Configuração de redes com linguagem natural apoiadas na identificação dos dispositivos: Um mapeamento sistemático

Alex Santos[1], Geraldo P. Rocha Filho[2], Rodolfo Ipolito Meneguette[3], Roger Immich[1]

[1]Instituto Metrópole Digital
Universidade Federal do Rio Grande do Norte (UFRN)

[2]Departamento de Ciências Exatas e Tecnológicas
Universidade Estadual do Sudoeste da Bahia (UESB)

[3]Instituto de Ciências Matemáticas e de Computação
Universidade de São Paulo (ICMC-USP)

Resumo. *A proliferação de dispositivos inteligentes em ambientes residenciais tem ensejado uma crescente demanda por controle de acesso à rede. Este requisito pode ser motivado por usuários que almejam bloquear o acesso a determinados dispositivos por razões de controle parental ou para prevenir acessos não autorizados. Apesar de os roteadores domiciliares disponibilizarem recursos para tal controle, muitos usuários leigos se deparam com entraves, uma vez que não detêm conhecimento técnico suficiente para configurar estes equipamentos e, frequentemente, a interface de configuração carece de padronização. Com o intuito de superar este obstáculo, a linguagem natural emerge como alternativa mais acessível para usuários não técnicos, permitindo-lhes descrever regras de controle de acesso para seus dispositivos utilizando sua língua materna. Neste trabalho, é proposta uma revisão sistemática que tem como objetivo analisar o estado da arte e as lacunas em relação à configuração de redes utilizando linguagem natural, além de investigar o conhecimento atual sobre a identificação automática de tipos de dispositivos e questões de segurança em redes residenciais. Os resultados da pesquisa indicam que o uso da linguagem natural para configuração de redes ainda é limitado às redes definidas por software e pouco explorado no contexto residencial. Ademais, foi constatado que nenhum estudo até o momento buscou integrar o controle de acesso por tipo de dispositivo por meio da linguagem natural.*

Palavras-chave.: linguagem natural, linguagem de intenção, configuração de rede, identificação de dispositivos, tipo de dispositivo

Abstract. *The proliferation of intelligent devices in residential environments has led to a growing demand for network access control. This requirement may be motivated by users who seek to block access to certain devices for reasons of parental control or to prevent unauthorized access. Although home routers provide resources for such control, many non-technical users face obstacles, as they lack sufficient technical knowledge to configure these devices and often the configuration interface lacks standardization. In order to overcome this issue, natural language emerges as a more accessible alternative for non-technical users, allowing them to describe access control rules for their devices using their native*

language. This work proposes a systematic review aimed at analyzing the state of the art and gaps in relation to the configuration of networks using natural language, as well as investigating current knowledge about automatic identification of device types and security issues in residential networks. The research results indicate that the use of natural language for network configuration is still limited to software-defined networks and little explored in the residential context. Moreover, it was found that no study to date has sought to integrate device-type access control through natural language.
keywords.: natural language, intent language, network configuration, device identification, device-type

1. Introdução

Com difusão dos dispositivos inteligentes e das redes sem fio em ambientes residenciais, a presença de tais dispositivos nesses ambientes tem aumentado consideravelmente [do Prado and et al 2021, Rodrigues and et al 2019]. Como consequência, tem-se observado uma crescente demanda pelo controle de acesso desses dispositivos à rede [Fiorenza and et al 2021]. Esse controle de acesso pode ser motivado por usuários que desejam bloquear o acesso de determinados dispositivos ou restringir o acesso a certos horários, como no caso de pais que buscam controlar o acesso à Internet por seus filhos [Alsudais and Keller 2017], ou ainda impedir acessos indesejados aos dispositivos de IoT (*Internet of Things*) [Pisani and et al 2020] presentes na casa.

Não obstante alguns fabricantes de roteadores residenciais oferecerem recursos de configuração que permitem um certo nível de controle de acesso ou controle parental, esses equipamentos não seguem um padrão uniforme de interface de configuração. Tal disparidade representa um obstáculo para usuários leigos, os quais frequentemente carecem de conhecimento técnico para configurar seus dispositivos [Jacobs et al. 2019]. A falta de padronização das interfaces de configuração pode, ainda, apresentar dificuldades de acessibilidade para usuários com baixa visão ou outras limitações visuais [Cesário and et al 2022, Neto and et al 2021].

As interfaces de configuração dos roteadores residenciais geralmente apresentam recursos limitados e fornecem apenas informações básicas sobre os dispositivos conectados à rede do usuário. É importante ressaltar que elas não são capazes de identificar ou especificar as funcionalidades dos dispositivos presentes na rede e, por conseguinte, não é possível determinar com precisão quais tipos de acesso devem ser permitidos a esses dispositivos. Dessa forma, os usuários enfrentam uma barreira para estabelecer regras de controle de acesso com base nas funcionalidades dos dispositivos, tais como proibir o acesso externo a câmeras IP, bloquear o acesso a jogos eletrônicos em horários pré-definidos ou priorizar o tráfego de rede de determinados tipos de dispositivos, como computadores.

É imprescindível destacar que usuários leigos não possuem a obrigação de possuir conhecimento técnico para configurar uma rede. No entanto, seria desejável que tais usuários pudessem, por meio de texto ou voz, fornecer instruções em alto nível, a fim de que o sistema possa traduzi-las para uma linguagem configurável em equipamentos de rede [Cesário and et al 2022].

Nesse contexto, o objetivo do presente estudo é realizar uma revisão sistemática acerca do emprego da linguagem natural para o controle de acesso de dispositivos em

roteadores residenciais, bem como dos métodos de identificação automática desses dispositivos na rede e das questões associadas à segurança de dispositivos IoT em redes residenciais. Este estudo identificou uma lacuna na literatura sobre a utilização de linguagem natural para configuração de redes domésticas e a identificação de tipos de dispositivos para controle de acesso. Embora existam trabalhos sobre esses temas, nenhum estudo abordou ambos de forma integrada até o momento, sendo essa a principal contribuição deste trabalho.

O artigo está organizado da seguinte forma. A Seções 2 e 3 apresentam a metodologia e processos de pesquisa empregados no mapeamento sistemático. A Seção 4 apresenta caraterísticas pertinentes identificadas nos estudos selecionados e respondem às questões de pesquisas propostas. Finalmente, a Seção 5 sintetiza as descobertas identificadas e apresenta uma proposta para uma lacuna identificada com a pesquisa.

2. Metodologia adotada

Foi realizada uma pesquisa exploratória com a análise de artigos científicos publicados entre 2017 e 2022. O protocolo utilizado para orientar a revisão sistemática seguiu o planejamento, condução e análise estabelecidos por [Kitchenham and Charters 2007]. Para orientar o mapeamento sistemático, foram utilizadas as técnicas de triagem e classificação adotadas por [Petersen et al. 2008], que serão explicadas com mais detalhes nas próximas seções.

Este levantamento seguiu três abordagens principais, que são (1) analisar o estado da arte com relação a configuração de redes tendo como entrada a linguagem natural, (2) explorar o estado atual de conhecimento da identificação automática de tipos de dispositivos e (3) investigar questões ligadas à segurança desses dispositivos inteligentes em relação a outros dispositivos em redes residenciais. As questões de pesquisa consideradas neste trabalho para delimitar a pesquisa estão descritas abaixo.

- QP1 A linguagem natural vem sendo utilizada para capturar as intenções dos usuários em realizar controle de acesso e/ou permitir a configuração de roteadores/firewall residenciais?
- QP2 Qual o propósito da utilização da linguagem natural nos estudos existentes e quais são os usuários dessas soluções?
- QP3 Quais os métodos de identificação de dispositivos são utilizados na literatura para identificação do tipo de dispositivo em uma rede? Qual categorizações de tipos de dispositivos são adotadas?
- QP4 Os dispositivos inteligentes podem tornar uma rede doméstica desprotegida. Considerando o contexto de identificação de dispositivos, o que tem sido feito para tornar as redes nas quais os dispositivos IoT estão inseridos mais seguras?

3. Processo de pesquisa

A busca por estudos foi realizada na base Scopus. Durante esta fase, foi percebido que as questões de pesquisa levariam a duas frentes de pesquisa distintas.

I. Uso da linguagem natural para configuração de equipamentos de rede domésticos;
II. Identificação de tipos de dispositivos e questões de segurança

A Tabela 1 exibe as *strings* de busca consideradas no processo de pesquisa.

No processo de definição da *string* de busca para a frente I, foi utilizado uma *string* (a) que procurou abranger trabalhos relacionados ao uso da linguagem natural ou da

intenção para configuração e gerenciamento de redes, e que fossem capazes de responder às questões de pesquisa QP1 e QP2.

Tabela 1. Seleção das *strings* de busca

#	String de busca	Resultados
a	((“natural language” OR nlp) OR (intent AND (language OR recognition OR detection))) AND (“network management” OR “access control” OR “network configuration” OR “home network*” OR “smart device*”)	301
b	(“device-type identification” AND security)	17

Para a frente II de pesquisa, foi utilizada a *string* de busca (b) que listasse trabalhos que procuraram responder as questões de pesquisa QP3 e QP4. Portanto, este trabalho utilizou as *strings* de busca (a) e (b), pesquisando por o título, resumo e palavras-chave.

3.1. Critérios de Inclusão e Exclusão

Os critérios de inclusão foram (1) Estudos cujo conteúdo esteja relacionado ao contexto de pesquisa; e (2) Estudos que respondem as perguntas da pesquisa. Os critérios de exclusão foram (1) Trabalhos duplicados; (2) Trabalhos do tipo *Conference review*; (3) Trabalhos publicados antes de 2017; (4) Trabalhos não escritos em inglês ou português; (5) Trabalhos em que o título ou resumo não tem relação ao contexto da pesquisa; e (6) Trabalhos com condição de acesso mediante pagamento ou inacessíveis.

3.2. Coleta de dados e seleção de trabalhos

Com a *string* de busca (a), esperou-se obter a maior quantidade de trabalhos que utilizaram linguagem natural ou de intenção para realizar qualquer tipo de gestão de rede, controle de acesso, configuração de rede, utilização em redes domésticas ou no contexto de dispositivos inteligentes. A partir desta busca obteve-se $n = 301$ estudos. Os n estudos foram exportados para a ferramenta Parsifal para o refinamento de acordo com os critérios definidos na pesquisa. Ao aplicar os critérios de inclusão e exclusão, 31 foram selecionados para uma nova seleção. Os 31 artigos tiveram seus dados de referência e documentos PDF (*Portable Document Format*) coletados com o auxílio da ferramenta Zotero para facilitar a próxima etapa de seleção: a leitura na íntegra de resumo, introdução e conclusão.

Com a *string* de busca (b), utilizou-se as mesmas ferramentas e esperou-se obter estudos que realizaram a identificação de tipos de dispositivos e que fizessem alguma proposta de segurança relacionado a QP4.

Após a leitura da introdução e conclusão dos 31 trabalhos da frente I para identificar o contexto das publicações, foram selecionados 15 como trabalhos primários. Com relação a frente de pesquisa II, foram selecionados 9 trabalhos de um total de 17 trabalhos. Dessa forma, um total de 24 estudos foram selecionados como trabalhos primários.

A Figura 1 ilustra as etapas do processo de refinamento para seleção dos trabalhos acadêmicos.

As Tabelas 2 e 3 listam os trabalhos selecionados para análise, bem como as suas principais características que serão discutidas.

4. Análise de dados e Resultados

Primeiramente, serão analisadas características identificadas nos estudos selecionados para a Frente I, apresentados na Tabela 2. Posteriormente, serão discutidos aspectos

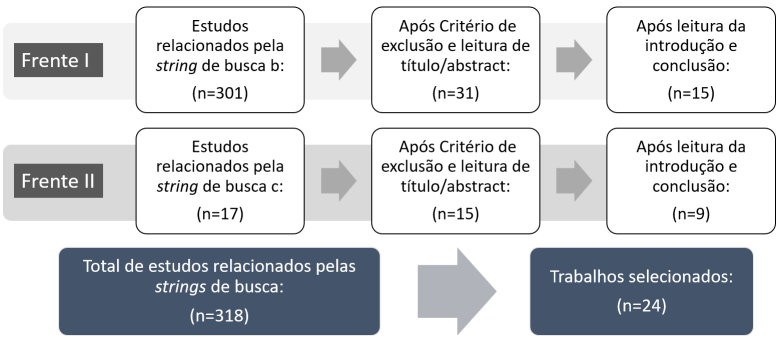

Figura 1. Etapas do mapeamento sistemático

Tabela 2. Trabalhos da Frente I

ID	Referência	Linguagem de intenção ou Representação intermediária	Interface de conversação	Propósito do uso da linguagem natural	Feedback	Descoberta de dispositivos	Segurança Smart devices	Rede de aplicação	Usuário autor da intenção
1	[Alsudais and Keller 2017]	-	Própria	Configuração de SDN	-	-	-	Redes corporativas; Redes domésticas*	Operadores de redes SDN
2	[Liu et al. 2017]	✔ (IR)	EasyACL	Configuração de ACLs	-	-	-	Redes corporativas	Operadores de redes cisco ou Juniper
3	[Jacobs et al. 2019]	✔ (Nile)	DialogFlow	Configuração de SDN	✔	-	-	Redes campus (CAN)	Operadores de redes SDN
4	[Tian and et al 2019]	✔ (LAI)*	Jinjing	Configuração de SDN	✔	-	-	Redes de operadoras WAN	Operadores de redes WAN
5	[Ribeiro 2020]	✔ (Nile)	SCRIBE	Exibição de configuração ACL e NAT de equip.	-	-	-	Redes de operadoras WAN; Redes corporativas	Operadores de redes
6	[Rivera et al. 2020]	✔ (POLANCO)*	-	Configuração de SDN	-	✔	-	Redes corporativas; Redes campus (CAN)	Criadores de políticas de segurança
7	[Zeydan and Turk 2020]	*Survey / Revisão*							
8	[Jacobs and et al 2021]	✔ (Nile)	DialogFlow	Configuração de SDN	✔	-	-	Redes campus (CAN)	Operadores de redes SDN
9	[Paudel and et al 2021]	-	-	Extração de funcionalidades de dispositivos IoT	-	-	✔	Redes domésticas	Usuários finais
10	[Shi et al. 2021]	✔ (JSON)	DialogFlow	Verificação de cumprimento de políticas de rede	-	✔	-	Redes corporativas; Redes campus (CAN)	Criadores de políticas de segurança; Operadores de rede
11	[Leivadeas and Falkner 2022]	*Survey / Revisão*							
12	[Nguyen et al. 2022]	✔ (Nile, SNIL)	Dialogflow	Configuração de SDN	✔	-	-	Internet exchange point (IXP)	Operadores de redes SDN
13	[Ribeiro and et al 2022]	✔ (Nile)	SCRIBE	Exibição de configuração ACL e NAT de equip.	-	-	-	Redes de operadoras WAN; Redes corporativas	Operadores de redes
14	[Saha and et al 2022]	-	Própria	Configuração de SDN	✔	-	-	Redes corporativas; Redes campus (CAN)	Operadores de rede SDN
15	[Xiao and et al 2022]	✔ (JSON)	Própria	Configuração de SDN	✔	-	-	Redes corporativas; Redes campus (CAN)	Operadores de rede SDN

Tabela 3. Trabalhos da Frente II e identificação de dispositivos (Frente I)

	Referência	Tipo de monito-ramento	Processamento da identificação	Métodos/Abordagem para descoberta de tipo de dispositivo	Origem das características (features) extraídas	Estado de atividade dos dispositivos para identificação	Classes definidas	Estratégia de segurança
Trabalhos relacionados à identificação de dispositivos da Frente I								
6	[Rivera et al. 2020]	Ativo; Passivo	Borda	Buscou somente identificar topologia e dispositivos individuais	Pacotes LLDB, BDDP, SNMP, ARP e DHCP	- Ativo; - Standby	-	-
10	[Shi et al. 2021]	Passivo	Borda	Buscou somente identificar dispositivos individuais	Presumivelmente cabeçalhos de pacotes das camadas 3 e 4	- Ativo; - Standby	-	-
Trabalhos da Frente II								
16, 17	[Miettinen and et al 2017]	Passivo	Fingerprinting: Borda Classificação: Nuvem	- Device Fingerprint; - Classificação supervisionada baseada em ML	Cabeçalhos de pacotes das camadas 2, 3, e 4 (23 features)	- Inicialização na rede; - Configuração	27 para tipos de dispositivo	- Isolamento de rede; - Regras de filtragem de tráfego; - Notificação ao usuário
18	[Salman et al. 2018]	Passivo	Pré-proc.: Borda Classificação: Borda	- Pré-processamento com extração de estatísticas do fluxo de rede - Classificação supervisionada baseada em ML	Dados estatísticos de cabeçalhos de pacotes da camadas 4 (36 features)	- Ativo	- 7 para tipos de dispositivo; - 3 para funções	- Detecção de tráfego anormal do dispositivo; - Regras de bloqueio de tráfego
19	[Yu and et al 2018]	Passivo	Fingerprinting: Borda Classificação: Borda	- Device Fingerprint; - Classificação supervisionada baseada em ML	Pacotes DHCP (18 features)	- Inicialização na rede;	21 para tipos de dispositivo	-
20	[Ammar et al. 2019]	Ativo; Passivo	Fingerprinting: Borda Classificação: manual	- Device Fingerprint (Bag Of Words) - Classificação manual	Pacotes DHCP, UPnP, HTTP e registros mDNS	- Inicialização na rede;	Não definiu classes. Apenas utilizou o dataset de #16, #17 para análises	- Isolamento de rede
21	[Marchal and et al 2019]	Passivo	Fingerprinting: Borda Classificação: Nuvem	- Device Fingerprint; - Classificação não supervisionada baseada em ML	Todo o fluxo até camada 4 (33 features)	- Inicialização na rede; - Standby; - Interação com usuário;	Não define. As classes são geradas automati-camente com rótulos genéricos sequenciais	- Políticas baseadas em tipo de dispositivo; - Isolamento de rede; - Detecção de tráfego anormal do dispositivo
22	[Hohum and et al 2021]	Passivo	Pré-proc.: Borda Classificação: Borda	- Pré-processamento baseado no endereçamento; - Classificação supervisionada baseada em ML	Cabeçalhos de pacotes das camadas 3 e 4.	- Inicialização na rede; - Ativo; - Standby;	8 para tipos de dispositivo	-
23	[Sánchez and et al 2021]	*Survey / Revisão*						
24	[Salman and et 2022]	Passivo	Pré-proc.: Borda Classificação: Borda	- Pré-processamento baseado no tamanho do pacote, timestamp, direção e protocolo de transporte; - Classificação supervisionada baseada em ML	Dados estatísticos de cabeçalhos de pacotes das camadas 3 e 4 (4 features)	- Ativo	- 7 para tipos de dispositivo; - 3 para funções; - 7 para tipos de ataque; - 2 para classe do tráfego	- Detecção de tráfego anormal do dispositivo

relevantes identificados na Frente II e apresentados na Tabela 3. É importante observar que a maioria dos trabalhos que utilizam a linguagem natural como entrada preveem em sua arquitetura um mecanismo de tradução que identifica palavras-chaves da descrição em alto nível e procura extrair entidades predefinidas (todos os trabalhos, exceto #5, #6 e #9). Em alguns casos este trabalho é facilitado com a utilização de inteligência artificial (#3, #8, #10 e #12).

Nos mecanismos de tradução avaliados na Frente I, frequentemente é adotada uma linguagem (ou representação) intermediária de alto nível como transição entre a linguagem natural e o subsistema de execução da intenção. Para este efeito, podemos destacar a linguagem de definição de intenção Nile, criada no trabalho #3 e utilizada ou estendida em #5, #8, #12 e #13. Outras criações de linguagem de intenção também foram propostas no trabalho #4 com LAI (*Language for ACL Intents*) e #6 com POLANCO (*POlicy LANguage for Campus Operations*). O trabalho #2 também utilizou uma estratégia de representação intermediária, porém bem mais simplificada. A primeira proposta de desenho de arquitetura com a previsão de uma camada de abstração entre os usuários e a rede foi apresentada em #1.

O processo de tradução de linguagem de alto nível, como a voz ou texto, é passível a imprecisões ou erros seja por ausência de informação por parte do usuário ou falha no processo de tradução [Shi et al. 2021]. Para lidar com isso, parte dos trabalhos propõe algum tipo de revisão, por parte do usuário, sobre a intenção traduzida. Os autores denominam esse processo como *feedback*. Em alguns trabalhos, este processo é utilizado apenas para validação/confirmação da intenção. Em outros, o sistema baseado em intenção pode ser capaz de aprender com eles, e isso é feito nos trabalhos #8, #12 e #15.

Com relação aos métodos de identificação de tipos de dispositivos e questões de segurança, (Frente II - Tabela 3), é pertinente pontuar que para a realização de identificação de dispositivos, em todos os estudos foram utilizadas estratégias passivas de monitoramento do tráfego, como o TCPdump. Por este motivo, os desenhos de arquitetura habitualmente planejam um módulo de análise de tráfego fazendo parte de um *gateway*. Assim, existirá maior garantia de que todo o fluxo de dados irá passar pelo analisador de pacotes da solução.

Outra particularidade é que as propostas monitoram o fluxo de dados do dispositivo para extrair características e criar uma impressão digital deste. Essa abordagem é chamada de *device fingerprint* ou *fingerprinting* (#16, #17, #19, #20 e #21). Outros autores chamaram esta etapa de pré-processamento (#18, #22 e #24). No escopo deste trabalho, trataremos esses termos como sinônimos para definir uma forma de extrair características de um fluxo de dados que possam identificar um dispositivo como pertencente a uma classe. Nos métodos passivos o endereço MAC de origem tende a utilizado para delimitar o pertencimento dos fluxos que passam pelo *gateway* a um dispositivo. Essa conduta foi explicitamente adotada nos trabalhos #16, #17, #21 e #22.

4.1. [QP1] A linguagem natural vem sendo utilizada para capturar as intenções dos usuários em realizar controle de acesso e/ou permitir a configuração de roteadores/firewall residenciais?

Alguns trabalhos, como #3 e #8, apontam que os usuários domésticos poderiam fazer uso de uma linguagem de intenção para configuração de redes domésticas. Embora esses trabalhos citem os usuários redes domésticas como potenciais utilizadores de um sistema baseado em intenção, seus trabalhos não exploraram isso. O trabalho #1 cita que

a linguagem natural poderia ser utilizada para controle parental no contexto residencial. Na prática, a implantação da proposta do autor no contexto residencial exigiria que o usuário tivesse conhecimento técnico dos endereços IP dos dispositivos alvo. A proposta #9 trata da possibilidade de extração de funcionalidades de dispositivos IoT a partir de informações públicas sobre os dispositivos. O objetivo dessa extração seria informar ao usuário se aquele dispositivo tem alguma funcionalidade não óbvia que poderia comprometer sua segurança com relação à privacidade. Desta forma, foi constatado que existe uma falta de exploração da linguagem natural na ocasião de configuração, gestão de redes domésticas ou controle de acesso de dispositivos inteligentes.

4.2. [QP2] Qual o propósito da utilização da linguagem natural ou de intenção nos estudos existentes e quais são os usuários dessas soluções?

A maior parte das propostas sugere a criação de um sistema baseado em intenção para gerenciar uma rede baseada por intenção, programada por meio de um controlador SDN (*Software-Defined Networking*), como apresentado na Figura 2. Além disso, conforme Figura 3, a maioria dos usuários de soluções de configuração de redes por intenção é de uma categoria de usuários conhecida como operadores de rede *(network operators)*.

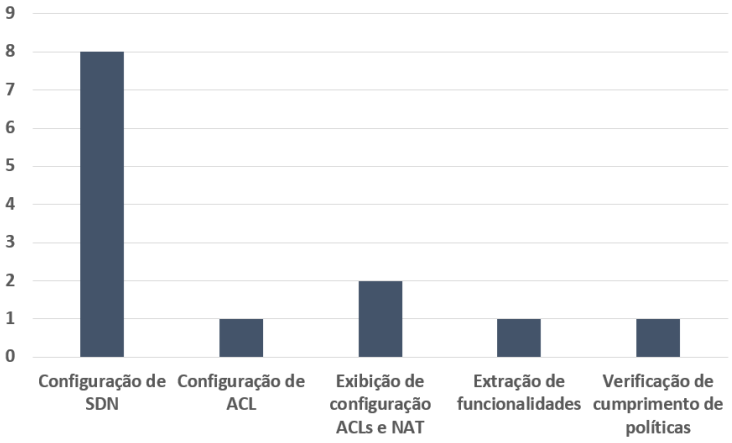

Figura 2. Propósito do uso da linguagem natural

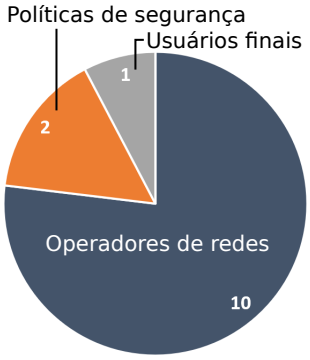

Figura 3. Usuários das propostas

Duas soluções, #6 e #10, apontam como seus usuários os criadores de políticas de segurança ou referenciam estes usuários com nomenclaturas similares como *Policy Writing Committees*. Estes utilizadores são usuários não especialistas da infraestrutura e definem suas declarações de segurança em alto nível. O único trabalho que tem como utilizador um usuário final não especialista e potencialmente residencial é o trabalho #9. Este porém, foca em extrair informações públicas de funcionalidade dos dispositivos inteligentes, e não para a configuração da rede na qual dos dispositivos estariam implantados. Portanto, as soluções existentes de sistemas baseados em intenção estão focadas principalmente em operadores de redes mais complexas e pouco estão focadas em usuários finais, principalmente os residenciais.

4.3. [QP3] Quais os métodos de identificação de dispositivos são utilizados na literatura para identificação do tipo de dispositivo em uma rede? Que categorizações de tipos de dispositivos são adotadas?

Na Frente I, apenas dois trabalhos utilizaram um método para identificação de dispositivos na rede (estudos #6 e #10). No primeiro, o autor utiliza recursos disponíveis em

um controlador SDN para realizar descoberta de topologia e fazer inspeção de pacotes. O segundo trabalho utilizou ferramentas como Netflow, TCPdump e ELK stack com elasticsearch para a coleta de fluxo de rede a nível de pacote. Nenhum desses dois trabalhos mencionou uma arquitetura de tipos de dispositivos.

Já os trabalhos da Frente II apresentam predominantemente duas abordagens, que são a geração de impressão digital dos dispositivos (*fingerprinting*/pré-processamento) com a extração de características do tráfego e a classificação baseada em aprendizado de máquina (*machine learning*). Como técnica de classificação, nos trabalhos selecionados, observou-se a abordagem supervisada com classes rotuladas como adoção mais comum, presente nos trabalhos #16, #17, #18, #19, #22 e #24.

Nos trabalhos #16 e #17 os autores realizam um *fingerprinting* a partir da extração de 23 características dos cabeçalhos de pacotes de rede do dispositivo IoT. No estudo #18 foi adotada a ferramenta netmate-flowcalc para realizar a extração de 36 características estatísticas do fluxo de rede TCP e UDP para treinar um modelo de aprendizado de máquina supervisionado. O estudo #19 utiliza unicamente o protocolo DHCP para fazer a impressão digital de dispositivos, extraindo 18 características. No estudo #20 é proposto a utilização de métodos ativos e passivos para extrair textos em linguagem natural de pacotes de rede.

O trabalho #21 propôs um sistema chamado AuDI para identificar de maneira autônoma o tipo de dispositivo em uma rede IoT ao analisar o tráfego da rede. Por outro lado, o trabalho #22 realizou o pré-processamento extraindo características (*features*) de oito atributos do TCP/IP. No estudo #24 é realizada a extração de 4 características por pacote em 16 pacotes consecutivos do fluxo de rede. Por fim, no trabalho de revisão #23 os autores citam que outras abordagens ou uma combinação de abordagens tem sido utilizadas para a identificação de dispositivos baseado no seu comportamento através da análise do fluxo de rede: baseadas em regras, baseada em estatísticas, baseadas em conhecimento, baseadas em *machine learning/deep learning* e em séries temporais.

Nenhum trabalho propôs ou definiu uma classificação que represente de forma mais ampla os tipos de dispositivos utilizando nomenclaturas como *smart TVs*, computadores ou *smartphones*. Foi possível observar que em todos os trabalhos baseados em inteligência artificial, os autores tendem a treinar as inteligências artificiais para reconhecer padrões de mesmo fabricante e/ou variações de modelo de um número limitado de classes de dispositivos. Em contrapartida, nos trabalhos #20 e #21 os autores mostram seus resultados de classificação com algum nível de agrupamento. Contudo, esse agrupamento foi realizado como recurso visual, de maneira manual, para apresentação dos resultados (#20) ou para automaticamente criar rótulos de classes sequenciais (#21).

4.4. [QP4] Os dispositivos inteligentes podem tornar uma rede doméstica desprotegida. Considerando o contexto de identificação de dispositivos, o que tem sido feito para tornar as redes nas quais os dispositivos IoT estão inseridos mais seguras?

Da frente I, apenas um trabalho (#9) abordou os dispositivos IoT com alguma consideração sobre a segurança dos usuários, com foco na proteção da privacidade. Por outro lado, trabalhos da Frente II propuseram formas de mitigação a fim de proteger a rede do usuário contra dispositivos inteligentes inseguros. O IoT Sentinel (#16 e #17) é composto por um *Security Gateway* localizado na rede do usuário e um serviço de nuvem. Os dispositivos são classificados como confiáveis, estritos ou restritos e a rede é dividida

em duas zonas isoladas virtualmente: rede confiável e não-confiável. Dependendo da classificação de segurança do dispositivo, ele pode ser impedido de acessar a Internet ou ter sua comunicação limitada com outros dispositivos. A arquitetura proposta em #21 também prevê um modelo de classificação e segurança na nuvem com identificação de dispositivos na borda. Através da Internet, o serviço seria capaz de receber impressões digitais de diversos IoT e estabelecer políticas de restrição de acesso. Os demais trabalhos apenas sugerem que suas propostas de identificação de dispositivos poderiam ser desenvolvidas futuramente no contexto da segurança. Nesse sentido, houve menções para a detecção de intrusão pela identificação de tráfego anormal de dispositivos (#18, #24) e isolamento de dispositivos IoT da rede (#20).

5. Considerações finais

Neste trabalho foi apresentado um mapeamento sistemático com o objetivo de realizar uma revisão da literatura sobre o uso da linguagem natural para controle de acesso de dispositivos em ambientes residenciais, os métodos de identificação automática desses dispositivos na rede e também sobre questões ligadas à segurança de dispositivos IoT em redes residenciais. Foi constatado que a linguagem natural é amplamente utilizada nas redes baseadas em intenção, principalmente utilizada em SDN. O gerenciamento de redes utilizando linguagem natural em alto nível é um campo de pesquisa em crescimento nos últimos anos. No entanto, o interesse dos pesquisadores ainda está muito limitado ao gerenciamento de grandes redes, sendo ainda é pouco explorado na literatura para o ambiente doméstico. No cenário de identificação de tipos de dispositivos as duas técnicas que são utilizadas com mais frequência são o *fingerprinting* baseado na análise do comportamento do tráfego de rede dos dispositivos e a classificação desse comportamento utilizando aprendizado de máquina. Os modelos de classificação de tipos de dispositivos nos estudos analisados tendem a agrupar dispositivos da mesma marca na mesma classe. Para facilitar a configuração de regras em um roteador, uma abordagem válida seria se o modelo de classificação fosse capaz de detectar todos os dispositivos de uma classe, como *smartphones*, permitindo uma abstração para o usuário. Este trabalho buscou integrar duas áreas de pesquisa: a utilização de linguagem natural para configuração de redes domésticas e a identificação de tipos de dispositivos de maneira hierárquica para controle de acesso. Como contribuição, este estudo favorece o desenvolvimento de trabalhos que buscam integrar o uso da linguagem natural no gerenciamento de redes em conjunto com a identificação de dispositivos. Uma proposta para solucionar a lacuna identificada é a criação de uma arquitetura de um sistema baseado em intenção para redes residenciais, capaz de identificar e classificar automaticamente dispositivos por tipo. Essa solução permitiria que usuários sem conhecimentos técnicos realizem o controle de acesso de seus dispositivos usando linguagem natural, criando uma abstração da rede ao usuário.

Referências

Alsudais, A. and Keller, E. (2017). Hey network, can you understand me? In *IEEE Conference on Computer Communications Workshops (WKSHPS)*, pages 193–198.

Ammar, N., Noirie, L., and Tixeuil, S. (2019). Network-protocol-based iot device identification. In *Inter. Conf. on Fog and Mobile Edge Computing (FMEC)*, pages 204–209.

Cesário, H. and et al (2022). Arquitetura para gerenciamento de dispositivos através de assistentes virtuais comandados por voz. In *CoUrb*, pages 1–14. SBC.

do Prado, P. F. and et al (2021). *Mobile Edge Computing for Content Distribution and Mobility Support in Smart Cities*, pages 473–500. Springer International Publishing.

Fiorenza, M. and et al (2021). Representação e aplicação de políticas de segurança em firewalls de redes híbridas. In *SBRC*, pages 490–503. SBC.

Hohum, T. and et al (2021). Scottishfold: Catboost-enabled lightweight autonomous smart home device classification. In *Globecom Workshops*, pages 1–6.

Jacobs, A. S. and et al (2021). Hey, lumi! using natural language for intent-based network management. In *USENIX*, pages 625–639.

Jacobs, A. S., Pfitscher, R. J., Ferreira, R. A., and Granville, L. Z. (2019). Refining network intents for self-driving networks. *SIGCOMM Comp. Comm.*, 48(5):55–63.

Kitchenham, B. A. and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical Report EBSE 2007-001, Keele University and Durham University Joint Report.

Leivadeas, A. and Falkner, M. (2022). A survey on intent based networking. *IEEE Communications Surveys & Tutorials*, pages 1–1.

Liu, X., Holden, B., and Wu, D. (2017). Automated synthesis of access control lists. In *Inter. Conference on Software Security and Assurance (ICSSA)*, pages 104–109.

Marchal, S. and et al (2019). Audi: Toward autonomous iot device-type identification using periodic communication. *IEEE Journal on Sel. Areas in Communications*, 37(6).

Miettinen, M. and et al (2017). Iot sentinel: Automated device-type identification for security enforcement in iot. In *ICDCS*, pages 2177–2184.

Neto, E. P. and et al (2021). Seamless mano of multi-vendor sdn controllers across federated multi-domains. *Computer Networks*, 186:107752.

Nguyen, M.-T.-A., Souihi, S. B., Tran, H.-A., and Souihi, S. (2022). When nlp meets sdn : an application to global internet exchange network. In *ICC*, pages 2972–2977.

Paudel, U. and et al (2021). Context-aware iot device functionality extraction from specifications for ensuring consumer security. In *CNS*, pages 155–163.

Petersen, K., Feldt, R., Mujtaba, S., and Mattsson, M. (2008). Systematic mapping studies in software engineering. BCS Learning & Development.

Pisani, F. and et al (2020). Fog computing on constrained devices: Paving the way for the future iot. *Adv. in Edge Comp.: Massive Parallel Processing and Applications*, 35:22.

Ribeiro, R. H. (2020). A bottom-up approach for extracting network intents. Accepted: 2020-11-21T04:25:11Z.

Ribeiro, R. H. and et al (2022). A deterministic approach for extracting network security intents. *Computer Networks*, 214:109109.

Rivera, S., Fei, Z., and Griffioen, J. (2020). Polanco: Enforcing natural language network policies. In *Inter. Conf. on Comp. Comm. and Networks (ICCCN)*, pages 1–9.

Rodrigues, D. O. and et al (2019). Computação urbana da teoria à prática: Fundamentos, aplicações e desafios. Technical report, Minicurso SBRC.

Saha, B. K. and et al (2022). Intent-based industrial network management using natural language instructions. In *IEEE CONECCT*, pages 1–6.

Salman, O., Chaddad, L., Elhajj, I. H., Chehab, A., and Kayssi, A. (2018). Pushing intelligence to the network edge. In *SDS*, pages 87–92.

Salman, O. and et (2022). A machine learning based framework for IoT device identification and abnormal traffic detection. *Trans. on Emerging Telecommunications Technologies*, 33(3). Pub.: John Wiley & Sons.

Shi, P., Song, Y., Fei, Z., and Griffioen, J. (2021). Checking network security policy violations via natural language questions. In *ICCCN*, pages 1–9.

Sánchez, P. M. S. and et al (2021). A survey on device behavior fingerprinting: Data sources, techniques, application scenarios, and datasets. *IEEE Communications Surveys & Tutorials*, 23(2):1048–1077.

Tian, B. and et al (2019). Safely and automatically updating in-network acl configurations with intent language. SIGCOMM '19, page 214–226. ACM.

Xiao, Y. and et al (2022). Lightweight natural language driven intent translation mechanism for intent based networking. In *ICCCS*, pages 46–51.

Yu, L. and et al (2018). Wdmti: Wireless device manufacturer and type identification using hierarchical dirichlet process. In *IEEE MASS*, pages 19–27.

Zeydan, E. and Turk, Y. (2020). Recent advances in intent-based networking: A survey. In *2020 IEEE 91st Vehicular Technology Conference (VTC2020-Spring)*, pages 1–5.