $0,70$ de precisão, $0,67$ de precisão e $0,89$ de *recall*. Com 90% de dispositivos maliciosos, obtivemos $0,5$ de precisão, $0,15$ de precisão e $0,88$ de *recall*. A exatidão e a precisão diminuem com o aumento do número de dispositivos maliciosos, visto que nos concentramos em classificar os dispositivos lícitos e não os maliciosos. Portanto, as amostras lícitas tornam-se mais raras, o que reduz o número de verdadeiros positivos e falsos negativos. Apesar das variações apresentadas na exatidão e precisão, o *recall* permaneceu estável, indicando que a abordagem tem poucos casos de falso negativo, quando um dispositivo lícito deseja estabelecer comunicação, mas é impedido pela métrica de confiança que ficou abaixo do limiar. A Tabela 6.3 resume os indicadores de desempenho obtidos.

Todos os experimentos desta tese foram escritos usando as linguagens de programação Python 3.6 e ShellScript, e implementados em um computador equipado com CPU Intel Core i5 de 8ª geração e 8 GB de RAM.

Tabela 6.2: Matriz de confusão para avaliação da métrica de confiança proposta

Métrica de confiança	Dispositivo malicioso	Dispositivo lícito
Abaixo do Limiar	VP	FP
Acima do Limiar	FN	VN

Tabela 6.3: Indicadores de desempenho para diferentes porcentagens de dispositivos maliciosos

% de dispositivos maliciosos	Acurácia	Precisão	Cobertura (*Recall*)
10%	0.84	0.94	0.88
50%	0.7	0.67	0.89
90%	0.5	0.15	0.88

Capítulo 7

Conclusões e Sugestões para Pesquisas Futuras

Este capítulo apresenta as considerações finais sobre esta tese, destacando as contribuições e inovações que o presente trabalho trouxe para a linha de pesquisa abordada, além de apontar possíveis pontos de limitação do trabalho. Ademais, o capítulo traz potenciais vertentes de novos trabalhos inspirados nos resultados obtidos por esta pesquisa que podem ser explorados no futuro.

7.1 Resumo e Contribuições

Nesta tese apresentamos uma proposta de modelagem matemática para o conceito de confiança considerando o contexto de comunicação entre dispositivos da IoT. Abordar o conceito de confiança é um desafio, além de ser um tema extremamente relevante perante a comunidade científica, visto que se trata de um conceito multidimensional, encontrado em trabalhos científicos com diversas definições e aplicações em diferentes sistemas [15, 20, 35–37, 67]. Além disso, os aspectos de confiança são essenciais para a aceitação do paradigma de SIoT que vem se consolidando ao longo dos últimos anos. Com isso, consideramos a hipótese de ser possível definir uma métrica de confiança para o contexto de IoT.

Através de experimentos usando dados sintéticos e reais, conseguimos contribuir para a área de estudo considerada respondendo às questões de pesquisa levantadas no Capítulo 1:

- Como fornecer uma métrica de confiança na comunicação entre dispositivos IoT de modo a proteger os dispositivos lícitos de uma rede e isolar outros dispositivos que de alguma maneira tiveram sua segurança comprometida, sendo potencialmente maliciosos?
- Como quantificar e modelar matematicamente o conceito de confiança?

Através do modelo apresentado no Capítulo 3, respondemos tais questões de maneira alinhada com o ciclo de vida do processo de gerenciamento de confiança considerado [67]. Posteriormente, apresentamos um conjunto de experimentos que validam a proposta, bem como mostram o funcionamento e os indicadores de desempenho associados. Tais experimentos contemplam o estudo do comportamento da métrica de confiança proposta sob diferentes condições de tráfego de dispositivos, usando tanto tráfego sintético, quanto baseado em dados reais de aplicações IoT em um ambiente de *campus* universitário [73], como visto no Capítulo 4.

Também fizemos experimentos fornecendo a caracterização do tráfego de dispositivos IoT e a análise do tempo de contato entre estes nos Capítulos 5 e 6, respectivamente. Além disso, oferecemos os resultados de uma revisão sistemática da literatura, expandida com a adição de novos trabalhos relacionados no Capítulo 2, assim como uma comparação qualitativa entre os trabalhos encontrados e a nossa proposta.

Nossa proposta contribui para o avanço da área de pesquisa estudada considerando o uso de uma abordagem de dois níveis, reunindo de modo abrangente dados das camadas de rede e de aplicação, respectivamente para o Nível Baixo e o Nível Alto de nossa abordagem.

A principal contribuição desta tese consiste, sobretudo, em combinar as características da perspectiva de rede (Nível Baixo) através do uso da entropia relativa do fluxo de entrada de dados de um dispositivo IoT; e as características da perspectiva de aplicação (Nível Alto) com uma abordagem baseada em registro distribuído (cadeia de blocos) que fornece a reputação das identidades associadas aos dispositivos. Desta maneira, o modelo proposto consegue capturar mudanças no comportamento do tráfego dos dispositivos e ajustar os valores de confiança relacionados, sendo capaz de isolar os dispositivos que apresentam algum tipo de comportamento inesperado, podendo ser potencialmente maliciosos, desta forma, protegendo os dispositivos lícitos.

Em resumo, esta tese apresenta uma proposta inovadora com as seguintes contribuições:

- A abordagem de dois níveis para fornecer uma métrica de confiança baseada em informações provenientes tanto de características de rede, quanto de aplicação;
- O uso de uma estratégia baseada em livro-razão distribuído, mais especificamente cadeia de blocos, para fornecer um valor de confiança inicial e informações sobre características de aplicação, usado no *Nível Alto*;
- O uso de Teoria da Informação através do conceito de entropia relativa para fornecer confiança considerando a dinâmica das características do tráfego de rede dos dispositivos IoT, utilizado no *Nível Baixo*;

- A caracterização do tráfego de um dispositivo IoT e ajuste de distribuição;
- A quantificação e a modelagem matemática para o conceito de confiança.

7.1.1 Publicações da Tese

Dentre as contribuições alcançadas ao longo do doutorado e apresentadas nesta tese, destacamos os seguintes trabalhos:

- MACEDO, E. L. C., DE OLIVEIRA, E. A. R., SILVA, F. H., et al. "On the security aspects of Internet of Things: A systematic literature review", Journal of Communications and Networks, v. 21, n. 5, pp. 444–457, 2019. DOI: 10.1109/JCN.2019.000048;
- MACEDO, E. L. C., SILVA, R. S., DE MORAES, L. F. M., et al. "Trust Aspects of Internet of Things in the Context of 5G and Beyond". In: 4th Conference on Cloud and Internet of Things (CIoT), pp. 59–66, 2020. DOI: 10.1109/CIoT50422.2020.9244297;
- MACEDO, E. L. C., DELICATO, F. C., DE MORAES, L. F. M., FORTINO, G. "Assigning Trust to Devices in the Context of Consumer IoT Applications", Aceito para publicação no IEEE Consumer Electronics Magazine, 2022. DOI: 10.1109/MCE.2022.3154357;
- MACEDO, E. L. C., DELICATO, F. C., DE MORAES, L. F. M., FORTINO, G. "A Two-Level Integrated Approach for Assigning Trust Metrics to Internet of Things Devices", Aceito para publicação no 7th International Conference on Internet of Things, Big Data and Security (IoTBDS), 2022.

7.2 Limitações

Uma possível limitação de nossa abordagem é quando um dispositivo impõe um comportamento de tráfego específico para aumentar sua reputação perante os outros. No entanto, mesmo que alguns dispositivos possam ficar comprometidos por um tempo devido ao comportamento de tráfego forjado, o Nível Alto de nossa abordagem visa lidar com essa situação. Este componente atua reduzindo a confiança de tais dispositivos defeituosos ou mal-intencionados assim que o comportamento anômalo é relatado, sendo a respectiva reputação de tais dispositivos penalizada perante a comunidade de dispositivos. Desta forma, os dispositivos que apresentam comportamento anômalo são contidos.

Em outra perspectiva, um dispositivo que altera seu comportamento de rede devido à modificação lícita da aplicação associada, pode ter sua confiança impactada

enquanto o novo comportamento não é atualizado nos outros dispositivos. Além disso, é necessário fornecer inicialmente aos dispositivos os respectivos valores mínimos de confiança (limiar) que estes devem tolerar para que a comunicação seja estabelecida. Assim, nossa abordagem exige que uma configuração inicial seja feita nos dispositivos para estes possam operar de maneira adequada.

7.3 Sugestões para Pesquisas Futuras

Nesta tese, a abordagem em dois níveis para construção de uma métrica de confiança entre dispositivos IoT se mostrou uma técnica eficaz para proteger os dispositivos lícitos dos potencialmente maliciosos e evitar que determinados tipos de ataques se propaguem pela rede.

Como extensão deste trabalho, sugerimos a investigação de possíveis usos da métrica de confiança em outros contextos, como, por exemplo, nos fluxos de dados de um *backbone* de um provedor de serviços. Em particular, podemos citar a RedeRio/FAPERJ, cujos dados trafegados pelos roteadores de borda são coletados e analisados no Projeto IPTraf [75], cujo sistema de coleta se encontra disponível no Laboratório Ravel da COPPE/UFRJ [76]. Desta maneira, pode ser possível a identificação de membros da rede maliciosos, podendo ser aplicado a redes de provedores de maneira geral.

Além disso, é interessante investigar oportunidades de uso de mecanismos e modelos de aprendizado de máquina para o cálculo da estimativa de tráfego a ser recebido, o qual nesta tese é feito através de um Filtro de Kalman. Ao mesmo tempo, sugere-se o estudo dos possíveis benefícios que mecanismos de aprendizado de máquina poderiam obter através da aplicação da métrica de confiança proposta.

O componente Nível Baixo da proposta se mostrou eficaz em perceber as mudanças de comportamento do tráfego dos dispositivos, servindo como uma boa forma de ajustar os valores de confiança. Sugerimos como próximos trabalhos o estudo deste componente considerando o potencial de uso de outras medidas de entropia, conforme pode ser visto no Apêndice B, que possam também fornecer maneiras de acompanhar as variações de tráfego e ser de interesse para aplicações específicas. Além disso, como o componente Nível Baixo se vale de informações de tráfego, vale à pena investigar a criação de perfis de tráfego específicos de acordo com os domínios de aplicação de IoT, de modo a aprimorar a classificação do que é considerado tráfego lícito e tráfego anômalo, incluindo casos de dispositivos configurados com mais de um perfil de tráfego lícito.

O componente Nível Alto também se mostrou efetivo em ajustar os valores de confiança durante os períodos iniciais da comunicação entre os dispositivos, e quando estes, por algum motivo, tinham suas confianças reduzidas pelo Nível Baixo aquém

do limite preestabelecido. Uma possível vertente de trabalho promissora é experimentar outras estruturas de registro distribuído, como colocado no Apêndice D, de maneira a oferecer a reputação, não só de dispositivos IoT, mas também de outros membros da rede, como sugerido anteriormente.

Com a consolidação de SIoT, pode ser interessante também investigar como os modelos de ciência de redes (redes complexas), como, por exemplo, o modelo de Barabasi-Albert [77, 78], podem influenciar no comportamento dos dispositivos e impactar na confiança estabelecida entre estes.

Por fim, esta tese considerou o uso de fluxos de dados sintéticos e reais. Seria interessante engendrar um ambiente com dispositivos reais em um *testbed* [79–81], em que, não só a proposta desta tese possa se valer de tal infraestrutura para suas validações e aferições em um ambiente praticamente real, mas também outras propostas de outros problemas que venham ser elaborados.

Referências Bibliográficas

[1] ROBERTS, L. "The Arpanet and computer networks". In: *A history of personal workstations*, Association for Computing Machinery, pp. 141–172, New York, NY, USA, 1988.

[2] KLEINROCK, L. *Communication Nets; Stochastic Message Flow and Delay*. USA, Dover Publications, Inc., 1972. ISBN: 0486611051.

[3] SHANNON, C. E. "A Mathematical Theory of Communication", *Bell System Technical Journal*, v. 27, n. 3, pp. 379–423, 1948.

[4] ATZORI, L., IERA, A., MORABITO, G. "The Internet of Things: A survey", *Computer Networks*, v. 54, n. 15, pp. 2787 – 2805, 2010. ISSN: 1389-1286. doi: http://dx.doi.org/10.1016/j.comnet.2010.05.010. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1389128610001568>.

[5] PIRES, P. F., DELICATO, F. C., BATISTA, T., et al. "Plataformas para a Internet das Coisas". In: Villaça, R. (Ed.), *Minicursos SBRC 2015*, Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos (SBRC), cap. 3, pp. 110–169, Brasil, 2015.

[6] SINHA, S. *Global IoT market forecast*. Relatório técnico, IoT Analytics Research, Setembro 2021. Disponível em: <https://iot-analytics.com/number-connected-iot-devices>.

[7] ANDREWS, J. G., BUZZI, S., CHOI, W., et al. "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, v. 32, n. 6, pp. 1065–1082, 2014. doi: 10.1109/JSAC.2014.2328098.

[8] CASADEI, R., FORTINO, G., PIANINI, D., et al. "Modelling and simulation of Opportunistic IoT Services with Aggregate Computing", *Future Generation Computer Systems*, v. 91, pp. 252 – 262, 2019. ISSN: 0167-739X. doi: https://doi.org/10.1016/j.future.2018.09.005. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167739X18307246>.

[9] INTERNATIONAL, C. *Testing our trust: consumers and the Internet of Things 2017 review*. Relatório técnico, Consumer International, 2017.

[10] THE GUARDIAN. "DDoS attack that disrupted internet was largest of its kind in history, experts say". 2016. Disponível em: <https://www.theguardian.com/technology/2016/oct/26/ddos-attack-dyn-mirai-botnet>.

[11] DELICATO, F. C., PIRES, P. F. "Challenges in developing collaborative IoT systems". In: *2020 IEEE 6th Inter. Conf. on Collab. and Internet Comp. (CIC)*, pp. 25–33. IEEE, 2020.

[12] ABOMHARA, M., KØIEN, G. M. "Security and privacy in the Internet of Things: Current status and open issues". In: *2014 International Conference on Privacy and Security in Mobile Systems (PRISMS)*, pp. 1–8, 2014. doi: 10.1109/PRISMS.2014.6970594.

[13] SICARI, S., RIZZARDI, A., GRIECO, L., et al. "Security, privacy and trust in Internet of Things: The road ahead", *Computer Networks*, v. 76, pp. 146 – 164, 2015. ISSN: 1389-1286. doi: https://doi.org/10.1016/j.comnet.2014.11.008. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1389128614003971>.

[14] KHAN, M. A., SALAH, K. "IoT security: Review, blockchain solutions, and open challenges", *Future Generation Computer Systems*, v. 82, pp. 395–411, 2018. ISSN: 0167-739X. doi: https://doi.org/10.1016/j.future.2017.11.022. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167739X17315765>.

[15] MACEDO, E. L. C., DE OLIVEIRA, E. A. R., SILVA, F. H., et al. "On the security aspects of Internet of Things: A systematic literature review", *Journal of Communications and Networks*, v. 21, n. 5, pp. 444–457, 2019. doi: 10.1109/JCN.2019.000048.

[16] PALISZKIEWICZ, J. "Trust: A Multifaceted Notion". In: *Managing Public Trust*, pp. 9–23, Cham, Springer International Publishing, 2018. ISBN: 978-3-319-70485-2. doi: 10.1007/978-3-319-70485-2_2. Disponível em: <https://doi.org/10.1007/978-3-319-70485-2_2>.

[17] YAN, Z., ZHANG, P., VASILAKOS, A. V. "A survey on trust management for Internet of Things", *Journal of network and computer applications*, v. 42, pp. 120–134, 2014.

[18] BERTINO, E. "IoT Security A Comprehensive Life Cycle Framework". In: *2019 IEEE CIC*, pp. 196–203, 2019. doi: 10.1109/CIC48465.2019.00033.

[19] DEDEOGLU, V., JURDAK, R., PUTRA, G. D., et al. "A trust architecture for blockchain in IoT". In: *16th EAI Inter. Conf. on Mobile and Ubiquitous Sys.: Comp., Net. and Serv.*, pp. 190–199, 2019.

[20] ALTAF, A., ABBAS, H., IQBAL, F., et al. "Trust models of internet of smart things: A survey, open issues, and future directions", *Journal of Network and Computer Applications*, v. 137, pp. 93–111, 2019. ISSN: 1084-8045. doi: https://doi.org/10.1016/j.jnca.2019.02.024. Disponível em: <`https://www.sciencedirect.com/science/article/pii/S1084804519300839`>.

[21] JUNIOR, F. M. R., KAMIENSKI, C. A. "A Survey on Trustworthiness for the Internet of Things", *IEEE Access*, v. 9, pp. 42493–42514, 2021.

[22] BABAR, S., MAHALLE, P., OTHERS. "Trust Management Approach for Detection of Malicious Devices in SIoT", *Tehnički glasnik*, v. 15, n. 1, pp. 43–50, 2021.

[23] MOHAMMADI, V., RAHMANI, A. M., DARWESH, A. M., et al. "Trust-based recommendation systems in Internet of Things: a systematic literature review", *Human-centric Computing and Information Sciences*, v. 9, n. 1, pp. 1–61, 2019.

[24] GUO, J., CHEN, R., TSAI, J. J. "A survey of trust computation models for service management in internet of things systems", *Computer Communications*, v. 97, pp. 1–14, 2017.

[25] ALLADI, T., CHAMOLA, V., SIKDAR, B., et al. "Consumer IoT: Security vulnerability case studies and solutions", *IEEE Consumer Electronics Magazine*, v. 9, n. 2, pp. 17–25, 2020.

[26] ARABSORKHI, A., SAYAD HAGHIGHI, M., GHORBANLOO, R. "A conceptual trust model for the Internet of Things interactions". In: *2016 IST*, pp. 89–93, 2016. doi: 10.1109/ISTEL.2016.7881789.

[27] INTERNATIONAL, C. *Securing consumer trust in the internet of things – Principles and Recommendations*. Relatório técnico, Consumer International, 2017.

[28] ALDOWAH, H., UL REHMAN, S., UMAR, I. "Trust in IoT Systems: A Vision on the Current Issues, Challenges, and Recommended Solutions".

In: Saeed, F., Al-Hadhrami, T., Mohammed, F., et al. (Eds.), *Advances on Smart and Soft Computing*, pp. 329–339, Singapore, 2021. Springer Singapore. ISBN: 978-981-15-6048-4.

[29] LIU, L., LOPER, M. "Trust as a Service: Building and Managing Trust in the Internet of Things". In: *2018 IEEE HST*, pp. 1–6, 2018. doi: 10.1109/THS.2018.8574169.

[30] SAPUTRA, D. E. "Defining Trust in Computation". In: *2020 ICITSI*, pp. 161–166, 2020. doi: 10.1109/ICITSI50517.2020.9264918.

[31] STANKOVIC, J. A. "Research Directions for the Internet of Things", *IEEE Internet of Things Journal*, v. 1, n. 1, pp. 3–9, 2014. ISSN: 2327-4662. doi: 10.1109/JIOT.2014.2312291.

[32] AL-FUQAHA, A., GUIZANI, M., MOHAMMADI, M., et al. "Internet of Things: A Survey on Enabling Technologies, Protocols, and Applications", *IEEE Communications Surveys Tutorials*, v. 17, n. 4, pp. 2347–2376, 2015. ISSN: 1553-877X. doi: 10.1109/COMST.2015.2444095.

[33] PROKOFIEV, A. O., SMIRNOVA, Y. S., SILNOV, D. S. "The Internet of Things cybersecurity examination". In: *2017 Siberian Symposium on Data Science and Engineering (SSDSE)*, pp. 44–48, 2017. doi: 10.1109/SSDSE.2017.8071962.

[34] ZHANG, P., ZHOU, M., FORTINO, G. "Security and trust issues in Fog computing: A survey", *Future Generation Computer Systems*, v. 88, pp. 16 – 27, 2018. ISSN: 0167-739X. doi: https://doi.org/10.1016/j.future.2018.05.008. Disponível em: <`http://www.sciencedirect.com/science/article/pii/S0167739X17329722`>.

[35] FORTINO, G., MESSINA, F., ROSACI, D., et al. "ResIoT: An IoT social framework resilient to malicious activities", *IEEE/CAA Journal of Automatica Sinica*, v. 7, n. 5, pp. 1263–1278, 2020.

[36] FORTINO, G., FOTIA, L., MESSINA, F., et al. "Trust and Reputation in the Internet of Things: State-of-the-Art and Research Challenges", *IEEE Access*, v. 8, pp. 60117–60125, 2020. doi: 10.1109/ACCESS.2020.2982318.

[37] WANG, Y., VASSILEVA, J. "Trust and reputation model in peer-to-peer networks". In: *Proceedings Third International Conference on Peer-to-Peer Computing (P2P2003)*, pp. 150–157, 2003.

[38] NAKAMOTO, S. "Bitcoin: A peer-to-peer electronic cash system", *Working Paper*, 2008.

[39] CONOSCENTI, M., VETRÒ, A., MARTIN, J. C. D. "Blockchain for the Internet of Things: A systematic literature review". In: *2016 IEEE/ACS 13th International Conference of Computer Systems and Applications (AICCSA)*, pp. 1–6, 2016. doi: 10.1109/AICCSA.2016.7945805.

[40] BANERJEE, M., LEE, J., CHOO, K.-K. R. "A blockchain future to Internet of Things security: A position paper", *Digital Communications and Networks*, 2017. ISSN: 2352-8648. doi: https://doi.org/10.1016/j.dcan.2017.10.006. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2352864817302900>.

[41] FORTINO, G., MESSINA, F., ROSACI, D., et al. "Using Blockchain in a Reputation-Based Model for Grouping Agents in the Internet of Things", *IEEE Transactions on Engineering Management*, pp. 1–13, 2019. ISSN: 1558-0040. doi: 10.1109/TEM.2019.2918162.

[42] LI, W., SANTOS, I., DELICATO, F. C., et al. "System modelling and performance evaluation of a three-tier Cloud of Things", *Future Generation Computer Systems*, v. 70, pp. 104 – 125, 2017. ISSN: 0167-739X. doi: https://doi.org/10.1016/j.future.2016.06.019. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167739X16302047>.

[43] KSHETRI, N. "Can blockchain strengthen the internet of things?" *IT Professional*, v. 19, n. 4, pp. 68–72, 2017.

[44] GOODMAN, L. "Tezos a self-amending crypto-ledger", *Whitepaper*, 2014.

[45] SATO, H., KANAI, A., TANIMOTO, S., et al. "Establishing Trust in the Emerging Era of IoT". In: *2016 IEEE Symposium on Service-Oriented System Engineering (SOSE)*, pp. 398–406, 2016. doi: 10.1109/SOSE.2016.50.

[46] SFAR, A. R., NATALIZIO, E., CHALLAL, Y., et al. "A roadmap for security challenges in the Internet of Things", *Digital Communications and Networks*, v. 4, n. 2, pp. 118 – 137, 2018. ISSN: 2352-8648. doi: https://doi.org/10.1016/j.dcan.2017.04.003. Disponível em: <http://www.sciencedirect.com/science/article/pii/S2352864817300214>.

[47] LING, Z., LIU, K., XU, Y., et al. "An End-to-End View of IoT Security and Privacy". In: *GLOBECOM 2017 - 2017 IEEE Global Communications Conference*, pp. 1–7, 2017. doi: 10.1109/GLOCOM.2017.8254011.

[48] ALPHAND, O., AMORETTI, M., CLAEYS, T., et al. "IoTChain: A blockchain security architecture for the Internet of Things". In: *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, pp. 1–6. IEEE, apr 2018. ISBN: 978-1-5386-1734-2. doi: 10.1109/WCNC.2018.8377385. Disponível em: <https://ieeexplore.ieee.org/document/8377385/>.

[49] DAI, H.-N., ZHENG, Z., ZHANG, Y. "Blockchain for Internet of Things: A survey", *IEEE IoT Journal*, v. 6, n. 5, pp. 8076–8094, 2019.

[50] BERNABE, J. B., RAMOS, J. L. H., GOMEZ, A. F. S. "TACIoT: Multidimensional Trust-Aware Access Control System for the Internet of Things", *Soft Comput.*, v. 20, n. 5, pp. 1763–1779, maio 2016. ISSN: 1432-7643. doi: 10.1007/s00500-015-1705-6. Disponível em: <https://doi.org/10.1007/s00500-015-1705-6>.

[51] OGUNDOYIN, S. O., KAMIL, I. A. "A trust management system for fog computing services", *Internet of Things*, v. 14, pp. 100382, 2021. ISSN: 2542-6605. doi: https://doi.org/10.1016/j.iot.2021.100382.

[52] DE FILIPPI, P., MANNAN, M., REIJERS, W. "Blockchain as a confidence machine: The problem of trust & challenges of governance", *Technology in Society*, v. 62, pp. 101284, 2020. ISSN: 0160-791X. doi: https://doi.org/10.1016/j.techsoc.2020.101284.

[53] TANG, B., KANG, H., FAN, J., et al. "IoT Passport: A Blockchain-Based Trust Framework for Collaborative Internet-of-Things". In: *Proceedings of the 24th ACM Symposium on Access Control Models and Technologies*, SACMAT '19, p. 83–92, New York, NY, USA, 2019. Association for Computing Machinery. ISBN: 9781450367530. doi: 10.1145/3322431.3326327. Disponível em: <https://doi.org/10.1145/3322431.3326327>.

[54] CHEN, Z., TIAN, L., LIN, C. "Trust Model of Wireless Sensor Networks and Its Application in Data Fusion", *Sensors*, v. 17, n. 4, 2017. ISSN: 1424-8220. doi: 10.3390/s17040703. Disponível em: <https://www.mdpi.com/1424-8220/17/4/703>.

[55] HONGJUN, D., ZHIPING, J., XIAONA, D. "An Entropy-based Trust Modeling and Evaluation for Wireless Sensor Networks". In: *2008 International*

Conference on Embedded Software and Systems, pp. 27–34, July 2008. doi: 10.1109/ICESS.2008.31.

[56] KHAN, Z. A., ULLRICH, J., VOYIATZIS, A. G., et al. "A Trust-Based Resilient Routing Mechanism for the Internet of Things". In: *Proceedings of the 12th International Conference on Availability, Reliability and Security*, ARES '17, New York, NY, USA, 2017. Association for Computing Machinery. ISBN: 9781450352574. doi: 10.1145/3098954.3098963. Disponível em: <https://doi.org/10.1145/3098954.3098963>.

[57] CAMINHA, J., PERKUSICH, A., PERKUSICH, M. "A Smart Trust Management Method to Detect On-Off Attacks in the Internet of Things", *Security and Communication Networks*, v. 2018, pp. 1–10, 04 2018. doi: 10.1155/2018/6063456.

[58] BARBOSA, J. C. *A Proposal of Using Elliptic Curve in Identity Based Cryptography and its Application for Secure Message Exchange*. M.s. thesis, Federal University of Rio de Janeiro, 2005.

[59] YANG, L., YU, P., BAILING, W., et al. "IOT secure transmission based on integration of IBE and PKI/CA", *International Journal of Control and Automation*, v. 6, n. 2, pp. 245–254, 2013.

[60] ZHOU, B., LI, H., XU, L. "An Authentication Scheme Using Identity-based Encryption Blockchain". In: *2018 IEEE Symposium on Computers and Communications (ISCC)*, pp. 00556–00561, June 2018. doi: 10.1109/ISCC.2018.8538446.

[61] OTTE, P., DE VOS, M., POUWELSE, J. "TrustChain: A Sybil-resistant scalable blockchain", *Future Generation Computer Systems*, v. 107, pp. 770 – 780, 2020. ISSN: 0167-739X. doi: https://doi.org/10.1016/j.future.2017.08.048. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0167739X17318988>.

[62] KRAVARI, K., BASSILIADES, N. "StoRM: A social agent-based trust model for the internet of things adopting microservice architecture", *Simulation Modelling Practice and Theory*, v. 94, pp. 286 – 302, 2019. ISSN: 1569-190X. doi: https://doi.org/10.1016/j.simpat.2019.03.008. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1569190X19300322>.

[63] MACEDO, E. L. C., SILVA, R. S., DE MORAES, L. F. M., et al. "Trust Aspects of Internet of Things in the Context of 5G and Beyond". In: *2020*

4th Conference on Cloud and Internet of Things (CIoT), pp. 59–66, 2020. doi: 10.1109/CIoT50422.2020.9244297.

[64] KULLBACK, S. *Information Theory and Statistics*. New York, Wiley, 1959.

[65] SHI, W., CAO, J., ZHANG, Q., et al. "Edge Computing: Vision and Challenges", *IEEE Internet of Things Journal*, v. 3, n. 5, pp. 637–646, 2016. doi: 10.1109/JIOT.2016.2579198.

[66] ABBAS, N., ZHANG, Y., TAHERKORDI, A., et al. "Mobile Edge Computing: A Survey", *IEEE Internet of Things Journal*, v. 5, n. 1, pp. 450–465, 2018. doi: 10.1109/JIOT.2017.2750180.

[67] KHAN, W. Z., ARSHAD, Q.-U.-A., HAKAK, S., et al. "Trust Management in Social Internet of Things: Architectures, Recent Advancements, and Future Challenges", *IEEE Internet of Things Journal*, v. 8, n. 10, pp. 7768–7788, 2021. doi: 10.1109/JIOT.2020.3039296.

[68] LEON-GARCIA, A. *Probability and Random Processes For EE's (3rd Edition)*. USA, Prentice-Hall, Inc., 2007. ISBN: 0131471228.

[69] GALLAGER, R. G. *Information Theory and Reliable Communication*. New York, NY, USA, John Wiley & Sons, Inc., 1968. ISBN: 0471290483.

[70] PRINCIPE, J. C. *Information theoretic learning: Renyi's entropy and kernel perspectives*. USA, Springer Science & Business Media, 2010.

[71] COVER, T. M., THOMAS, J. A. *Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing)*. USA, Wiley-Interscience, 2006. ISBN: 0471241954.

[72] MACEDO, E. L. C., DELICATO, F. C., DE MORAES, L. F. M., et al. "Assigning Trust to Devices in the Context of Consumer IoT Applications", *IEEE Consumer Electronics Magazine*, pp. 1–1, 2022. doi: 10.1109/MCE.2022.3154357.

[73] SIVANATHAN, A., SHERRATT, D., GHARAKHEILI, H. H., et al. "Characterizing and classifying IoT traffic in smart cities and campuses". In: *2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, pp. 559–564, 2017.

[74] MACEDO, E. L. C., DELICATO, F. C., DE MORAES, L. F. M., et al. "A Two-Level Integrated Approach for Assigning Trust Metrics to Internet of Things Devices". In: *Proceedings of the 7th International Conference on*

Internet of Things, Big Data and Security - IoTBDS,. INSTICC, SciTePress, 2022.

[75] ASSIS, F., COUTINHO, M., FILHO, J. S., et al. "IPTraf: Coleta e Detecção de Anomalias em Fluxos de Rede". In: *Anais do XXVI Workshop de Gerência e Operação de Redes e Serviços*, pp. 96–109, Porto Alegre, RS, Brasil, 2021. SBC. Disponível em: <https://sol.sbc.org.br/index.php/wgrs/article/view/17188>.

[76] RAVEL, L. "Projeto IPTraf". 2022. Disponível em: https://iptraf.ravel.ufrj.br/. Acessado em Janeiro de 2022.

[77] BARABÁSI, A.-L. "Network science", *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, v. 371, n. 1987, pp. 20120375, 2013.

[78] ALBERT, R., BARABÁSI, A.-L. "Statistical mechanics of complex networks", *Reviews of modern physics*, v. 74, n. 1, pp. 47, 2002.

[79] ADJIH, C., BACCELLI, E., FLEURY, E., et al. "FIT IoT-LAB: A large scale open experimental IoT testbed". In: *2015 IEEE 2nd World Forum on Internet of Things (WF-IoT)*, pp. 459–464, 2015. doi: 10.1109/WF-IoT.2015.7389098.

[80] ADJIH, C., BACCELLI, E., FLEURY, E., et al. "FIT IoT-LAB". 2022. Disponível em: https://www.iot-lab.info/. Acessado em Janeiro de 2022.

[81] VITERBI, U. O. S. C. "A Campus-wide Internet-of-Things Testbed". 2022. Disponível em: http://cci.usc.edu/index.php/cci-iot-testbed/. Acessado em Janeiro de 2022.

[82] SCHNEIER, B. *Applied Cryptography (2Nd Ed.): Protocols, Algorithms, and Source Code in C*. New York, NY, USA, John Wiley & Sons, Inc., 1995. ISBN: 0-471-11709-9.

[83] STALLINGS, W. *Cryptography and Network Security: Principles and Practice*. 6th ed. Upper Saddle River, NJ, USA, Prentice Hall Press, 2013. ISBN: 0133354695, 9780133354690.

[84] PETITCOLAS, F. A. P. "Kerckhoffs' Principle". In: *Encyclopedia of Cryptography and Security*, pp. 675–675, Boston, MA, Springer US, 2011. ISBN: 978-1-4419-5906-5. doi: 10.1007/978-1-4419-5906-5_487. Disponível em: <https://doi.org/10.1007/978-1-4419-5906-5_487>.

[85] SHAMIR, A. "Identity-Based Cryptosystems and Signature Schemes". In: Blakley, G. R., Chaum, D. (Eds.), *Advances in Cryptology*, pp. 47–53, Berlin, Heidelberg, 1985. Springer Berlin Heidelberg. ISBN: 978-3-540-39568-3.

[86] BONEH, D., FRANKLIN, M. "Identity-based encryption from the Weil pairing". In: *Annual international cryptology conference*, pp. 213–229. Springer, 2001.

[87] SCHERTLER, M. J., BOYEN, X. "Identity-Based Cryptography Standard (IBCS) #1: Supersingular Curve Implementations of the BF and BB1 Cryptosystems". RFC 5091, dez. 2007. Disponível em: <`https://rfc-editor.org/rfc/rfc5091.txt`>.

[88] HEZAM, A., KONSTANTAS, D., MAHYOUB, M. "A Comprehensive IoT Attacks Survey based on a Building-blocked Reference Mode", *International Journal of Advanced Computer Science and Applications*, v. Vol. 9, 04 2018.

[89] SINGH, S. R., KHAN, A. K., SINGH, S. R. "Performance evaluation of RSA and Elliptic Curve Cryptography". In: *2016 2nd International Conference on Contemporary Computing and Informatics (IC3I)*, pp. 302–306, 2016. doi: 10.1109/IC3I.2016.7917979.

[90] SHARMA, C., SUNANDA. "Performance Analysis of ECC and RSA for Securing CoAP-Based Remote Health Monitoring System". In: Perez, G. M., Tiwari, S., Trivedi, M. C., et al. (Eds.), *Ambient Communications and Computer Systems*, pp. 615–628, Singapore, 2018. Springer Singapore. ISBN: 978-981-10-7386-1.

[91] KOBLITZ, N. "Elliptic Curve Cryptosystems", *Mathematics of Computation*, v. 48, n. 177, pp. 203–209, 1987. ISSN: 0025-5718.

[92] SAEKI, M. *Elliptic curve cryptosystems*. Tese de Doutorado, Citeseer, 1997.

[93] TSALLIS, C. "Possible generalization of Boltzmann-Gibbs statistics", *Journal of Statistical Physics*, v. 52, pp. 479–487, 07 1988. doi: 10.1007/BF01016429.

[94] DOS SANTOS, R. J. "Generalization of Shannon's theorem for Tsallis entropy", *Journal of Mathematical Physics*, v. 38, n. 8, pp. 4104–4107, 1997.

[95] TSALLIS, C. "Beyond Boltzmann–Gibbs–Shannon in Physics and Elsewhere", *Entropy*, v. 21, n. 7, 2019. ISSN: 1099-4300. doi: 10.3390/e21070696. Disponível em: <`https://www.mdpi.com/1099-4300/21/7/696`>.

[96] MASZCZYK, T., DUCH, W. "Comparison of Shannon, Renyi and Tsallis entropy used in decision trees". In: *International Conference on Artificial Intelligence and Soft Computing*, pp. 643–651. Springer, 2008.

[97] DE ALBUQUERQUE, M. P., ESQUEF, I. A., MELLO, A. G. "Image thresholding using Tsallis entropy", *Pattern Recognition Letters*, v. 25, n. 9, pp. 1059–1065, 2004.

[98] KHAN, R., KHAN, S. U., ZAHEER, R., et al. "Future Internet: The Internet of Things Architecture, Possible Applications and Key Challenges". In: *2012 10th International Conference on Frontiers of Information Technology*, pp. 257–260, 2012. doi: 10.1109/FIT.2012.53.

[99] WU, M., LU, T.-J., LING, F.-Y., et al. "Research on the architecture of Internet of Things". In: *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE)*, v. 5, pp. V5–484–V5–487, 2010. doi: 10.1109/ICACTE.2010.5579493.

[100] RAY, P. "A survey on Internet of Things architectures", *Journal of King Saud University - Computer and Information Sciences*, v. 30, n. 3, pp. 291 – 319, 2018. ISSN: 1319-1578. doi: https://doi.org/10.1016/j.jksuci.2016.10.003. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1319157816300799>.

[101] ANDREWS, J. G., BUZZI, S., CHOI, W., et al. "What Will 5G Be?" *IEEE Journal on Selected Areas in Communications*, v. 32, n. 6, pp. 1065–1082, 2014. ISSN: 0733-8716. doi: 10.1109/JSAC.2014.2328098.

[102] BOCCARDI, F., HEATH, R. W., LOZANO, A., et al. "Five disruptive technology directions for 5G", *IEEE Communications Magazine*, v. 52, n. 2, pp. 74–80, 2014. ISSN: 0163-6804. doi: 10.1109/MCOM.2014.6736746.

[103] DELICATO, F. C., PIRES, P. F., BATISTA, T., et al. "Towards an IoT Ecosystem". In: *Proceedings of the First International Workshop on Software Engineering for Systems-of-Systems*, SESoS '13, p. 25–28, New York, NY, USA, 2013. Association for Computing Machinery. ISBN: 9781450320481. doi: 10.1145/2489850.2489855.

[104] ALABA, F. A., OTHMAN, M., HASHEM, I. A. T., et al. "Internet of Things security: A survey", *Journal of Network and Computer Applications*, v. 88, pp. 10 – 28, 2017. ISSN: 1084-8045. doi: https://doi.org/10.1016/j.jnca.2017.04.002. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1084804517301455>.

[105] ALI, I., AHMED, A. I. A., ALMOGREN, A., et al. "Systematic literature review on IoT-based botnet attack", *IEEE Access*, 2020.

[106] KHAN, Z. A., HERRMANN, P. "A Trust Based Distributed Intrusion Detection Mechanism for Internet of Things". In: *2017 IEEE 31st International Conference on Advanced Information Networking and Applications (AINA)*, pp. 1169–1176, 2017. doi: 10.1109/AINA.2017.161.

[107] RAD, M. M., RAHMANI, A. M., SAHAFI, A., et al. "Social Internet of Things: vision, challenges, and trends", *Human-centric Computing and Information Sciences*, v. 10, n. 1, pp. 1–40, 2020.

[108] P., K., SRIDHAR, R. "Social Internet of Things (SIoT): Techniques, Applications and Challenges". In: *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)(48184)*, pp. 445–450, 2020. doi: 10.1109/ICOEI48184.2020.9142908.

[109] KHAN, W. Z., AALSALEM, M. Y., KHAN, M. K., et al. "When social objects collaborate: Concepts, processing elements, attacks and challenges", *Computers & Electrical Engineering*, v. 58, pp. 397–411, 2017. ISSN: 0045-7906. doi: https://doi.org/10.1016/j.compeleceng.2016.11.014. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0045790616307728>.

[110] KHANFOR, A., HAMROUNI, A., GHAZZAI, H., et al. "A trustworthy recruitment process for spatial mobile crowdsourcing in large-scale social iot". In: *2020 IEEE Technology & Engineering Management Conference (TEMSCON)*, pp. 1–6. IEEE, 2020.

[111] ATZORI, L., IERA, A., MORABITO, G., et al. "The Social Internet of Things (SIoT) – When social networks meet the Internet of Things: Concept, architecture and network characterization", *Computer Networks*, v. 56, n. 16, pp. 3594–3608, 2012. ISSN: 1389-1286. doi: https://doi.org/10.1016/j.comnet.2012.07.010. Disponível em: <https://www.sciencedirect.com/science/article/pii/S1389128612002654>.

[112] ALAM, K. M., SAINI, M., SADDIK, A. E. "Toward Social Internet of Vehicles: Concept, Architecture, and Applications", *IEEE Access*, v. 3, pp. 343–357, 2015. doi: 10.1109/ACCESS.2015.2416657.

[113] LIN, Z., DONG, L. "Clarifying trust in social internet of things", *IEEE Transactions on Knowledge and Data Engineering*, v. 30, n. 2, pp. 234–248, 2017.

[114] AHMED, A. I. A., AB HAMID, S. H., GANI, A., et al. "Trust and reputation for Internet of Things: Fundamentals, taxonomy, and open research challenges", *Journal of Network and Computer Applications*, v. 145, pp. 102409, 2019.

[115] PRATHAPCHANDRAN, K., RUTRAVIGNESHWARAN, P. "Trust Based Security Mechanisms for Resource-Constrained Internet of Things-A Review", *Journal of Physics: Conference Series*, v. 1850, n. 1, pp. 012042, may 2021. doi: 10.1088/1742-6596/1850/1/012042. Disponível em: <https://doi.org/10.1088/1742-6596/1850/1/012042>.

[116] DIN, I. U., GUIZANI, M., KIM, B.-S., et al. "Trust management techniques for the Internet of Things: A survey", *IEEE Access*, v. 7, pp. 29763–29787, 2018.

[117] LAMPORT, L., SHOSTAK, R., PEASE, M. "The Byzantine generals problem", *ACM Transactions on Programming Languages and Systems (TOPLAS)*, v. 4, n. 3, pp. 382–401, 1982.

[118] SUNYAEV, A. "Distributed Ledger Technology". In: *Internet Computing: Principles of Distributed Systems and Emerging Internet-Based Technologies*, pp. 265–299, Cham, Springer International Publishing, 2020. ISBN: 978-3-030-34957-8. doi: 10.1007/978-3-030-34957-8_9. Disponível em: <https://doi.org/10.1007/978-3-030-34957-8_9>.

[119] WILD, J., ARNOLD, M., STAFFORD, P. "Technology: Banks seek the key to blockchain". 2015. Disponível em: https://www.ft.com/content/eb1f8256-7b4b-11e5-a1fe-567b37f80b64. Acessado em Dezembro de 2021.

[120] BRADBURY, D. "The problem with Bitcoin", *Computer Fraud & Security*, v. 2013, n. 11, pp. 5 – 8, 2013. ISSN: 1361-3723. doi: https://doi.org/10.1016/S1361-3723(13)70101-5. Disponível em: <http://www.sciencedirect.com/science/article/pii/S1361372313701015>.

[121] GUARDIAN, T. "NSA Prism program taps in to user data of Apple, Google and others". 2013. https://www.theguardian.com/world/2013/jun/06/us-tech-giants-nsa-data.

Apêndice A

Noções de Criptografia Baseada em Identidades

A criptografia é o estudo de técnicas matemáticas que proporcionam uma comunicação segura, enviando mensagens de forma indecifrável por terceiros. Essas técnicas matemáticas estão relacionadas a aspectos de segurança da informação, como confidencialidade, integridade de dados, autenticação e a inegabilidade da autoria de dados.

Segundo Schneier [82], a encriptação é o propósito original da criptografia, que consiste no processo de transformar uma mensagem inteligível em uma mensagem cifrada ininteligível. Ao mesmo tempo, o processo oposto é a decifração, que consiste em recuperar a mensagem inteligível da mensagem cifrada correspondente. A criptografia depende de algoritmos criptográficos e chaves criptográficas para os processos de criptografia e decifração. As chaves criptográficas são definidas como um grande número que dificilmente pode ser descoberto, exigindo sucessivas tentativas de localizá-las.

Enquanto isso, os algoritmos criptográficos baseados em chave podem ser classificados em simétricos ou assimétricos. Algoritmos simétricos usam a mesma chave criptográfica, tanto para o processo de criptografia quanto para o processo de decifração. Esses algoritmos podem ser aplicados a um fluxo contínuo de dados, operando em um *bit* (ou *byte*) de cada vez; ou em blocos, operando em grupos de bits (ou bytes) [83]. Os algoritmos assimétricos, por sua vez, usam uma chave diferente para cada processo criptográfico. A chave usada para criptografar é a chamada *chave pública*, com a qual qualquer remetente pode criptografar uma mensagem, mas apenas o destinatário, que possui a chave de decifração (*chave privada*), pode abri-la. Também é possível que um remetente criptografe uma mensagem com sua chave privada para fins de assinatura digital, enquanto todos com a chave pública correspondente podem ter certeza da origem da mensagem.

Para que uma mensagem seja recuperada por um terceiro (possível invasor), se-

ria necessário que este soubesse qual algoritmo e chave criptográfica foram usados. Porém, o princípio de Kerckhoff [84] diz que a segurança de um sistema criptográfico deve depender apenas do sigilo das chaves, e não do sigilo do algoritmo. Assim, para que o processo de criptografia seja bem-sucedido, é necessário que as chaves criptográficas sejam protegidas e geradas de forma robusta, ou seja, não sejam facilmente descobertas. Para permitir que as chaves geradas sejam robustas o suficiente para que não possam ser facilmente descobertas, os sistemas de chaves públicas são baseados em problemas matemáticos de difícil solução computacional, uma vez que algoritmos eficientes para suas resoluções de tempo polinomial não foram descobertos.

Dois tipos de problemas encontrados na literatura são comumente usados em criptografia, a saber, o problema de fatoração de inteiros e o problema de logaritmo discreto. Este último problema ainda apresenta uma versão baseada em curvas elípticas, uma das áreas abordadas nesta tese. Nessa versão, o problema é saber quantas vezes uma determinada operação é realizada para obter um resultado. A Seção A.2 cobrirá os aspectos das curvas elípticas com mais detalhes.

O tamanho da chave criptográfica irá variar dependendo do problema matemático escolhido para suportar o sistema criptográfico. Dado um nível de segurança a ser alcançado, algoritmos baseados no problema de fatoração de grande número requerem uma chave maior (mais *bits*) do que a chave para algoritmos baseados no problema de logaritmo discreto. Assim, considerando a aplicação em IoT e as restrições comumente encontradas em dispositivos IoT, uma abordagem mais adequada é a utilização de chaves menores, a fim de consumir o mínimo possível os recursos do dispositivo. Por isso, este trabalho utiliza técnicas criptográficas baseadas no problema de logaritmos discretos com curvas elípticas.

A.1 Criptografia Baseada em Identidade

A Criptografia Baseada em Identidade (*Identity-Based Encryption – IBE*) foi proposta por Shamir [85], sendo inicialmente implementada através do trabalho de Boneh e Franklin [86], que usaram curvas elípticas para criar sistemas criptográficos baseados em identidade. O IBE está documentado na RFC 5091 [87], que serviu de base para a descrição do funcionamento do IBE ao longo do texto.

IBE é um tipo de criptografia de chave pública, com a diferença de que as chaves são calculadas e não geradas aleatoriamente como em outros esquemas de chaves públicas (como PKI, por exemplo). No IBE, a chave pública é calculada a partir de uma identidade e a chave privada é calculada com base na chave pública. Como as chaves públicas são geradas a partir de informações públicas (por exemplo, endereços de dispositivos) e as chaves privadas são geradas com base em chaves públicas,

então qualquer dispositivo seria capaz de gerar a chave privada de qualquer outro dispositivo. Para não permitir que isso aconteça, o IBE considera a presença de um gerador central de chaves (Private Key Generator - PKG), responsável por gerar as chaves privadas e distribuí-las aos dispositivos com suas respectivas chaves públicas. Comparado com a criptografia de chave pública típica, este esquema reduz muito a complexidade do processo de criptografia para usuários e administradores. Apesar de dispensar a necessidade de uma autoridade de certificação, o IBE depende de um servidor de terceiros para gerar chaves privadas. Tal vulnerabilidade permite ataques *Sybil*, nos quais um nó IoT malicioso finge ser o servidor PKG [88]. A Figura A.1 dá uma visão geral do processo IBE.

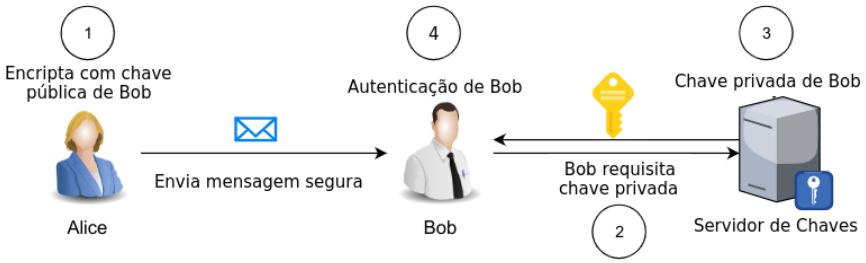


Figura A.1: (1) - Alice envia uma mensagem cifrada para Bob. (2) - Bob pergunta ao servidor PKG. (3) - O servidor PKG envia uma chave privada para Bob para permitir que ele descriptografe a mensagem de Alice

O esquema de criptografia baseado em identidade permite o estabelecimento de comunicação segura entre quaisquer duas partes (dispositivos, por exemplo) e a verificação de identidade sem a necessidade de troca de chave pública. O esquema é análogo ao envio de uma carta a uma pessoa, em que qualquer pessoa pode enviar uma carta a outra pessoa apenas sabendo o nome do destinatário e o respectivo endereço. Desta forma, este esquema simplifica o processo de encriptação, reduzindo o número de mensagens para troca de chaves criptográficas, o que tem um impacto positivo no contexto da IoT, visto que serão utilizados menos recursos de rede, bem como menor consumo de bateria dos dispositivos.

Para tornar o IBE possível, Boneh e Franklin [86] propuseram o uso de mapas bilineares, descritos a seguir.

Sejam G_1 e G_2 dois grupos de ordem p para algum grande número primo p. O mapa bilinear $e : G_1 \times G_1 \rightarrow G_2$ deve satisfazer as seguintes propriedades:

- **Bilinearidade**: O mapeamento $e : G_1 \times G_1 \rightarrow G_2$ é bilinear se:

$$e(aP, bQ) = e(P, Q)^{ab} \quad \forall P, Q \in G_1 \quad e \quad \forall a, b \in \mathbb{Z};$$

- **Não degeneração**: não há mapeamento para todos os pares de $G_1 \times G_1$ para o elemento de identidade de G_2;

- **Viabilidade computacional**: um algoritmo eficiente é necessário para calcular o mapeamento e para qualquer $P, Q \in G_1$.

Quando as propriedades acima são satisfeitas, diz-se que existe um mapeamento "permitido". Como Boneh e Franklin [86] mostraram, os pontos que definem o *locus* de uma curva elíptica podem ser aplicados como grupos "permitidos".

Em relação ao funcionamento do IBE, as etapas para o envio de uma mensagem cifrada, conforme RFC 5091 [87], são:

1. Obtenção dos parâmetros públicos do destinatário pelo remetente, por exemplo, a identificação utilizada para gerar a respectiva chave pública do destino;

2. Construir e enviar a mensagem cifrada usando os parâmetros públicos obtidos;

3. Obtenção dos parâmetros públicos do destinatário por si só, bem como sua respectiva chave privada do PKG;

4. Decifração da mensagem recebida.

A Figura A.2 ilustra o processo de troca de mensagens no esquema IBE. Inicialmente, o remetente obtém os parâmetros públicos do destinatário de um servidor de parâmetros, que pode ser acessado através de um endereço previamente conhecido. Se os parâmetros públicos consistirem em informações intrínsecas à comunicação entre o remetente e o destinatário (um e-mail ou endereço IP, por exemplo), essa consulta ao servidor de parâmetros não é necessária.

Caso seja necessária tal consulta, a comunicação com o servidor de parâmetros deve ser feita de forma segura, utilizando um protocolo de autenticação como o *Transport Layer Security* (TLS) , por exemplo. A construção da mensagem cifrada é então feita pelo remetente usando a chave pública do destinatário, calculada com base nos parâmetros públicos obtidos.

Do ponto de vista do destinatário, para receber uma mensagem cifrada no esquema IBE, o próprio destinatário também deve obter os parâmetros públicos relacionados a si mesmo para o cálculo da mesma chave pública que o remetente computou. Com base nessa chave, o destinatário pode obter a respectiva chave privada do PKG, que calcula as chaves privadas de todos os usuários no esquema IBE. Essa comunicação com o PKG também deve ser feita por meio de um protocolo de autenticação. De posse da chave privada, o destinatário pode então decifrar a mensagem recebida.

Entre as vantagens do uso do IBE, o fato de não haver necessidade de armazenamento de chaves públicas é positivo, uma vez que podem ser calculadas pelos próprios membros da rede a partir das informações sobre a identidade de outros membros com os quais desejam se comunicar. Inclui também a comunicação de

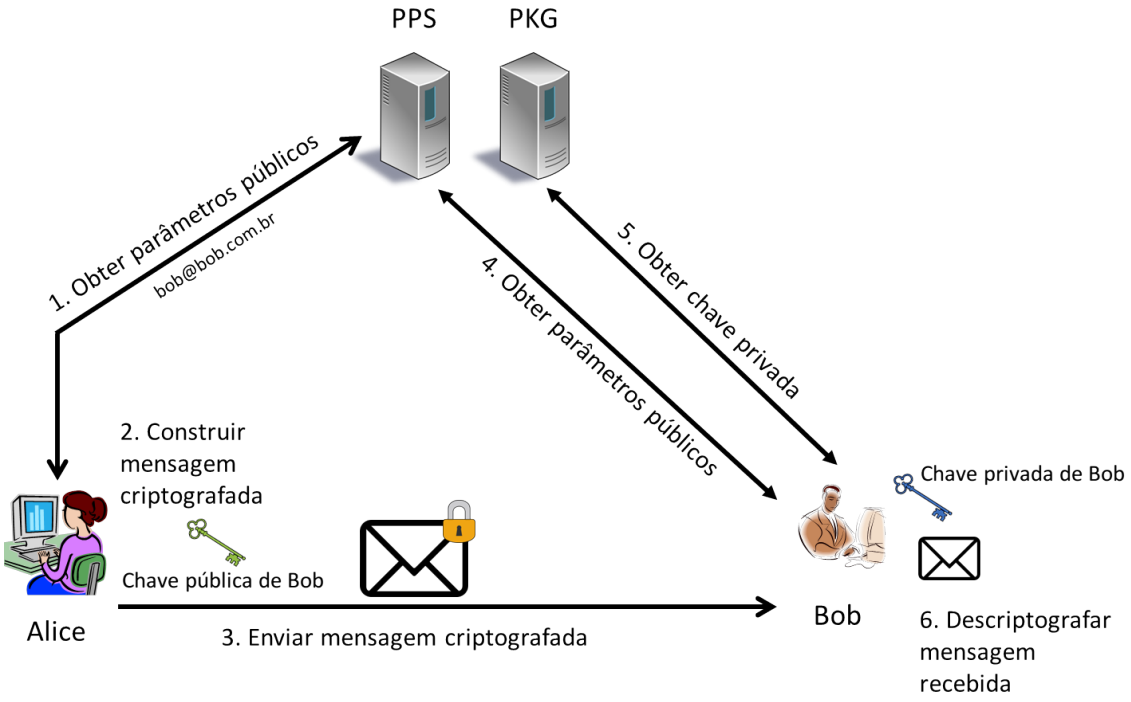


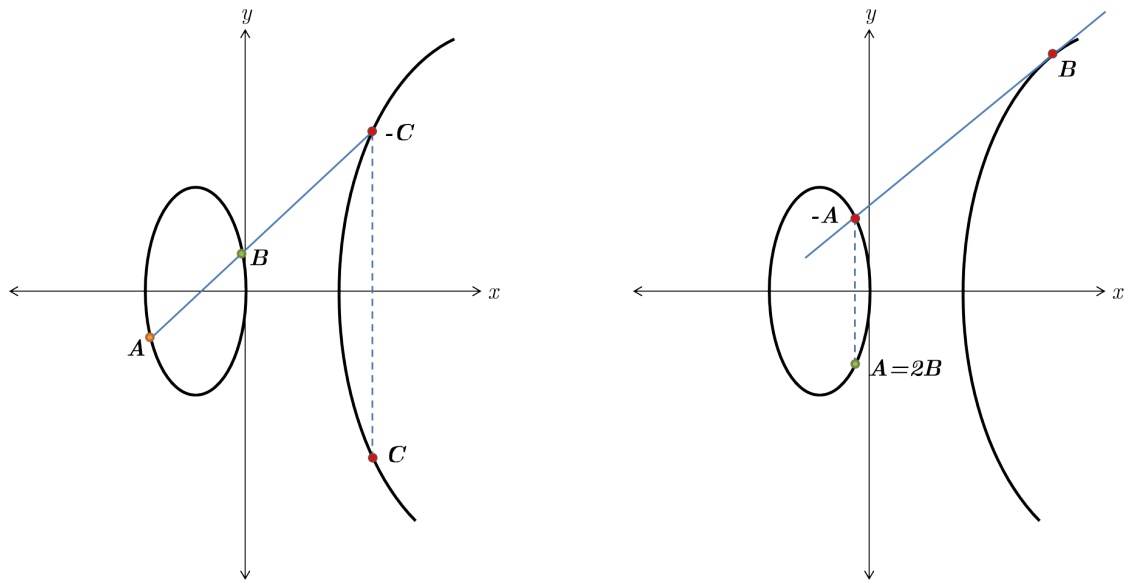
Figura A.2: Processo de troca de mensagens no esquema IBE

membros que ainda não iniciaram a sua participação na rede, que não requer a troca de informações prévias (chaves criptográficas, por exemplo), o que reduz o *overhead* de comunicação. Além disso, existe a possibilidade de recuperação de chaves privadas, uma vez que são geradas pelo PKG a qualquer momento. Porém, o que funciona como uma vantagem no IBE também acaba sendo um ponto fraco, em relação ao gerador de chave privada. Como o PKG pode gerar chaves privadas, ele também pode divulgá-las ou usá-las, fazendo-se passar por usuários legítimos. Outra desvantagem é que o PKG pode acessar o conteúdo de todas as mensagens de todos os usuários do sistema, o que compromete a questão básica da privacidade.

A.2 Curvas Elípticas

O problema do logaritmo discreto em curvas elípticas tem sido amplamente utilizado para técnicas criptográficas, devido à sua capacidade de fornecer certo nível de segurança com o uso de uma chave relativamente menor do que em outros tipos de problema.

Por meio de curvas elípticas é possível oferecer criptografia com poucas operações aritméticas, o que leva a um desempenho com baixo custo computacional em comparação com outras abordagens como RSA [89, 90]. Além disso, também oferece um sistema criptográfico que envolve um maior custo computacional para a inversão de chaves privadas.



(a) Exemplo de "soma" em curvas elípticas

(b) Exemplo de multiplicação por escalar em curvas elípticas

Figura A.3: Operações em curvas elípticas

As curvas elípticas não são elipses, mas herdam esse nome de uma conexão com o problema do comprimento do arco de uma elipse. As curvas elípticas são geralmente definidas no plano $\mathbb{R}^2$, [91]:

$$y^2 + cxy + dy = x^3 + ax + b, \qquad a, b, c, d \in \mathbb{R}^2 \tag{A.1}$$

Com base nesta equação, obtém-se o *locus* formado pelos pontos que satisfazem esta equação, formando uma curva cúbica em x simétrica em relação ao eixo x. O conjunto desses pontos forma um grupo [69], que apresenta características e operações fundamentais para criptografia baseada em curvas elípticas.

A.2.1 Operações em Curvas Elípticas

Seja uma curva elíptica representada por $y_2 = x_3 + ax + b$ e os pontos $A = (x_1, y_1)$, $B = (x_2, y_2)$ e $C = (x_3, y_3)$, a operação de "soma" é definida como [92]:

$$\begin{aligned} \lambda &= \frac{y_2 - y_1}{x_2 - x_1}; \\ x_3 &= \lambda^2 - x_1 - x_2; \\ y_3 &= \lambda(x_1 - x_3) - y_1 \end{aligned} \tag{A.2}$$

Desta forma, a operação $A + B$ resulta no ponto C. A Figura A.3a ilustra a operação de "soma".

A operação de multiplicação escalar é definida como uma sucessão de operações de "soma" e pode ser representada da seguinte forma:

$$\begin{gathered} A = B + B + ... + B \quad \text{(n vezes)} \\ A = nB \end{gathered} \tag{A.3}$$

Para viabilizar o uso de curvas elípticas para cifragem, é essencial a existência de operações originárias de estruturas algébricas como grupos, corpos, anéis, etc. Neste contexto, a operação de adição em uma curva elíptica ocorre através da existência de um grupo abeliano implicitamente associado, que vem da limitação da Equação A.1 com a operação de módulo, conforme abaixo:

$$(y^2 + cxy + dy) \mod p = (x^3 + ax + b) \mod p, \qquad a, b, c, d \in \mathbb{R}^2 \tag{A.4}$$

A dificuldade em resolver o problema do logaritmo discreto neste caso reside em encontrar o número de operações de "soma", ou seja, n, dados A e B dentro do grupo abeliano definido.

Apêndice B

Noções de Teoria da Informação

Na Teoria da Informação, Claude Shannon [3] conceituou de modo distinto uma representação probabilística que quantifica e define limites para transmitir mensagens de forma otimizada em um canal de comunicação com ruído. Os limites estabelecidos pela teoria de Shannon contemplam a compressão máxima da informação (entropia) e a capacidade máxima de transmissão de um canal de comunicação [70].

A teoria simples e brilhante desenvolvida por Shannon teve um impacto indubitável no projeto de sistemas de comunicação proporcionando um arcabouço matemático para formular e quantificar interações além das leis físicas [70].

B.1 Entropia de Shannon

Shannon procurou uma medida quantitativa de quanto a ocorrência de uma saída $y = b_j$ diz sobre a ocorrência de uma entrada $x = a_k$, considerando um conjunto de mensagens de entrada $X = a_1, ..., a_k$ e as mensagens de saída $Y = b_1, ..., b_j$ através de um canal de comunicação ruidoso. Em termos probabilísticos, a ocorrência de $x = a_k$ é dada pela probabilidade *a priori* $P[x]$ [69].

A medida de informação investigada por Shannon está relacionada a quanto saber sobre uma saída $y = b_j$ altera a probabilidade de ocorrência da entrada $x = a_k$, ou seja, qual é a probabilidade *a posteriori* $P[x|y]$ [69]. Portanto, esta medida de informação, denotada por $I(x, y)$, pode ser calculada como o logaritmo da razão entre as probabilidades *a posteriori* e *a priori*, como na Equação B.1.

$$I(x, y) = \log \frac{P[x|y]}{P[x]} \tag{B.1}$$

Devido à simetria que existe entre $I(x, y)$ e $I(y, x)$, a Equação B.1 define o conceito de informação mútua. Observe que, quando $P[x|y] = p[x]$, temos $I(x, y) = 0$, o que indica que não há diminuição na incerteza, ou em outras palavras, não há ganho de informação [70]. Um caso especial de informação mútua é quando a

ocorrência de $y = b_j$ especifica exclusivamente $x = a_k$, fazendo $P[x|y] = 1$ [69]. Assim, a auto-informação é definida como na Equação B.2.

$$I(x) = \log \frac{1}{P[x]} = -\log P[x] \tag{B.2}$$

Com isso, o conceito de entropia pode ser interpretado como o valor esperado da auto-informação, dado na Equação B.3.

$$H(x) = \sum P[x] \log \frac{1}{P[x]} = -\sum P[x] \log P[x] \tag{B.3}$$

Para uma variável aleatória B que segue a distribuição de Bernoulli, por exemplo, onde p é a probabilidade de ocorrência e, consequentemente, $1-p$ a probabilidade de não ocorrência, o cálculo da entropia $H(B)$ é dado pela Equação B.4. A Figura B.1 ilustra a variação da entropia em relação à probabilidade de ocorrência de um evento binário.

$$H(B) = -p \log p - (1-p) \log (1-p) \tag{B.4}$$

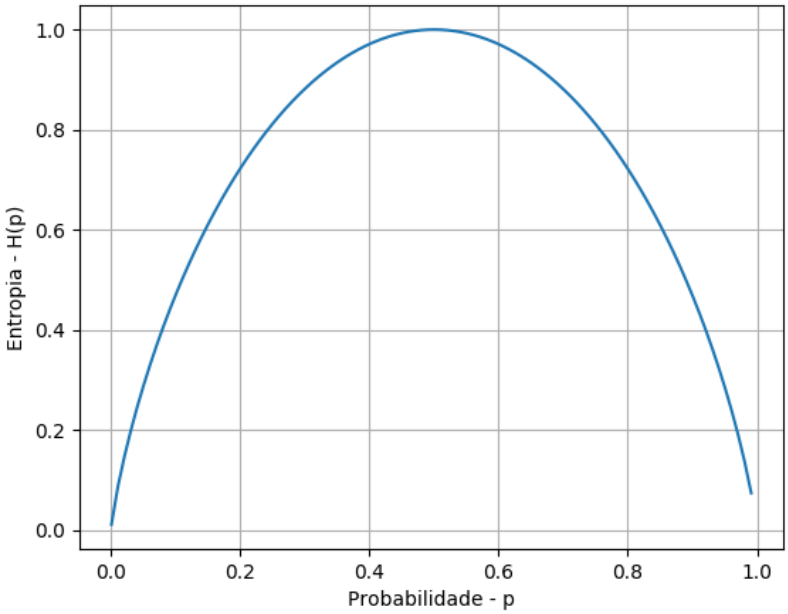


Figura B.1: Variação de entropia para uma variável aleatória de Bernoulli

Para se ter uma noção do conceito de entropia, considere uma analogia com os alunos de uma aula de Teoria da Informação, na qual todo o conteúdo ensinado é novo na primeira aula. Há muitas informações novas e muitas incertezas (entropia) sobre o que será ensinado. No entanto, o professor decide repetir exatamente a mesma lição indefinidamente. O que acontece é que o nível de incerteza diminui à

medida que os alunos vão às aulas, pois aprendem cada vez mais sobre o conteúdo. Assim, quanto maior for a quantidade de informação a ser obtida, maior será a entropia. Em outras palavras, à medida que os alunos têm menos conteúdo para aprender, a incerteza sobre a aula diminui (a entropia diminui). Por outro lado, quanto menor a entropia (menos informação), mais certeza se tem sobre o conteúdo lecionado. De maneira geral, ao ter a probabilidade de ocorrência de um evento muito alta ou muito baixa, tem-se uma diminuição da incerteza (como ilustrado na Figura B.1).

Considerando tais conceitos, é possível propor seu uso para inferir, por exemplo, sobre o comportamento do tráfego de um dispositivo para classificá-lo como confiável ou não. No Capítulo 3, vemos mais detalhes sobre como a Teoria da Informação pode ajudar na atribuição de métricas de confiança para IoT.

B.2 Entropias de Tsallis e de Rényi

Além do conceito de entropia de Shannon largamente difundido, outros dois tipos de entropia fazem parte do conjunto de ferramentas de Teoria da Informação, a saber a entropia de Tsallis e a entropia de Rényi.

A entropia de Tsallis [93, 94] surgiu como proposta de generalização da entropia de Boltzmann–Gibbs–Shannon para sistemas físicos não-extensíveis. Através da inserção de um novo parâmetro α, a noção de entropia é generalizada para prover mais informação sobre eventos específicos, como eventos muito comuns ou eventos raros. Assim, tal parâmetro provê controle sobre a importância dada a tais eventos, de modo que potências das probabilidades de ocorrência dos eventos $\sum_{i=1}^{n} P[x_i]^{\alpha}$ passam a ser utilizadas. Quanto maior o valor de α, mais sensível a entropia fica para eventos que ocorrem com certa frequência, enquanto para valores negativos de α, mais importância é dada a eventos raros.

A entropia de Tsallis é dada pela seguinte expressão na Equação B.5:

$$T_{\alpha} = \frac{1}{\alpha - 1}\left(1 - \sum_{i=1}^{n} P[x_i]^{\alpha}\right) \tag{B.5}$$

Similarmente, a entropia de Rényi também considera o parâmetro α em sua formulação, dada pela seguinte expressão na Equação B.6:

$$R_{\alpha} = \frac{1}{1 - \alpha}\log\left(\sum_{i=1}^{n} P[x_i]^{\alpha}\right) \tag{B.6}$$

Diversos trabalhos foram desenvolvidos usando as entropias de Tsallis e Rényi como fundamento, desde aplicações físicas [94, 95], arvores de decisão [96], a segmentação de imagens [97].

Para ilustrar os conceitos apresentados, os gráficos da Figura B.2 mostram os valores para as entropias de Tsallis e Rényi para diferentes valores de α (positivos e negativos), considerando duas probabilidades p_1 e $p_2 = 1 - p_1$.

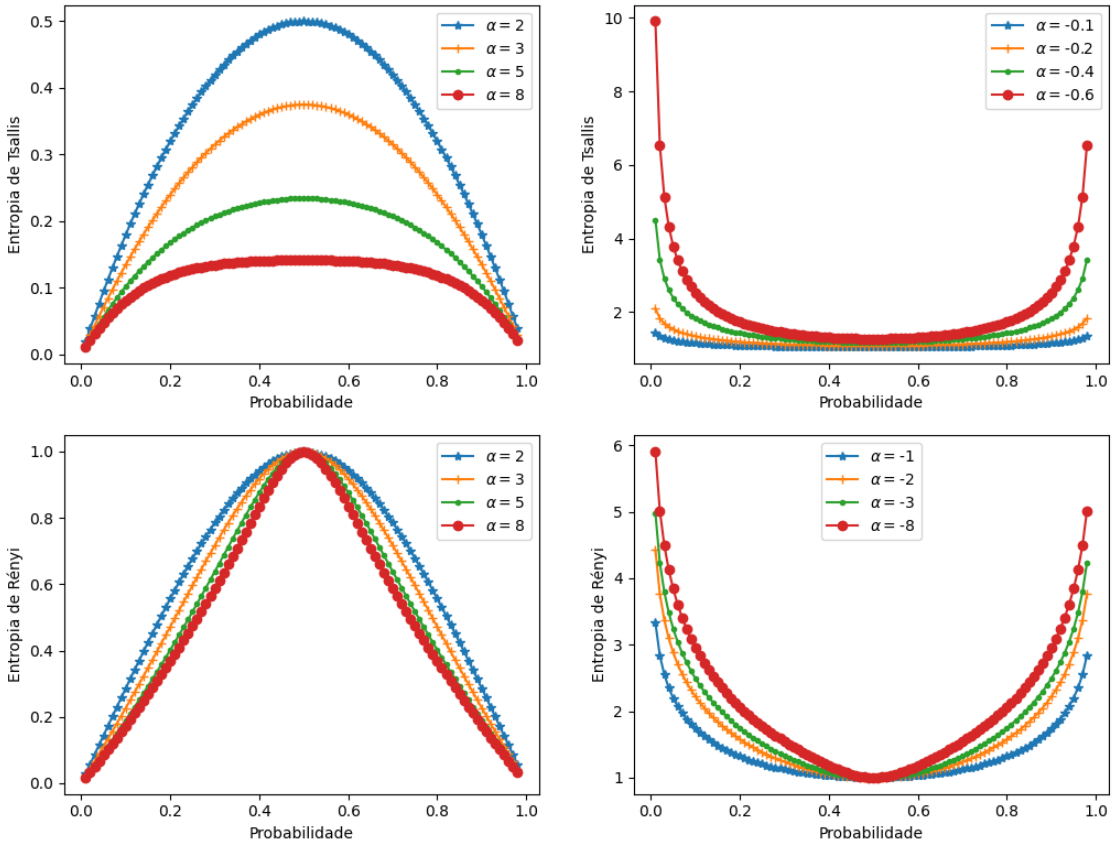


Figura B.2: Variação das entropias de Tsallis e Rényi para diferentes valores de α considerando duas probabilidades

Apêndice C

Noções de Internet das Coisas

C.1 Internet das Coisas

A Internet das Coisas (*Internet of Things – IoT*) [4] é um paradigma que promove a evolução das redes de computadores habilitando a comunicação para quaisquer categorias de objetos, tornando-os dispositivos inteligentes. Para estruturar a IoT, diversas arquiteturas são propostas na literatura [98–100], variando em taxonomias lógicas ou físicas e em número de camadas. Uma arquitetura comumente usada [98] considera cinco camadas fundamentais, sendo elas:

- **Camada de Percepção:** é a camada física onde os dispositivos IoT se encontram. Os dados são obtidos ("produzidos") a partir desta camada, na qual se encontram não apenas os dispositivos sensores que geram massa de dados, mas também os atuadores, que de alguma forma modificam o estado do ambiente em que se encontram;

- **Camada de Rede:** é a camada que habilita a conectividade e provê infraestrutura para os dispositivos IoT se comunicarem com suas respectivas aplicações. Esta camada contém as mais diversas tecnologias de conectividade, incluindo redes cabeadas, redes sem fio, redes móveis 3G, 4G e 5G, entre outras. Em particular, as redes 5G surgem como a infraestrutura de conectividade que será uma realidade para o futuro das redes, apresentando diversos estudos sobre inovações nesta área [101, 102];

- **Camada de *Middleware*:** é a camada de *software* que permite a integração entre componentes IoT heterogêneos, que possuem suas especificidades e não podem se comunicar exceto por meio dessa camada. Assim, a camada de *middleware* é essencial para o requisito de interoperabilidade na IoT, que possibilita a comunicação transparente entre sistemas e recursos por meio da criação de uma interface comum entre os dispositivos.

- **Camada de Aplicação:** reúne os mais diferentes domínios de aplicações de IoT, entre os quais podemos citar cidades inteligentes, IoT industrial, agricultura de precisão, saúde, casas inteligentes, redes veiculares, entre outros. Nesta camada são desenvolvidas as aplicações que permitem, que coleta de dados de dispositivos IoT seja realizada em quantidades sem precedentes;

- **Camada de Negócio:** no nível mais alto, é a camada onde se encontra o valor agregado das aplicações IoT, tanto para monetização, quanto para processos de tomada de decisão. Esta camada se beneficia dos dados gerados pelos dispositivos, além de despachar comandos de atuação que influenciam os ambientes nos quais os dispositivos relacionados se encontram.

Além das camadas fundamentais, outra camada que é essencial, mas não é muito abordada é a **camada de segurança**. Em particular, esta camada se diferencia pelo fato de permear todas as outras camadas, visto que os aspectos de segurança devem ser considerados em todos os níveis da arquitetura, não apenas em camadas específicas, como, por exemplo, se os protocolos de rede sem fio seguros são usados na camada de rede, ou se os controles de acesso são implementados na camada de aplicação. Todas as camadas devem incluir soluções de segurança para que todo o sistema esteja seguro, desde o projeto e desenvolvimento das aplicações, até seus usos e implementações reais. No entanto, ainda existem lacunas significativas na IoT em relação à segurança. Uma dessas lacunas é o aspecto da confiança entre os dispositivos, que este trabalho aborda. Assim, temos o que podemos chamar arquitetura de $5 + 1$ camadas, destacando a camada de segurança, perpendicular às outras. A Figura C.1 ilustra a arquitetura de $5 + 1$ camadas.

Outra forma de representar a arquitetura IoT é considerar o aspecto da localização física onde as soluções de segurança são implementadas [15, 42]. Conforme mostrado na Figura C.2, três camadas são identificadas, nas quais, dependendo da solução de segurança, a implementação pode ocorrer em uma ou mais camadas. Soluções que envolvem proteção de dados, por exemplo, requerem maior proximidade do local onde os dados são gerados e todo o caminho que podem percorrer, estabelecendo um compromisso entre a coleta e a disseminação dos dados. Assim, soluções desse tipo são implementadas desde a camada de dispositivos, até a camada de computação em nuvem, passando pela camada de borda. Enquanto isso, as soluções de controle de acesso não precisam necessariamente ser implementadas em dispositivos, mas em locais onde há mais recursos computacionais e de rede, como camadas de borda e de nuvem. Também não devemos descartar os requisitos de rede que as aplicações possam ter, não sendo tolerantes a atrasos, por exemplo, caso em que as implementações devem ocorrer o mais próximo possível dos dispositivos. A recomendação geral é que, sempre que possível, as implementações sejam feitas

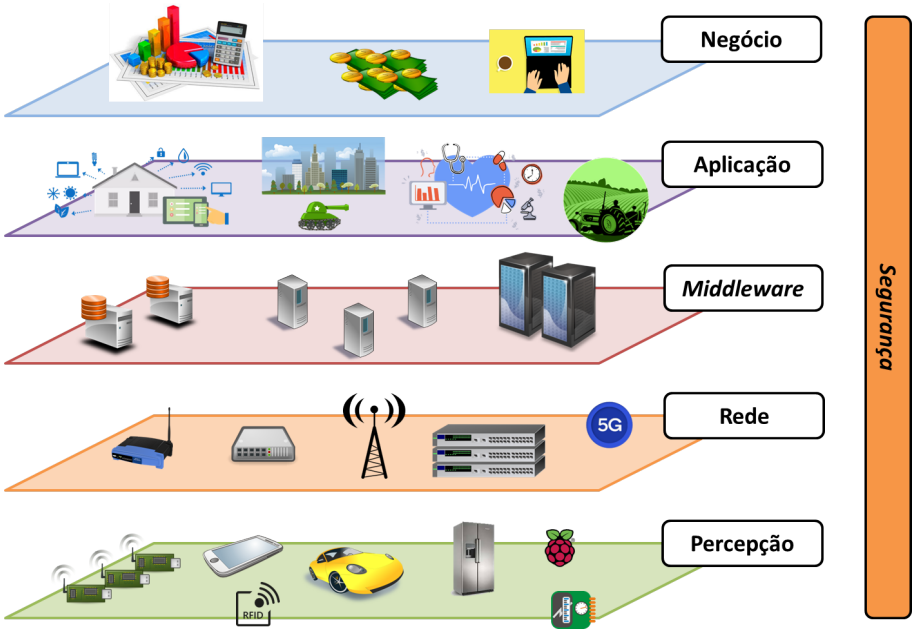


Figura C.1: Arquitetura IoT com 5 + 1 camadas

fora dos dispositivos para não comprometer seus recursos, como processamento e bateria, que comumente são limitados.

Diversos desafios estão presentes na realização do paradigma de IoT, como a descoberta de dispositivos, descoberta de redes para conectividade, identificação única de dispositivos, infraestrutura para comunicação e heterogeneidade entre plataformas, monitoramento dos dispositivos, gerenciamento de interoperabilidade, entre outros [103]. Além disso, os aspectos de privacidade e segurança também são objetivos importantes, assim como os tipos de ataques que os dispositivos podem sofrer

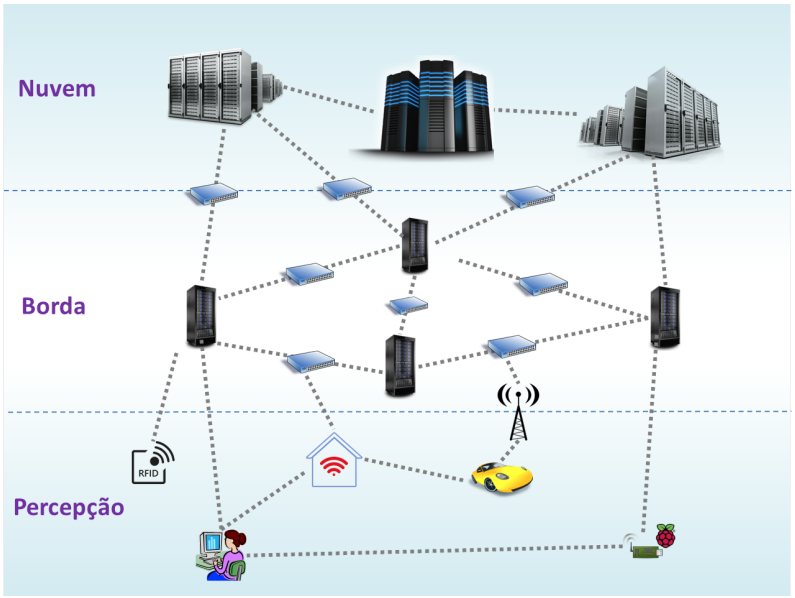


Figura C.2: Arquitetura IoT de três camadas

ou serem usados como vetores de ataques. Nesse contexto, vários trabalhos relacionados comentam os aspectos de segurança como essenciais e fundamentais para IoT [14, 16, 104]. Em [15], nós apresentamos uma revisão sistemática que reuniu diversos trabalhos demonstrando os esforços da comunidade acadêmica na busca por soluções para os problemas de segurança inerentes à IoT. Tal revisão confirma a importância das pesquisas nesta área. Em especial, a atribuição de confiança na comunicação entre dispositivos IoT ainda é um problema pouco explorado e carece de soluções.

Além das vantagens e benefícios que IoT traz, há também as desvantagens. Ao passo que IoT proporciona coleta de dados de maneira massiva com o crescimento vertiginoso do número de dispositivos conectados, pelo mesmo fato IoT também potencializa ataques cibernéticos, visto que aumenta significativamente o número de dispositivos vulneráveis conectados à Internet, expandindo a superfície de ataque. Ataques cibernéticos que afetam as aplicações IoT ou que usam os próprios dispositivos para realizar ataques estão crescendo e não mostram sinais de parar [105]. Os ataques cibernéticos em IoT podem ser classificados em: (i) ataques contra dispositivos IoT; e (ii) ataques que usam dispositivos IoT para realizar os ataques. Os ataques contra a IoT podem ser dos mais diversos tipos, como [14, 106]:

- Infraestrutura de comunicação: que inviabiliza a comunicação entre os dispositivos, tornando-os indisponíveis;
- *Sybil*: caracterizado por um ataque de personificação, quando um nó (dispositivo) tenta se passar por outro;
- *Sink role*: quando um determinado nó absorve todo o tráfego para si, não o encaminhando para os demais nós;
- Encaminhamento seletivo: quando os nodos enviam apenas mensagens de controle para inviabilizar um serviço;
- Auto-promoção: quando um dispositivo dissemina boas recomendações sobre ele mesmo para aumentar sua reputação;
- *Ballot stuffing (good mouth)*: quando um ou mais dispositivos maliciosos reportam bons comportamentos sobre outro(s) dispositivo(s) para aumentar a reputação destes;
- *Bad mouth*: o caso oposto do ataque *good mouth*, quando um ou mais dispositivos maliciosos se unem para reportar negativamente sobre outro(s) dispositivo(s) para reduzir a reputação destes.

Por outro lado, os atacantes também usam dispositivos IoT para realizar ataques, aproveitando suas vulnerabilidades para usá-los como *bots*, agrupando-os em *botnets* para atingir maior potencial de ataque. Os ataques não necessariamente apresentam novas características, mas são fortalecidos pelo grande número de dispositivos vulneráveis utilizados. Os ataques realizados com dispositivos IoT são:

- Acesso a ambientes: acesso a ambientes privados, como residências ou empresas, através de câmeras, microfones ou sensores conectados à Internet de forma insegura;
- Sistemas de controle: ataques com dispositivos que controlam semáforos, automóveis, sistemas de irrigação, ou qualquer ambiente inteligente, como uma cidade inteligente, por exemplo;
- Negação de Serviço (*Denial of Service – DoS*): ataque de negação de serviço realizado através de um dispositivo de geração de tráfego para tornar os serviços indisponíveis;
- Negação de Serviço Distribuído (*Distributed Denial of Service – DDoS*): ataque distribuído de negação de serviço executado através de vários dispositivos que geram tráfego para tornar os serviços indisponíveis.

Assim, propostas foram desenvolvidas para abordar as vulnerabilidades da IoT, abordadas com mais detalhes nos trabalhos relacionados na Seção 2. Além disso, esta tese apresenta uma proposta de atribuição de confiança na comunicação entre dispositivos IoT, aumentando a barreira de entrada de dispositivos não confiáveis, ou removendo-os da rede, caso já estejam conectados e passem a operar de maneira anômala.

A quantidade exorbitante de dados também contribui para o desafio de transferi-los de maneira contínua, confiável e segura. Além disso, como visto na Figura C.1, os dados gerados frequentemente passam por várias camadas, o que contribui para aumentar a complexidade do sistema. Como resultado, os ataques cibernéticos também podem ocorrer em camadas diferentes, identificando assim os ataques nos seguintes níveis:

- **Físico**: ataques que visam dispositivos, através da obtenção de controle não autorizado, devido ao acesso físico aos dispositivos (furto, por exemplo);
- **Lógico**: ataques possibilitados por vulnerabilidades de *software*, que permitem a instalação de *malwares* ou vírus;
- **Criptográfico**: ataques contra mecanismos de criptografia, geralmente por "força bruta" para quebra de senha ou descoberta de chave privada;

- **Rede**: ataques vindos da rede, que podem até usar os próprios dispositivos, como negação de serviço (DoS), ataques por meio de protocolos de roteamento e vulnerabilidades intrínsecas da mídia (rede sem fio, por exemplo).

A adaptação de aplicações IoT para se defender contra os diferentes tipos de ataques mencionados é uma tarefa homérica que requer trabalho colaborativo entre todas as partes envolvidas no processo. No entanto, se cada sistema IoT estiver inerentemente vinculado à rede, seria interessante se a própria rede pudesse se proteger e, como resultado, também proteger todos os sistemas conectados a ela. Dessa forma, a tarefa muda de defender-se contra ataques para evitar que ataques sejam viáveis de forma proativa. A abordagem proposta nesta dissertação visa concretizar essa estratégia, apresentando diretrizes para a sua realização.

C.2 *Social IoT*

Os aspectos sociais são inerentes à natureza humana, com relações sociais das mais diversas, complexas e dinâmicas possíveis. As relações entre as pessoas formam comunidades com base em interesses em comum, sendo tais comunidades fortalecidas à medida que interações e colaborações ocorrem entre os membros das comunidades. No contexto de IoT, os dispositivos conectados, em geral, apresentam também um objetivo em comum, dependendo da aplicação com a qual estão associados, e cooperam entre si através de interações a fim de alcançar tal objetivo. Quando o conceito de redes sociais é integrado à IoT, surge o termo *Social IoT* (SIoT) [28, 67, 107–109]. Considerado um novo paradigma de evolução da IoT e das redes sociais, a SIoT vem se consolidando na comunidade acadêmica, habilitando os dispositivos a construírem relacionamentos segundo interações e colaborações definidas em regras e políticas configuradas pelas aplicações associadas [110].

Assim como temos relações sociais de amizade, família, negócios, entre outros, para os humanos, podemos expandir tais relações também para o contexto de IoT. Dois dispositivos podem ser "amigos" entre si, enquanto é "primo" de outro dispositivo e "irmão" de um terceiro. Enquanto dispositivos maliciosos podem ser entendidos como "inimigos" de outros dispositivos. Atzori *et al.* [111] define os tipos de relações possíveis no contexto de SIoT, como as relações *parental object relationship* (POR) entre dispositivos de um mesmo período de fabricação; *co-location object relationship* (CLOR) entre dispositivos fisicamente próximos; *co-work object relationship* (CWOR) entre dispositivos que trabalham em conjunto para uma aplicação em comum; entre outros tipos. Além disso, outras relações foram expandidas por outros autores para outros domínios de aplicação, como, por exemplo, a *guardian object relationship* (GOR) destinada para as relações entre veículos e as infraestruturas

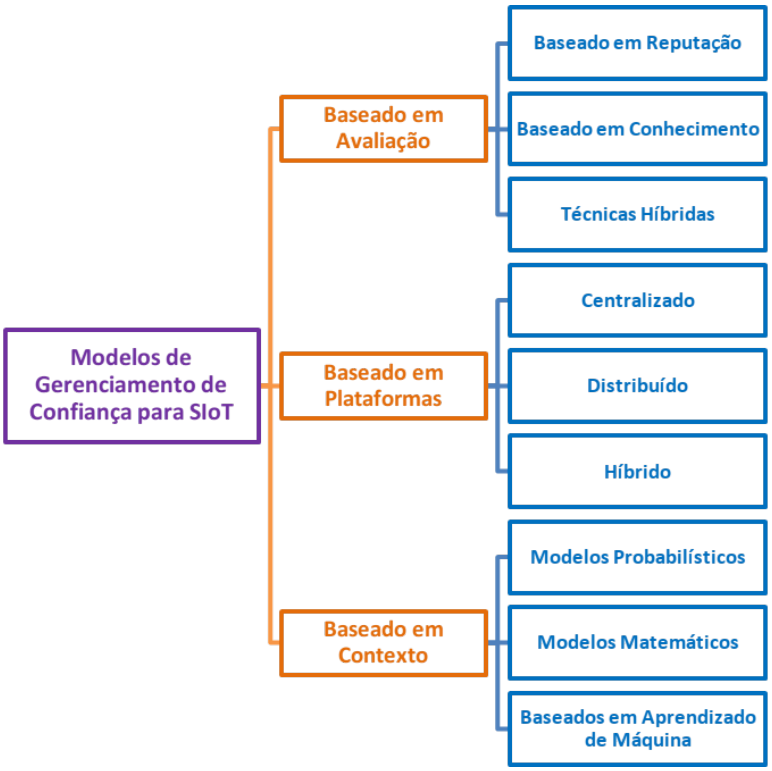


Figura C.3: Classificação dos modelos de gerenciamento de confiança para SIoT (Adaptado de [67]

de estradas em cenários de Internet de Veículos (IoV) [109, 112].

É preciso considerar também a existência de dispositivos mal-intencionados que queiram se valer das relações que SIoT proporciona para desempenhar ataques contra outros dispositivos. Assim, é necessário levar em consideração o aspecto essencial da confiança durante as interações entre os dispositivos IoT, assim como temos também a confiança como base dos relacionamentos entre as pessoas. Esse é exatamente o objeto de estudo desta tese, que se aplica inclusive neste contexto de SIoT. Uma métrica de confiança tem papel vital para a adoção e o correto funcionamento de SIoT.

Em [67] é apresentada uma classificação dos modelos de gerenciamento de confiança existentes na literatura, conforme ilustrado na Figura C.3. Considerando esta classificação, podemos perceber que nossa proposta de métrica de confiança se classifica como um modelo baseado em contexto, visto que propomos um modelo matemático para tal métrica.

C.2.1 Desafios em Social IoT

Social IoT apresenta desafios específicos, somados aos que a própria IoT já apresenta, como interoperabilidade e escalabilidade. Dentre os desafios temos [67, 109]:

- Escolha do amigo certo: pelo paradigma de SIoT, os dispositivos podem estabelecer relacionamentos de modo autonômico. Por conta da grande quantidade de dispositivos heterogêneos, saber com qual dispositivo se associar se torna um desafio. Novamente, o conceito de confiança se mostra indispensável para atacar este desafio;

- Escolha dos modelos de redes sociais: com a inclusão dos possíveis relações entre os dispositivos, escolher qual modelo de relacionamento [77] se ajusta a cada tipo de aplicação SIoT é um desafio relevante e essencial para o funcionamento adequado de SIoT;

- Aceitação dos usuários: sistemas baseados em SIoT ainda têm suas aceitações impactadas pelo alto risco de possíveis vazamentos de informações, não só pelos dispositivos de posse dos usuários de tais sistemas, mas também dos outros dispositivos com os quais são estabelecidos relacionamentos;

- Gerenciamento de confiança: o gerenciamento de confiança entre os dispositivos é fundamental em SIoT e torna-se um desafio quantificar de maneira objetiva tal conceito que é naturalmente subjetivo. Em particular, também é um desafio oferecer tal gerenciamento de modo distribuído segundo uma abordagem baseada em cadeia de blocos;

- Tolerância a falhas: por conta das características dinâmicas originárias de IoT, interrupções nas comunicações (relacionamentos) podem ocorrer, por exemplo, por falta de bateria nos dispositivos, ou devido à mobilidade destes. Assim, as soluções de SIoT devem ser cientes destas características;

Em especial, o desafio de estabelecer confiança entre os dispositivos de SIoT se mostra central e indispensável de ser abordado, de modo que tal paradigma possa continuar evoluindo [15, 23, 24, 113–116].

Apêndice D

Noções de Registro Distribuído

D.1 Registro Distribuído

A tecnologia de registro distribuído (*distributed ledger technology – DLT*) é uma das tecnologias mais promissoras para as próximas gerações de redes, sendo parte essencial, por exemplo, do projeto da sexta geração das redes, o 6G. As DLTs vêm sendo cada vez mais adotadas desde o surgimento da criptomoeda *Bitcoin* que promulgou o uso de cadeia de blocos (um tipo de DLT), de modo a atenderem às novas necessidades de armazenamento e gerenciamento dos dados de sistemas.

Historicamente, os dados de um sistema são organizados e armazenados em base de dados, geralmente relacionais, as quais permitem que operações básicas de criação, leitura, atualização e deleção (*create, read, update, delete – CRUD*) sejam feitas. Todas as operações (transações) realizadas sobre uma base de dados são feitas de maneira atômica, de modo a garantir a consistência dos dados. Tais bases de dados consideram os três modelos básicos de arquitetura, ilustrados na Figura D.1. A arquitetura centralizada oferece o benefício de maior controle dos processos de gerência dos membros (nós) da rede, mas também apresenta como desvantagem ter um único ponto de falha, o qual inviabiliza toda a rede caso fique indisponível. Além disso, nesta arquitetura a escalabilidade do sistema fica restrita de acordo com a capacidade do nó central.

Na arquitetura descentralizada, o controle da gerência dos membros da rede que armazenam os dados é desmembrado em vários centros de controle, o que permite maior escalabilidade. Entretanto, se um desses centros se tornar indisponível, todos os membros dependentes desse centro ficarão fora do sistema.

Já na arquitetura distribuída, os nós que armazenam os dados são conectados a dois ou mais nós, perfazendo uma conectividade típica de um rede *mesh*. Na ocorrência de uma falha de um dos nós, os outros nós vizinhos que continuam a operar podem responder às requisições, o que torna essa estrutura resiliente, tolerante a

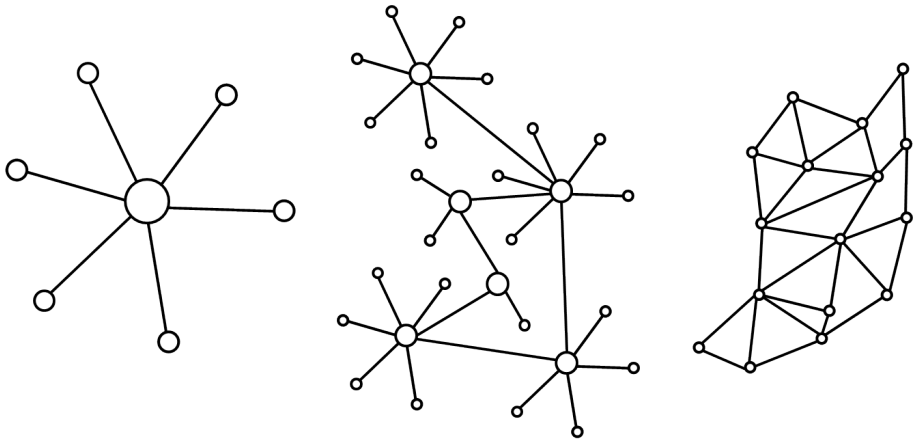


Figura D.1: Tipos de arquitetura de rede: centralizada, descentralizada e distribuída

falhas e altamente disponível. Contudo, há o desafio de manter os dados sincronizados em todos os nós da rede para que todos possam manter o mesmo estado e responder com os mesmos dados solicitados. Por conta da natureza distribuída, é necessário que haja um mecanismo que provisione consenso entre as bases de dados dos nós distribuídos a fim de ter um estado consistente na rede. Tal consenso pode ser impactado por conta de possíveis falhas durante a comunicação entre os membros da rede no processo de sincronização das bases, acarretando no que se chama falhas Bizantinas [117]. Este termo tem origem no problema dos Generais Bizantinos definido por Lamport *et al.* que trata de generais em busca de uma estratégia de comunicação entre eles que seja tolerável a alguns membros não confiáveis que modificam a mensagem original (falhas). Assim, pelo modelo *Practical Byzantine Fault Tolerance* (PBFT), é necessário um número mínimo de $3m + 1$ membros confiáveis para suportar um determinado número m de generais não confiáveis (falhas em membros da rede).

Como um caso especial de base de dados distribuída, as DLTs apenas aceitam que novos dados sejam adicionados, não sendo possível remover ou alterar os dados armazenados. Além disso, as DLTs também assumem a possibilidade de existência de membros maliciosos que tentam subverter as informações armazenadas para benefício próprio, sendo necessária a utilização de um protocolo de consenso para manter a consistência dos dados [118]. Por fim, como comentamos no início deste apêndice, um dos tipos de DLTs mais difundidos é a cadeia de blocos (*blockchain*). A próxima seção trata de mais detalhes sobre essa DLT.

D.2 Cadeia de Blocos

Cadeia de blocos (*blockchain*) é uma tecnologia emergente que permite o registro de transações entre entidades de maneira confiável, auditável e imutável [16]. Inicialmente, o conceito de cadeia de blocos foi empregado no desenvolvimento da primeira e mais difundida criptomoeda, o *Bitcoin* [38]. Logo após a publicação dessa nova criptomoeda, observou-se o potencial da tecnologia de cadeia de blocos, tendo sua utilização expandida e incorporada a aplicações de outros contextos, tais como votações, controle de produtos em cadeia de produção, autenticação de documentos, estabelecimento de confiança e consenso entre partes, entre outros [16].

A cadeia de blocos reúne diversas tecnologias, contemplando áreas como criptografia, algoritmos de consenso, banco de dados distribuído, redes *peer-to-peer* e uma série de outras áreas, as quais são comentadas ao longo do texto.

Como o próprio nome da tecnologia sugere, uma cadeia de blocos é formada por um encadeamento de blocos, no qual cada bloco é composto por um conjunto de transações. Essas transações são validadas (mineradas) por membros de uma rede de validadores (mineradores). Cada bloco é composto não só pelo conjunto de transações, mas também por outros campos necessários para a formação do encadeamento. Assim, os campos de um bloco são:

- uma marca temporal (*timestamp*);
- um número aleatório que só pode ser usado uma única vez, chamado *nonce*;
- um *hash* das transações presentes no bloco;
- o *hash* do bloco anterior.

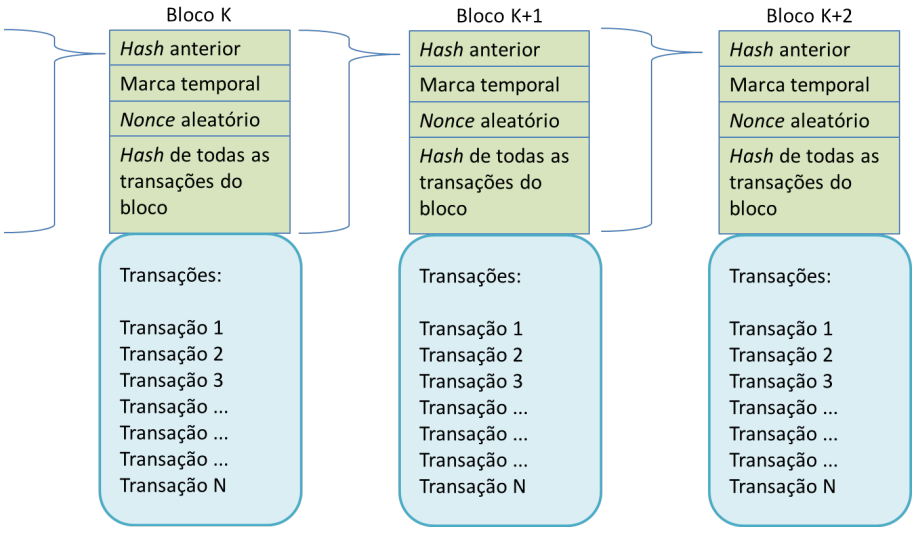


Figura D.2: Representação dos blocos em uma cadeia de blocos

A Figura D.2 ilustra a formação dos blocos de uma cadeia. Para armazenar os blocos de transações, é formada uma rede *peer-to-peer* na qual todos os membros dessa rede (chamados *full nodes*) possuem a mesma base de dados com todas as transações já validadas até então, compondo o que se chama registro distribuído (*distributed ledger*). Cada nova transação é adicionada a um bloco, o qual é validado pelos *full nodes* e, em seguida, adicionado ao histórico de blocos, formando assim a cadeia de blocos.

A tecnologia de cadeia de blocos tem como propósito fundamental armazenar transações entre entidades sem depender de uma terceira parte centralizada para certificar a veracidade de tais transações, como acontece em infraestruturas de chaves públicas (PKI), por exemplo. Os *full nodes*, de maneira distribuída, passam a ser os responsáveis por chegarem a um consenso sobre quais transações são válidas ou não. Esta rede armazena o estado atual considerado como a "verdade" para a rede. Durante uma transação, nenhuma das partes precisa confiar previamente na outra, pois cada uma pode confiar no que os *full nodes* consideram como transação válida.

Para ilustrar os conceitos mencionados, consideramos um cenário de uso da cadeia de blocos para criptomoedas em que Alice deseja transferir uma quantia para Bob. Neste cenário, Alice e Bob não necessariamente se conhecem ou confiam um no outro. Vale ressaltar que Alice e Bob podem ser pessoas, instituições, ou até mesmo um dispositivo, ou computador por detrás de cada uma destas identidades. Como ilustrado na Figura D.3, Alice inicia uma transação que indica que uma quantidade x de moeda deve ser transferida de sua carteira digital (endereço público, ou chave pública) para a carteira de Bob. Para isso, as identidades de Alice e Bob devem ser verificadas, sem, contudo, serem reveladas, o que é feito através do mecanismo de troca de chaves pública-privada. Depois disso, a transação é colocada em um conjunto de transações (*pool*), para então ser de fato validada pela rede.

Para que as transações sejam adicionadas à cadeia de blocos e passarem a fazer parte da "verdade" da rede, é preciso que os membros desta rede provem que o bloco de transações é válido e que pode ser inserido na cadeia. Para isso, os membros da rede competem entre si em uma "corrida" para resolver um problema matemático de difícil resolução[1], sendo dada uma recompensa ao primeiro que resolver tal problema. Este processo é chamado mineração de blocos e o problema a ser resolvido é baseado em um protocolo de consenso. Em particular, considerando a criptomoeda Bitcoin, o protocolo de consenso utilizado é chamado Prova de Trabalho (*Proof of Work* – PoW).

O protocolo de Prova de Trabalho define que deve ser usada uma função resumo (*hash*) no processo de mineração de um bloco. A dificuldade da mineração se dá

[1]Um problema matemático é difícil quando não se conhece um algoritmo eficiente para resolução em tempo polinomial, mas apenas em tempo exponencial.

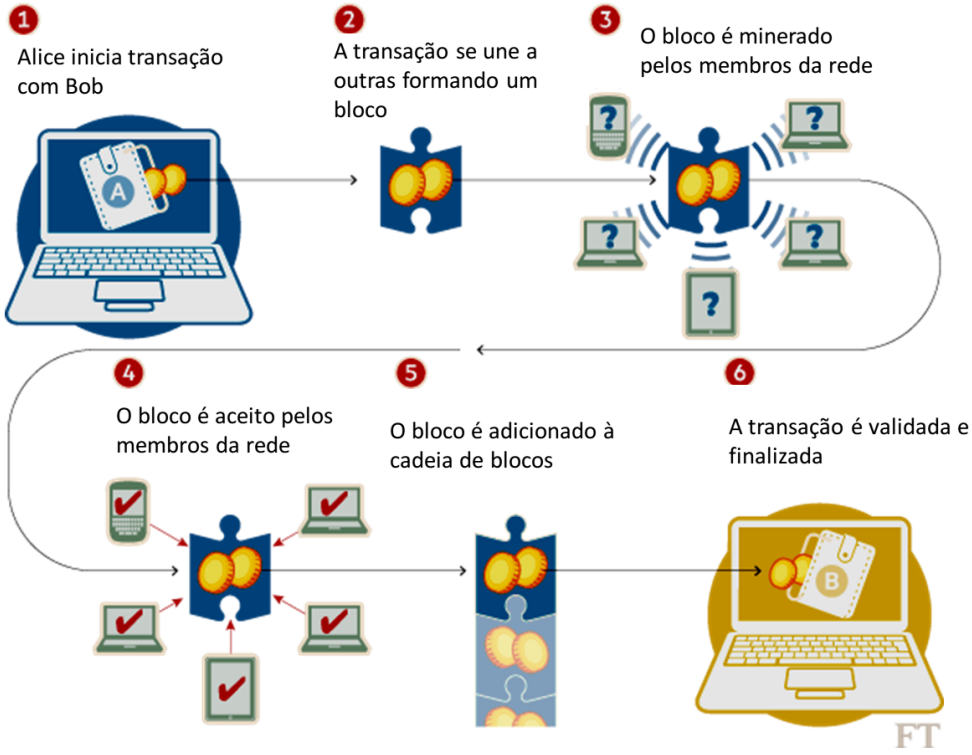


Figura D.3: Exemplo de transação em uma cadeia de blocos. Adaptado de [119]

pela função *hash* ser do tipo *one-way*, a qual dado um valor $y = f(x)$, é fácil obter y utilizando a função $f(.)$, porém é muito difícil obter x a partir de y. A única forma de descobrir x é por força bruta, o que consiste em tentar todas as possíveis soluções até encontrar a que resolva a tarefa. Analogamente, é como obter um número específico em um lançamento de dados, o que pode requerer várias tentativas antes de se alcançar o resultado desejado. Como os campos de um bloco não podem ser modificados, então inseriu-se o *nonce*[2] como campo para ser variado até se obter o *hash* correto. O protocolo de consenso PoW ainda especifica a característica que o *hash* alvo deve ter, definindo o número de zeros que o *hash* inicia. Assim, a mineração de um bloco consiste em descobrir um *nonce* que sirva para obter um *hash* que comece com uma quantidade de zeros predefinida para um bloco. A dificuldade de mineração é ajustada variando o número de zeros iniciais que o *hash* procurado deve conter.

Quando um membro da rede encontra o *nonce* que é solução para o *hash* do bloco sendo minerado, este divulga a solução para os outros membros da rede. Então, se uma quantidade significativa do poder computacional (acima de 50%) dos membros da rede concordarem com a solução apresentada, então o membro que achou a solução é recompensado com uma quantia da criptomoeda e os outros membros

[2]Um número aleatório que só pode ser usado uma única vez.

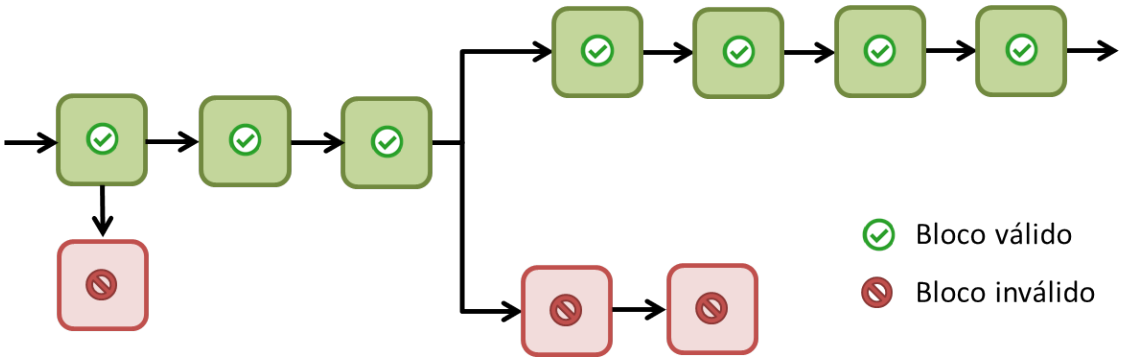


Figura D.4: Representação de ramificações de cadeias, com a cadeia longa sendo a vencedora.

atualizam seus bancos de dados, adicionando o novo bloco minerado. Desta forma, a rede entra em consenso, concordando com as novas transações pertencentes ao bloco.

Após estes passos, a transação do exemplo entre Alice e Bob foi inserida na cadeia de blocos. Entretanto, esta transação ainda pode ser substituída caso outros blocos sejam minerados em outro ramo da cadeia e esta "vença a corrida" pela cadeia mais longa, como ilustrado na Figura D.4. A cadeia mais longa é a que contém a "verdade" da rede, pois recebeu mais provas de trabalho em cada bloco minerado. Em outras palavras, a maioria dos membros da rede sempre concordará com a cadeia mais longa e terá esse conjunto de transações como sendo válido. Neste exemplo, a rede só poderia ser adulterada se pelo menos 51% do poder computacional dos membros da rede fosse direcionado a concordar com a cadeia inválida, o que caracteriza o ataque que uma cadeia de blocos pode sofrer, conhecido como ataque dos 51% [120].

Um dos motivos pelo qual a transação entre Alice e Bob pode não fazer parte da cadeia mais longa é devido ao problema do gasto duplo (*double spending*). Este problema pode ocorrer em uma cadeia de blocos, mas é resolvido pela PoW. Como se trata de um ativo digital (uma moeda no caso do *Bitcoin*), esta poderia ser facilmente copiada por um atacante, podendo enviar duas transações ao mesmo tempo para diferentes destinatários, consumindo o mesmo ativo duas vezes. Quando isso ocorre, uma ramificação é criada na cadeia, o que leva à necessidade de aguardar uma quantidade suficiente de confirmações da rede sobre a validade da transação, até considerá-la como válida de fato. As confirmações são baseadas na quantidade de blocos inseridos após o bloco que possui a transação desejada. A cada bloco adicionado, a probabilidade da transação ser originada a partir de um gasto duplo é reduzida exponencialmente, sendo seis a quantidade mínima sugerida de confirmações antes de considerar a transação como válida [38].

Por conta da inserção do *hash* de um bloco como parte da formação de um bloco seguinte, isso cria um encadeamento entre os blocos, o que torna a cadeia de

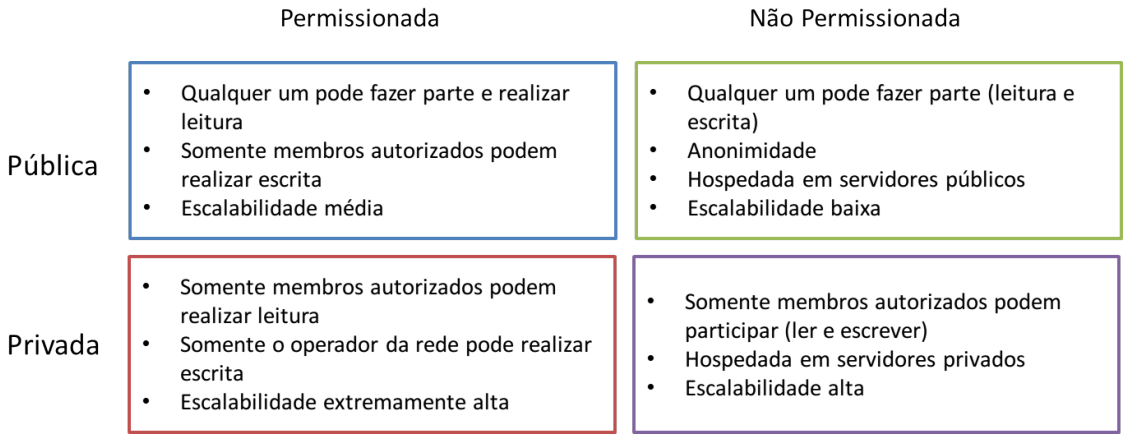


Figura D.5: Tipos de cadeias de blocos

blocos à prova de adulteração, visto que qualquer modificação feita em um bloco resultará em um novo valor de *hash*. Com isso, todos os *hashes* dos blocos seguintes precisam também ser recalculados e difundidos novamente por toda rede. Como essa é uma tarefa extremamente difícil (pelo ataque dos 51%), as informações mantidas na base de dados da rede são praticamente invioláveis. Além disso, a rede não pertence a um grupo ou instituição específicos, prevenindo que a segurança e a privacidade dos dados sejam comprometidas. É válido lembrar que em 2013, como revelado por Edward Snowden [121], programas de espionagem exploravam dados de pessoas com diversos propósitos, justamente por conta da centralização de serviços em companhias específicas.

A arquitetura distribuída, na qual a cadeia de blocos se baseia, compõe um sistema em forma de uma malha de conectividade, o que torna o sistema robusto ao ponto que, se um membro se torna indisponível, apenas aquele membro é afetado e a rede continua operacional. A cadeia de blocos se vale desta arquitetura distribuída através de uma rede *peer-to-peer*. Outros tipos de arquitetura não oferecem as mesmas características.

Em relação aos tipos de cadeias de blocos, existem as versões pública e privada, podendo ser permissionada ou não. Essa classificação está relacionada com quem terá a posse dos dados armazenados (se será público ou pertencente a uma instituição), bem como com as permissões de leitura e escrita que os participantes da rede terão (se qualquer membro pode ter acesso livre, ou se apenas membros autorizados podem fazer ações permitidas). A Figura D.5 ilustra as quatro opções de cadeias de blocos, apontando as principais características em cada uma das opções.

Apêndice E

Propriedade de Regularidade Estatística

A propriedade da regularidade estatística afirma que médias obtidas em longas sequências de repetição (tentativas, ou amostras) de experimentos aleatórios convergem para valores específicos [68]. Esses valores são dados inicialmente pela frequência relativa que varia cada vez menos à medida que o número de tentativas aumenta. No limite, quando o número de tentativas tende ao infinito, a frequência relativa tende à probabilidade do resultado específico. A lei dos grandes números reafirma esta propriedade.

Leon-Garcia em [68], ilustra a propriedade de regularidade estatística através de um experimento que reproduziremos aqui e depois utilizaremos como base para outro experimento usando o tráfego entre dispositivos IoT. No experimento apresentado por Leon-Garcia, três bolas idênticas rotuladas como 1, 2 e 3 são consideradas dentro de uma urna. Uma bola é selecionada após sacudir a urna (aleatoriedade do experimento), seu rótulo é anotado e a bola retorna à urna. Então, todos os resultados possíveis deste experimento são os rótulos das bolas, representado pelo conjunto $S = \{1, 2, 3\}$.

À medida que o experimento é repetido, os rótulos são anotados e o número de tentativas (amostras) aumenta, conforme ilustrado na Figura E.1. As frequências relativas começam com grandes variações, mas à medida que o número de amostras aumenta, cada frequência relativa de cada resultado possível converge para $0,3$. A Figura E.2 mostra o experimento com 1.000 amostras.

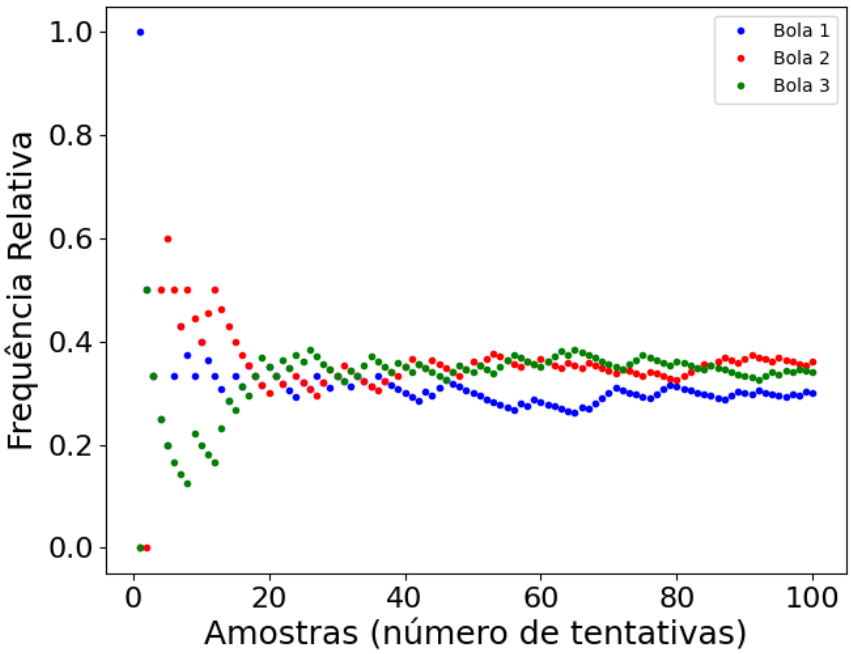


Figura E.1: Experimento com 100 amostra (repetições)

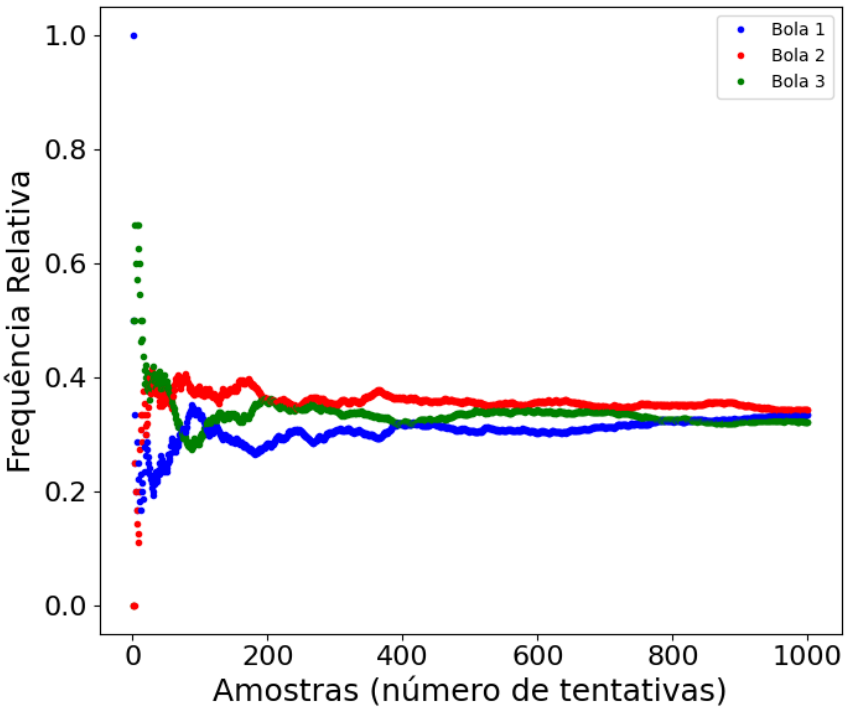


Figura E.2: Experimento com 1000 amostra (repetições)

E.1 Experimento com Tráfego entre Dispositivos IoT

Semelhante ao experimento anterior, desenvolvemos um experimento usando o tráfego entre dispositivos IoT. Com até 600 amostras, vemos na Figura E.3 as frequências relativas do tráfego de um dispositivo para outro à medida que o número de amostras aumenta.

Consideramos que os valores de tráfego foram quantizados em 10 intervalos de 10 KBps cada, com tráfego máximo de 100 KBps, *i. e.*, o tráfego pode ficar em 10 faixas possíveis, de 0 a 10 KBps, de 10 KBps a 20 KBps, e assim por diante. Isto seria o equivalente a termos 10 bolas possíveis de serem sorteadas no experimento anterior. Próximo da amostra 300 vemos que há alguma anomalia a partir desta amostra, cujo impacto altera a frequência relativa durante um período.

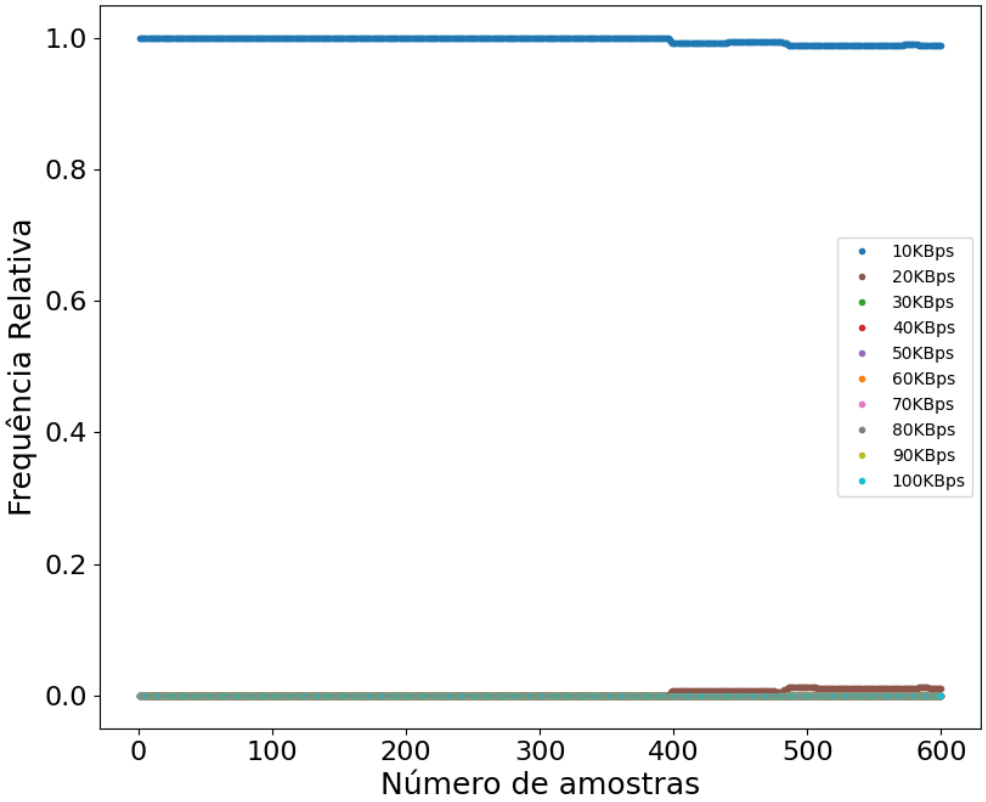


Figura E.3: Experimento das frequências relativas do tráfego com 600 amostras

Fazendo outro experimento, desta vez com 25000 amostras de tráfego, podemos perceber que, a partir da amostra 10000, a frequência relativa do primeiro intervalo de 0 a 10 KBps sofre uma mudança significativa, por conta de uma anomalia em tal dispositivo. Este comportamento viola a lei dos grandes números, indicando que as frequências relativas não convergiram para os respectivos valores de probabilidade. Novamente, de maneira análoga ao experimento das bolas, é como se a quantidade

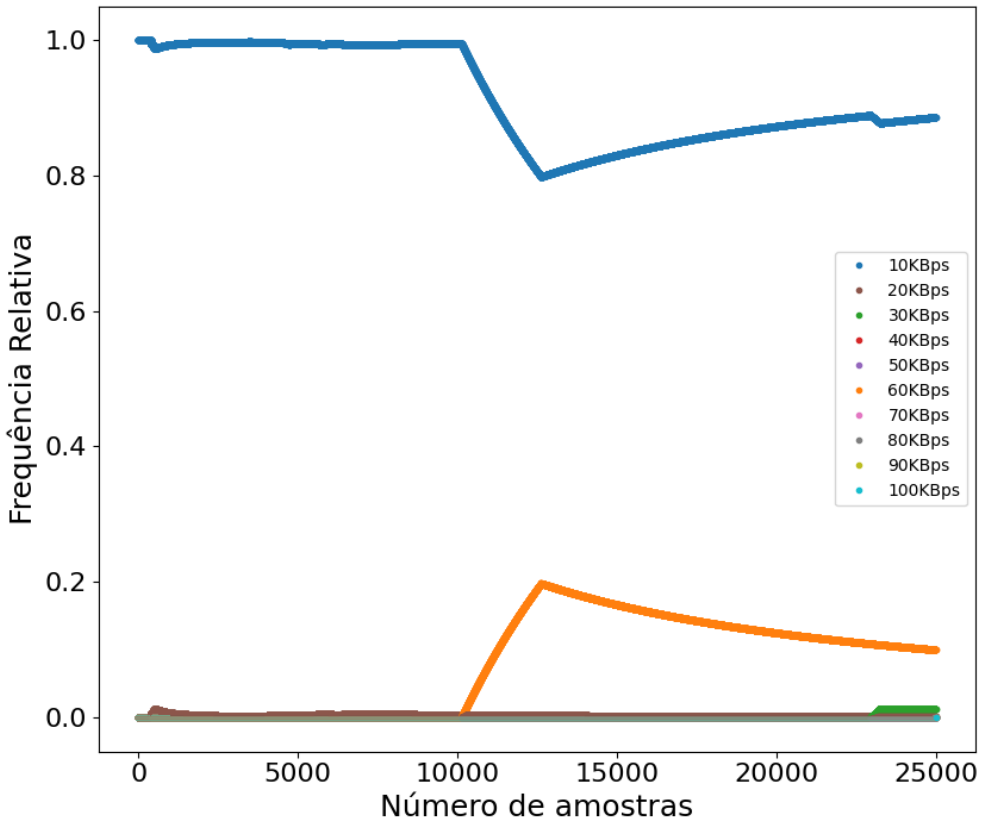


Figura E.4: Experimento das frequências relativas do tráfego com 25000 amostras

de bolas de um determinado rótulo mudasse dentro da urna ao longo do tempo. Assim, o objetivo deste experimento é mostrar que, no caso de tráfego de dispositivos, o aumento do número de amostras não garante a manutenção da propriedade de regularidade estatística, violando a lei dos grandes números. Por conta disso, adotamos um número suficientemente grande de amostras, como no exemplo anterior com 600 amostras, mas não tanto quanto 25000. Isso também corrobora com a utilização de um número razoável de amostras considerando as restrições de capacidade muito comuns dos dispositivos IoT.