Equipamentos para *Smart Home*: O que eles querem saber sobre nós?

Thiago Adriano Coleti[1], Omar Ali Mahmoud[2], Victor Hugo Sotti[2]
André Luís Andrade Menolli[1], Marcelo Morandini[3], Renato Balancieri[2]

[1]Universidade Estadual do Norte do Paraná – Bandeirantes – PR – Brasil

[2]Universidade de São Paulo – São Paulo – SP – Brasil.

[3]Universidade Estadual do Paraná – Apucarana – PR – Brasil

thiago.coleti@uenp.edu.br, omarmahmoud3611@gmail.com, vhsotti@gmail.com

menolli@uenp.edu.br, m.morandini@usp.br, renato.balancieri@unespar.edu.br

Abstract. *This paper presents an exploratory research that analyzed the Privacy and Security Policies and the Instruction Manuals of 59 home automation equipment for Smart Home in order to verify which personal data was handled. The analysis was conducted with a quantitative approach followed by a qualitative analysis. The surveys identified the following types of personal data: Identification, Financial, Devices and Location. The results presented greater interest in identification and financial data, although location is also used in some cases. This can help data subjects analyze and decide how their privacy is being affected.*

Resumo. *Este artigo apresenta uma pesquisa exploratória que analisou as Políticas de Privacidade e Segurança e os Manuais de Instruções de 59 equipamentos de automação residencial para Smart Home a fim de verificar quais dados pessoais eram manipulados. A análise foi conduzida com uma abordagem quantitativa seguida de análise qualitativa. Os levantamentos identificaram os seguintes tipos de dados pessoais: Identificação, Financeiro, Dispositivos e Localização. Os resultados apontaram maior interesse por dados de identificação e financeiro, embora a localização também é utilizada em alguns casos. Isso pode ajudar os titulares de dados a analisar e decidir sobre como sua privacidade está sendo afetada.*

1. Introdução

Smart Home é a capacidade de automatizar a execução de equipamentos domésticos/residenciais tais como portões, relógios, equipamentos de limpeza, ar-condicionado, gestão do consumo de energia dentre outros, utilizando recursos computacionais. Nesse contexto, também compreende a personalização da execução das tarefas e da experiência do usuário, de forma que o mesmo possa fazer o uso de recursos que se adéquem às suas necessidades e preferências [Wanzeler et al. 2016].

Entretanto, esses benefícios podem requerer a manipulação (coleta, processamento, compartilhamento etc) de dados pessoais, o que pode levantar questões e/ou preocupações relativas à privacidade, segurança e liberdade dos indivíduos

[Guhr et al. 2020]. Dados pessoais são registros que permitem a identificação de um usuário de software, tais como: dados cadastrais, registros de compras, comandos de voz, histórico de navegação e fotos [Mortier et al. 2016].

O conhecimento sobre a manipulação dos dados é um mecanismo de proteção dos usuários, no entanto, essas informações são, geralmente, pouco ou nada acessíveis, além de utilizarem termos técnicos e judicializados [Mortier et al. 2016]. Além disso, a dificuldade em acessar essas informações faz com que os usuários instalem ou adquiram produtos e serviços sem ter o conhecimento das ações de uso de seus dados, o que pode interferir de forma severa em sua privacidade.

Este artigo apresenta uma pesquisa na qual as Políticas de Privacidade e Segurança (PPS) e os Manuais de Instruções (MI) de um conjunto de dispositivos para *Smart Home* foram analisados a fim de identificar quais dados pessoais eram manipulados por eles. Os dados foram disponibilizados em uma planilha *online* para que serem consultados por usuários e desenvolvedores.

O artigo está estruturado da seguinte forma: a Seção 2 descreve a Fundamentação Teórica que norteou essa pesquisa; a Seção 3 apresenta os Materiais e Métodos utilizados nesse trabalho; a Seção 4 apresenta os Resultados e Discussões e a Seção 5 traz as Considerações Finais.

2. Fundamentação Teórica

Esta seção apresenta a fundamentação teórica desta pesquisa, que contempla assuntos como *Smart Home*, Privacidade e a Manipulação de Dados Pessoais e os Trabalhos Relacionados.

2.1. *Smart Home*

Smart Home é um conceito que descreve a habilidade de controlar, personalizar tarefas e automatizar equipamentos domésticos utilizando recursos computacionais. Esses controles utilizam de eletrônica, algoritmos e mecanismos de comunicação de dados para prover, não somente a automação, mas a capacidade de proporcionar inteligência e capacidade de reconhecimento e adaptação ao contexto [Wanzeler et al. 2016].

A existência de hardware acessível como microcontroladores Arduino e Raspberry aceleraram o acesso e o desenvolvimento de tecnologias para construir ambientes personalizados e inteligentes de Smart Home [Huq et al. 2018][Jang et al. 2019]. Nesse contexto, o uso de técnicas de Inteligência Artificial (IA) para apoiar a implementação desses recursos tornou-se comum [Freitas et al. 2010][Luor et al. 2015].

Nos ambientes de *Smart Home*, a automação pode reconhecer contextos de uso e adaptar-se de forma a conhecer, antecipar e executar preferências e comportamentos dos usuários a fim de fornecer uma experiência de uso personalizada [Freitas et al. 2010][Jang et al. 2019].

As experiências personalizadas e focadas nos usuários ocorrem com o uso massivo de dados pessoais, que são manipulados por algoritmos, técnicas de *Machine Learning*, Mineração de Dados, Internet, Bluetooth, dentre outros recursos de computação e eletrônica [Basarudin et al. 2017]. São exemplos de serviços entregues por ambientes de *Smart Home*:

- Câmeras de vídeo com reconhecimento facial e identificação de pessoas e objetos para fins de segurança [Thabet and Amor 2015];
- Sensores que identificam e aprendem os comportamentos dos usuários tais como horários e preferências, e ajustam a operação do equipamento e o gerenciamento de energia [Assaf et al. 2012];
- Recursos de multimídia que identificam as preferências musicais, filmes e séries dos usuários. Esses equipamentos também podem se conectar com outros na residência e serem controlados por aplicativos de celular [Lamjane and Rojatkar 2018].

Pode-se afirmar que o conceito de Automação Residencial Inteligente (*Smart Home)* passa por forte crescimento e as perspectivas são de continuidade. A pesquisa conduzida por [Singh et al. 2018] destaca as perspectivas em relação a automação residencial para dar suporte à questões como saúde, segurança e bem estar. Há expectativas de crescimento na utilização de ambientes inteligentes, especialmente entre idosos para fins de cuidados de saúde e auxílio em tarefas que podem se tornar complexas. Embora um pequeno percentual (5%) dos participantes revelou uma grande preocupação com a manipulação e compartilhamento de seus dados pessoais, os demais acreditam que os benefícios proporcionados por isso superam os perigos

Na próxima seção são discutidos assuntos relacionados à privacidade e a manipulação de dados pessoais.

2.2. Manipulação de Dados Pessoais

Tornou-se comum na utilização das aplicações, os usuários fornecerem dados pessoais que representam suas características, comportamentos, ações e preferências [Toledo 2020]. Esses dados proporcionam *insights* para empresas oferecerem produtos e serviços dentre outras possibilidades e assim obter vantagem competitiva, financeira e poder de decisão e barganha [Bataineh et al. 2016].

A manipulação de dados pessoais ocorre por meio da coleta dos dados em sensores, redes sociais, *websites* e aplicativos móveis e é praticamente impossível interagir com uma aplicação sem ter um dado pessoal coletado [Maus 2015]. Porém sua utilização indevida ou incorreta pode afetar severamente a privacidade, segurança e liberdade dos mesmos, uma vez que o acesso e/ou o compartilhamento de forma indevida pode comprometer o titular dos dados [Schneier 2015]

A disponibilidade de informações sobre a manipulação dos dados dos usuários é fortemente exigida pelas regulamentações para uso de dados pessoais, uma vez que é um direito dos indivíduos, que tem se preocupado cada vez mais com sua privacidade [Audich et al. 2021] [Toledo 2020]. As Políticas de Privacidade e Segurança (PPS) são amplamente conhecidas e utilizadas para prover informações sobre direitos e deveres dos usuários, assim como informar sobre ações em seus dados pessoais [Efroni et al. 2019]. Porém, elas são pouco acessadas e utilizadas em virtude do volume dos textos e a complexidade dos conteúdos [Zeng et al. 2019].

Assim, a preocupação com a manipulação e compartilhamento de dados tende a exigir de desenvolvedores de equipamentos e aplicações de automação uma atenção especial às necessidades de privacidade dos usuários, em especial, a capacidade de pro-

porcionar conhecimento e autonomia para que os titulares possam analisar e avaliar a manipulação de seus dados [Zeng et al. 2019].

A próxima subseção discute os trabalhos relacionados, utilizados como base de estudos para este artigo.

2.3. Trabalhos Relacionados

A preocupação com a manipulação de dados pessoais tem se tornado uma constante para pesquisadores. Pesquisas nesse sentido são identificadas desde o começo do século XXI, época em que a Internet teve grande expansão e as pessoas aumentaram o interesse em aplicações de software.

O trabalho apresentado por [Earp et al. 2005] discute o impacto dos textos das políticas de privacidade ao informar o usuário das ações sobre seus dados e como tais informações podem aumentar a confiabilidade do usuário em relação ao sistema. Dentre outras questões de pesquisa, os autores buscaram verificar se as políticas de privacidade forneciam informações que os usuários desejavam saber. Foram analisados 24 *websites* e os dados permitiram afirmar que os *websites*, principalmente os norte-americanos, não disponibilizavam PPS adequadas a *Fair Information Practice* (FIP) e que muito ainda deveria ser feito em busca de proporcionar políticas de privacidade mais adequadas para os usuários.

No trabalho de [Subahi and Theodorakopoulos 2018] os autores realizaram dois processos de análise de políticas de privacidade. Para a pesquisa, foram propostos 8 critérios, que deveriam ser aplicados por fabricantes de equipamentos de *Internet of Things* (IoT) em PPS. No primeiro, as PPS de 11 equipamentos foram analisadas de forma manual e verificado o quanto se enquadravam nos critérios estalecidos, que mostrou que a adequação aos critérios propostos era fraca. Os autores também desenvolveram uma aplicação de software, que monitorou pacotes de dados entre os dispositivos de IoT e a nuvem, o que permitiu concluir uma aderência superior a 63% dos equipamentos aos critérios analisados.

Os trabalhos supracitados buscaram analisar políticas de privacidade, considerando também a IoT, classe na qual se enquadra a *Smart Home*. Porém, não foram identificados trabalhos que analisassem um volume relativamente grande de equipamentos com forte apelo comercial e com foco em *Smart Home*.

A próxima seção apresenta os materiais e métodos utilizados neste trabalho.

3. Materiais e Métodos

Considerando o cenário de manipulação de dados pessoais em ambientes de *Smart Home* e a possível interferência na privacidade e segurança dos usuários; e a obrigatoriedade de publicidade das informações de manipulação, requerida pela Lei Geral de Proteção de Dados Pessoais (LGPD), esta pesquisa foi desenvolvida com o objetivo de responder as seguintes questões:

- Q1 - Quais dados pessoais são de maior interesse dos tipos de equipamentos de *Smart Home* analisados?
- Q2 - Quais dados pessoais são de maior interesse das marcas (fabricantes) dos equipamentos analisados (Tabela 1)?

- Q3 - Quais os três fabricantes que manipulam o maior número de dados pessoais?
- Q4 - Quais dados pessoais são pouco utilizados, mas que podem ter potencial futuro?

Para responder as questões apontadas, foram conduzidas as ações descritas nas próximas subseções.

3.1. Seleção de equipamentos residenciais *Smart*

A primeira etapa envolveu as ações de selecionar os dispositivos de *Smart Home*. A seleção dos equipamentos foi aleatória e ocorreu no ano de 2021, com buscas via Google por termos como *"Equipamentos de automação residencial"*. Os critérios para inclusão foram: (1) o equipamento deveria ser comercialmente vendido; e (2) o equipamento deveria permitir interação com o usuário de maneira direta, por meio de aplicativos ou pela coleta de dados via sensores. Foram selecionados 59 equipamentos, os quais não cobrem todos disponíveis no mercado, mas serviram de amostra para esta pesquisa. A lista de equipamentos, classificada por tipo e fabricante é mostrada na Tabela 1.

Tabela 1. Tipos de equipamentos e fabricantes

Tipo	**Qtde**.	**Fabricante**
Assistentes Virtuais	8	Amazon (2), Facebook, Google (2), Intelbras, Positivo e Samsung
Câmeras de vídeo	3	Ekaza, Logiteh and Positivo
Equipamentos Pet	2	PetKit
Fechaduras e Campainhas	5	Elsys, Intelbras(2), Netatmo and Smarteck
Players de áudio e vídeo	4	Apple, LG (2) and Sonos
Sensores	8	Ecobbe (2), Houseeasy, Sensative, Simplehuman, Sonoff (2) and Tuya
Tomadas, Lâmpadas e Conectores	11	Novadigital, Philips, Positivo (2), Ring, Smarteck (2), Sonoff (3) and Tuya
Utensílios domésticos	18	CHEF, Eufy, IRobot, Kohler, LG (4), Philips, Positivo, Rachio, Samsung (2), Sensative, Simplehuman, SmartMi and Tuya (2)

Os equipamentos selecionados pertencem a vinte e oito fabricantes diferentes, sendo que quatorze deles tiveram mais de um equipamento selecionado, conforme mostrado na Tabela 2. Destaca-se que não houve preferência por algum fabricante específico, uma vez que o processo de seleção dos equipamentos foi realizado única e exclusivamente pelos critérios já citados.

A análise dos dados e os resultados que serão apresentados posteriormente limita-se à lista de equipamentos citada e pode sofrer mudanças à medida que a mesma seja alterada.

3.2. Análise das Políticas de Privacidade e Manuais de Instruções

Nessa etapa, foi realizada a leitura das PPS e dos MIs dos equipamentos. Optou-se pela leitura dos dois materiais, pois assumiu-se que as informações sobre a manipulação dos dados pessoais deveriam estar inclusas nesses documentos.

Tabela 2. Quantidade de equipamentos por fabricante

Qt.	**Fabricantes**
01	Apple, CHEF, Ekaza, Elsys Facebook, Houseeasy, Irobot, KHOLER, Logitech, Netatmo, Novadigital, Rachio, Ring, Sonos.
02	Amazon, Ecobee, Google, PetKit, Philipps, Sensative, Simplehuman.
03	Intelbras, Samsung, Smarteck.
04	Tuya.
05	Positivo, Sonoff.
06	LG.

Esses materiais consistem, basicamente, em textos, o que levou à identificação de conteúdos objetivos e claros, mas também foram identificados conteúdos subjetivos, que deixaram dúvidas em relação a quais dados pessoais são manipulados, tais como: *"Também podemos coletar informações sobre seu computador"*; ou *"Alguns dados são coletados para melhorar a experiência do usuário"*

Assim, a interpretação e a inferência do conteúdo por parte dos pesquisadores foi necessária para decidir sobre seu contexto de manipulação. Ocorreu, também, a observação de dados com semelhança semântica, os quais foram classificados[1] em virtude da informação final apresentada. Por exemplo: o dado pessoal **Endereço** foi classificado como dado de identificação, pois permite identificar o usuário em uma aplicação de software. Já o dado pessoal **Localização** foi classificado como Dado de Localização, pois refere-se às coordenadas geográficas do usuário em determinado contexto de uso e é comumente utilizado para auxiliar o usuário em ações específicas, além de ter ampla variabilidade.

No total foram identificados 30 dados pessoais apontados como manipuláveis por equipamentos de *Smart Home*, que foram classificados nos seguintes grupos:

- **Dados de Identificação** - foram identificados 19 dados que permitem identificar o usuário, suas ações, costumes e preferências. Os dados foram: nome, endereço, e-mail, país, apelido, número de telefone, dados da empresa que trabalha ou estuda, cargo, dados comportamentais, número de documentos, data de nascimento, gênero, dados do telefone, dados biométricos, imagens, foto, áudios, idioma, estado civil, e identificadores específicos (ID Apple, por exemplo);
- **Dados Financeiros** - foram relacionados 03 dados que permitem identificar comportamentos financeiros de um indivíduo, tais como: dados do cartão de crédito, dados da compra (produto, serviço, valor, parcelamento) e dados de cobrança;
- **Dados do Dispositivo** - totalizaram 05 dados que permitem a identificação do dispositivo e de suas características de utilização, sendo eles: dados do dispositivo, dados de conexão, dados de uso, dados de desempenho e histórico do navegador;
- **Dados de Localização** - dados que permitem obter a localização do usuário ao utilizar um equipamento. Esse grupo compreendeu 03 dados: posição geográfica (latitude e longitude), informações da localização e dados operacionais do sistema.

A próxima seção apresenta a análise e a discussão realizada sobre os apontamentos de manipulação de dados pessoais.

[1]Classificação realizada com base no conhecimento empírico dos pesquisadores em relação ao objetivo final da manipulação.

4. Análise e Discussão

A análise e a discussão foram conduzidas considerando a indicação da utilização de um dado pessoal por um equipamento. Não foram realizadas análises detalhadas para verificar se são utilizados com maior frequência ou de alguma forma específica, mas somente a indicação da utilização. Também não foi considerado o esforço para encontrar a indicação de uso, questão essa a ser tratada em trabalhos futuros.

Para responder a **Q1**, assumiu-se como um dado pessoal muito utilizado aquele indicado por, pelo menos, 42 equipamentos analisados (70%). Esse percentual foi escolhido pelos pesquisadores, de forma aleatória considerando que um dado seria manipulado por ampla maioria dos equipamentos. Os dados pessoais que se enquadraram nesse critério são mostrados na Tabela 3.

Tabela 3. Dados pessoais com mais indicações de uso

Grupo	Dado Pessoal	Qtde.
Identificação	Nome	51
	E-mail	51
	Número de Telefone	47
	País	44
Dispositivo	Dados da conexão	43
	Dados do dispositivo	44

A ampla utilização dos dados **Nome** e **E-mail** não foi considerada como surpresa, uma vez que é o mínimo necessário para qualquer pessoa efetuar um registro em um produto ou serviço digital. Dos oito equipamentos que não indicaram a utilização dos campos citados, três deles não foram encontradas indicações sobre a manipulação de qualquer dado pessoal[2]. Já para os demais, há indicação de utilização de outros dados que eventualmente poderiam substituir os dados apontados. Por exemplo, ao invés do dado *Nome*, poderia ser considerado o *Apelido*.

Já com relação ao dado pessoal **E-mail**, os equipamentos que não indicaram a manipulação dos mesmos aparentam não demandar pelo dado, como é o caso de lâmpadas de led e de controles infravermelho. Assumiu-se também, que os equipamentos estão conectados a aplicativos que já fazem a coleta do e-mail. Este cenário também foi considerado para os campos *País* e *Número do Telefone*.

Em relação aos *Dados do Dispositivo*, assumiu-se que sua manipulação é necessária para a conexão dos equipamentos com aplicativos de controle ou outros equipamentos. Já os grupos Localização e Financeiro não apresentaram dados com indicação de uso por mais de 43 equipamentos.

Em relação aos dados que não tiveram indicação mínima de 70%, para o grupo de **Identificação**, o dado pessoal, o Endereço teve indicação de uso por 35 equipamentos e os demais dados tiveram indicação igual ou inferir à 16 equipamentos. Para o grupo de **Dados Financeiros**, os 03 dados pessoais supracitados foram indicados por 24, 11 e 16 equipamentos. Quanto ao grupo de **Dados do Dispositivo**, o histórico de navegador foi apontado por 32 equipamentos, já os dados de desempenho e de uso por 13 e 12 equipamentos, respectivamente. Os **Dados de Localização** também não tiveram indicação de

[2]A discussão sobre equipamentos que não indicaram dados pessoais será feita posteriormente.

uso por mais de 70% dos dispositivos, pois os dados de localização e a informação da localização foram apontados por 26 e 28 equipamentos, respectivamente e o dados operacionais, por 03 equipamentos. Portanto, assumiu-se que esses dados são de interesse dos dispositivos de **Smart Home**, mas são manipulados para atender demandas específicas, o que justifica a baixa quantidade de equipamentos que os indicaram.

Para responder a **Q2**, foram exploradas as indicações de manipulação de dados pessoais organizados por tipo de equipamento (Tabela 1), o que permitiu inferir que:

- Em relação ao grupo *Dados de Identificação*, todos os tipos de equipamentos apresentaram indicações de manipulação semelhante à analise realizada para responder a Q1, com exceção de 01 equipamento do grupo *Fechaduras e Campainhas* que indicou a manipulação de dados incomuns como foto, imagem e áudio;
- Já o grupo *Dados Financeiros* têm dados manipulados por grupos que têm interação direta com o usuário, como os *Virtual Assistants*, *Players* de Áudio e Video e Equipamentos domésticos. Assumiu-se que isso ocorre em virtude desses equipamentos intermediarem a aquisição de produtos e serviços pagos para os usuários. Porém, dispositivos como sensores, conectores e tomadas também apontaram o uso de dados financeiros, o que levanta questões sobre o interesse por esses dados, dada as características desses dispositivos;
- No grupo do *Dados do Dispositivo*, 89% dos equipamentos apontaram para a manipulação de pelo menos um dado desse grupo. Já 06 equipamentos apontaram para a manipulação de todos os dados de dispositivos, sendo eles de duas categorias: (1) *Tomadas, lâmpadas, conectores*; e (2) *Sensores*;
- Por fim, no grupo *Dados de Localização* a predominância é pela manipulação dos dados de localização e informações da localização. É bastante variada a utilização pelos tipos de equipamentos, porém estes estão usualmente relacionados a equipamentos com funcionalidades de movimentação como câmeras, *Virtual Assistants*, e Equipamento doméstico. Entretanto, há equipamentos de grupos como sensores, tomadas, lâmpadas e conectores, que também indicaram a manipulação de dados de localização. Destaca-se que, somente 3 equipamentos manipulam dados operacionais.

Assim, a Q2 pode ser respondida considerando que, para os dados pessoais apontados na Tabela 3, o perfil de manipulação é mantido quando analisado por tipo de equipamento. Já os demais dados têm uma variação mais significativa, uma vez que não há um padrão consistente de dados indicados, mesmo para tipos semelhantes de equipamentos. Assumiu-se que deve-se às funcionalidades dos equipamentos, e, também, ao fato de que há equipamentos que complementam outros (puro aspecto comercial). Pode-se afirmar, também, que o fato de poucos dados serem apontados por mais de 70% dos equipamentos pode estar fortemente relacionada a falta de um padrão e ao interesse específico de um equipamento por um dado pessoal. Também pode-se considerar que as informações desejadas podem estar disponíveis em documentos ou recursos não analisados pelos pesquisadores.

Para responder a **Q3** foi calculada a média aritmética da quantidade de dados manipulados pelos equipamentos de cada fabricante. O resultado é mostrado na Tabela 4.

Com base nos resultados da Tabela 4, os fabricantes Intelbras, Samsung e Logitech foram aqueles que apontaram o maior número de dados pessoais. Os três fabrican-

Tabela 4. Quantidade média de dados utilizados por fabricante.

Fabricante/Média	Fabricante/Média	Fabricante/Média
Amazon: 12	Apple: 13	CHEF: 6
Ecobee: 6	Ekaza: 13	Elsys: 11
Eufy: 9	Facebook: 10	Google: 11
Houseeasy: 11	Intelbras: 15	Irobot: 9
KOHLER: 10	LG: 12	Logitech: 14
Netatmo: 5	Novadigital: 12	PETKIT: 10
Philips: 5	Positivo: 13	Rachio: 13
Ring: 8	Samsumg: 14	Sensative: 9
Simplehuman: 8	Smarteck: 0	SmartMi: 9
Sonoff: 11	Sonos: 7	Tuya: 10

tes manipulam dados em todos os grupos estudados nesse artigo, com diferenças sutis na manipulação de dados específicos em cada grupo. Os produtos desses fabricantes estão nos grupos de *Fechaduras e campainhas*, *Players de áudio e vídeo* e *Utensílios domésticos*.

Já as fabricantes que indicaram o menor número de dados manipulados foram a Netatmo com 02 dados de *Identificação*, 02 de *Dispositivo* e 01 de *Localização*; e a Phillips, com indicações de manipulação de 05 dados, sendo a Netatmo Philips com 04 dados na categoria *Identificação* e 01 dado *Financeiro*.

Para responder a **Q4** os pesquisadores avaliaram dados com pouca indicação (abaixo de 20% dos equipamentos) de uso a fim de propor cenários de como eles poderiam ser utilizados em benefício da empresa controladora e/ou usuário, assim como poderia ser prejudicial para os usuários, caso manipulado de maneira incorreta e poderiam ser de grande potencial para manipulação e obtenção de informações. Três dados pessoais foram selecionados pelos pesquisadores com base em seu conhecimento prévio em pesquisas em Ciência de Dados e Interação Humano-Dados. Os dados selecionados foram:

- **Foto**: é um dado pessoal que permite a identificação direta do usuário. Sua utilização seria benéfica considerando eventos como: controle de acesso seguro; identificação de pessoas desaparecidas; identificação de reações faciais para avaliação de produtos ou exames de saúde. Porém, a manipulação desse dado poderia ser extremamente invasiva, uma vez que permitira localizar e/ou identificar a pessoas em vários locais ou atitudes de sua vida privada, sem seu conhecimento e consentimento;
- **Dados de compra**: Esses dados não criam uma situação de grande risco para os titulares, mas possibilidades de incômodos, uma vez que empresas podem utilizar informações de compra para oferecer outros produtos, identificar preferências dos clientes, traçar e prever intensões de compra, dentre outros aspectos comerciais. Embora as ações citadas sejam comuns em *websites* de *e-commerce*, no contexto dos equipamentos analisados nessa pesquisa, esse dado foi pouco apontado;
- **Dados de uso**: com foco na utilização do equipamento de automação, esse dado permite identificar comportamentos do usuário e assim personalizar a experiência de uso do mesmo. Entretanto, seu compartilhamento inadequado pode comprometer a privacidade do usuário, pois outros equipamentos, fornecedores e serviços

teriam informações sobre a vida da pessoa dentro da casa.

Os dados analisados mostraram que os equipamentos de *Smart Home* estudados, embora tenham grande capacidade e possibilidades de manipulação de dados, apontaram para uma quantidade relativamente pequena de dados pessoais.

Os dados de identificação, como nome, e-mail, país e telefone foram os mais apontados, seguidos por dados do dispositivo e dados de conexão. Esses dados são muito apontados, pois tratam-se do mínimo necessário para um usuário realizar seu registro, configurar e utilizar o equipamento.

Embora os equipamentos de *Smart Home* tenham ampla capacidade de oferecer experiências diferenciadas aos seus usuários, assumiu-se que os dados apontados são simples e nem sempre permitiriam o aprendizado, a previsão e a execução de experiências diferenciadas. Esperava-se, por exemplo, a manipulação de certos tipos de dados considerados comuns como os áudios, fotos, números de documentos e dados de cartão, entretanto, o mesmos foram pouco apontados. Para esses dados, as seguintes inferências foram feitas pelos pesquisadores para justificar o baixo número de apontamentos:

- A manipulação dos demais dados apontados são feitas de maneira muito específica em cada equipamento, embora diversos dados pessoais tenham sido apontados, poucos são efetivamente manipulados pelos equipamentos;
- As informações sobre quais dados são manipulados não foram identificados e/ou estão em outros documentos; ou mesmo o fato dos fabricantes não estarem totalmente ajustados às políticas de transparência no uso dos dados pessoais das regulamentações como a LGPD e GDPR, o que pode criar dificuldades para um usuário identificar e analisar como seus dados são manipulados;
- A manipulação ocorre de maneira terceirizada, em aplicativos de celulares, *tablets* e *websites*. Nesses casos, os equipamentos funcionam como atuadores para ligar e desligar componentes eletrônicos. Esse cenário é bastante cogitado por pesquisadores, uma vez que aplicativos instalados nos celulares dos usuários apresentam uma capacidade maior de coleta e manipulação dos dados, além de uma eventual combinação com serviços em nuvem;
- A quarta possibilidade seria o fato de que, mesmo com poucos dados pessoais apontados, a combinação de dois ou mais dados, em situações específicas também podem produzir informações relevantes sobre os indivíduos. Para esse caso, assumiu-se que algoritmos poderiam realizar combinações e associações entre os dados de forma a produzir as informações desejadas com o mínimo de dados coletados.

As planilhas com os dados utilizados nesta pesquisa estão disponíveis no endereço `bit.ly/3oDHGC3`, para ser consultado por desenvolvedores e usuários.

5. Considerações Finais

Esse artigo apresentou um levantamento e análise dos dados pessoais manipulados por equipamentos de *Smart Home*. Os equipamentos de *Smart Home* demandam por manipulação dos dados pessoais para execução de suas funcionalidades e os dados levantados na pesquisa mostraram que os mesmos têm se utilizado de diversos deles.

Esse trabalho teve foco em identificar os dados e não em analisar sua possível manipulação. Foi possível constatar que os equipamentos manipulam um conjunto consi-

derável de dados pessoais, em especial, dados de identificação. Entretanto, se manipulados de maneira adequada, podem melhorar significativamente a experiência dos usuários, mas o uso inadequado pode ser muito invasivo para a privacidade das pessoas, pois os dados refletem comportamentos domésticos.

Com as informações obtidas nesta pesquisa, espera-se poder apoiar usuários na identificação de quais dados pessoais são manipulados por equipamentos de automação residencial. Os dados referem-se aos equipamentos analisados e podem variar caso outros equipamentos sejam estudados, mas pode-se considerar uma amostra relevante para análise.

Destaca-se que, a tendência de manipulação de dados por equipamentos de *Smart Home* tende a crescer e assim, a necessidade de métodos e técnicas para garantir a privacidade do usuário e a transparência da manipulação é um desafio futuro dentro da área da Computação, uma vez que a manipulação compreende uma forte relação entre pessoas, ambientes e negócios.

Como trabalhos futuros, pretende-se ampliar as informações do catálogo de maneira a melhorar a transparência das informações para usuários dos equipamentos, assim como prover métodos e técnicas para melhorar a implementação da transparência no contexto de *Smart Home*. Pretende-se, também, realizar ações para verificar como são realizadas manipulações dos dados pelos equipamentos.

6. Agradecimentos

Agradecemos a Fundação Araucária do Estado do Paraná e a Universidade Estadual do Paraná (UNESPAR) pelo apoio financeiro a essa pesquisa.

Referências

Assaf, M. H., Mootoo, R., Das, S. R., Petriu, E. M., Groza, V., and Biswas, S. (2012). Sensor based home automation and security system. *IEEE Instrumentation and Measurement Tecnology Conference*, (April 2019):722–727.

Audich, D. A., Dara, R., and Nonnecke, B. (2021). Improving Readability of Online Privacy Policies through DOOP: A Domain Ontology for Online Privacy. *Digital*, 1(4):198–215.

Basarudin, N. A., Yeon, A. L., Yusoff, Z. M., Dahlan, N. H. M., and Author, N. M. (2017). Smart home users' information in cloud system: A comparison between Malaysian personal data protection act 2010 and EU general data protection regulation. *Malaysian Construction Research Journal*, 2(2):209–222.

Bataineh, A. S., Mizouni, R., El Barachi, M., and Bentahar, J. (2016). Monetizing Personal Data: A Two-Sided Market Approach. *Procedia Computer Science*, 83(June):472–479.

Earp, J. B., Antón, A. I., Aiman-Smith, L., and Stufflebeam, W. H. (2005). Examining Internet privacy policies within the context of user privacy values. *IEEE Transactions on Engineering Management*, 52(2):227–237.

Efroni, Z., Metzger, J., Mischau, L., and Schirmbeck, M. (2019). Privacy icons: A risk-based approach to visualisation of data processing. *European Data Protection Law Review*, 5(3):352–366.

Freitas, C. C. S., Mesquita, B. D. R., Pereira, C. E., and Farias, V. J. C. (2010). Automação Residencial – Uma Abordagem Em Relação As Atuais Tecnologias E Perspectivas Para O Futuro. *V Congresso Norte-Nordeste de Pesquisa e Inovação (CONNEPI)*, (1):8.

Guhr, N., Werth, O., Blacha, P. P. H., and Breitner, M. H. (2020). Privacy concerns in the smart home context. *SN Applied Sciences*, 2(2):1–12.

Huq, S. M., Rahman, M. A., and Saleh, S. M. (2018). Application for integrating microcontrollers to Internet of Things. *20th International Conference of Computer and Information Technology, ICCIT 2017*, 2018-Janua(December):1–4.

Jang, I., Lee, D., Choi, J., and Son, Y. (2019). An approach to share self-taught knowledge between home IoT devices at the edge. *Sensors (Switzerland)*, 19(4).

Lamjane, K. S. and Rojatkar, D. V. (2018). Amazon Alexa Based Home Automation Using Particle Photon. *International Journal of Scientific Research in Science, Engineering and Technology IJSRSET*, 4(May):80–84.

Luor, T., Lu, H. P., Yu, H., and Lu, Y. (2015). Exploring the critical quality attributes and models of smart homes. *Maturitas*, 82(4):377–386.

Maus, G. (2015). Decoding, hacking, and optimizing societies: Exploring potential applications of human data analytics in sociological engineering, both internally and as offensive weapons. *Proceedings of the 2015 Science and Information Conference, SAI 2015*, pages 538–547.

Mortier, R., Haddadi, H., Henderson, T., Mcauley, D., Crowcroft, J., and Crabtree, A. (2016). Human-Data Interaction: The Encyclopedia of Human-Computer Interaction. *The Encyclopedia of Human-Computer Interaction*, pages 1–48.

Schneier, B. (2015). *Data and Goliath. The hidden battles to collect your data and control your world*. Norton, New York.

Singh, D., Psychoula, I., Kropf, J., Hanke, S., and Holzinger, A. (2018). Users' perceptions and attitudes towards smart home technologies. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10898 LNCS(August):203–214.

Subahi, A. and Theodorakopoulos, G. (2018). Ensuring Compliance of IoT Devices with Their Privacy Policy Agreement. *Proceedings - 2018 IEEE 6th International Conference on Future Internet of Things and Cloud, FiCloud 2018*, pages 100–107.

Thabet, A. B. and Amor, N. B. (2015). Enhanced smart doorbell system based on face recognition. In *2015 16th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA)*, pages 373–377.

Toledo, M. D. E. (2020). Lei Geral de Proteção de Dados. um guia completo.

Wanzeler, T., Fülber, H., and Merlin, B. (2016). Desenvolvimento de um sistema de automação residencial de baixo custo aliado ao conceito de Internet das Coisas (IoT). *Anais do XXXIV Simpósio Brasileiro de Telecomunicações*, pages 40–44.

Zeng, E., Mare, S., and Roesner, F. (2019). End user security & privacy concerns with smart homes. *Proceedings of the 13th Symposium on Usable Privacy and Security, SOUPS 2017*, pages 65–80.