Optimising user security recommendations for AI-powered smart-homes

Emma Scott, Sakshyam Panda, George Loukas, Emmanouil Panaousis

Internet of Things and Security Centre
University of Greenwich, London, UK
{e.scott, s.panda, g.loukas, e.panaousis}@greenwich.ac.uk

***Abstract*—Research in the context of user awareness has shown that smart-home occupants often lack cybersecurity awareness even when it comes to frequently used technologies such as online social networks and email. To cope with the risks, smart-homes must be equipped with adequate cybersecurity measures besides the *knowledge* and *time* required by smart-home occupants to implement security measures. In this paper, we explore potential threats in AI-powered smart-homes and identify a list of cybersecurity controls required to mitigate their potential impact considering attack vectors, as well as the time and knowledge required to implement a control. We use optimisation to identify the best set of controls to minimise the risk exposure considering these metrics. Our comparative analysis against a random selection approach highlight that our approach is at least 25% better at minimising risk. Finally, we show how improved knowledge or time impacts the risk.**

***Index Terms*—cybersecurity, smart-home, threats, control optimisation, risk assessment.**

I. Introduction

Artificial intelligence (AI)-based technologies are actively used in smart-homes to provide intelligent services and recommendations. Internet of Things (IoT) coupled with AI concepts has been applied to the home environment to make it safer, smarter and more convenient. The most popular application is the smart speakers. Besides, AI-driven logic and concepts has been employed to design Chatbots [1], email services [2], smart vacuum cleaners [3] and washing machine [4]. Furthermore, AI is used to manage energy usage in smart-homes [5]. Despite proving useful, these devices and applications introduce new vulnerabilities leading to cyber-physical risks.

smart-home devices have many constraints, including computational power, storage limitation, energy and hardware capabilities which complicate the implementation of traditional security solutions [6]. Besides, implementing security mechanisms to protect occupants from cyberattacks is challenging due to the heterogeneity of smart-homes and interaction between devices with different working mechanisms [7]–[9].

Our work mainly deals with identifying pre-emptive security controls to protect occupants of smart-homes from cybersecurity attacks. We first present the applications of AI in smart-homes, then discuss the most common cybersecurity threats and identify a list of cybersecurity controls to mitigate these threats. This research creates a repository of security controls using data from literature and technical reports. The controls are further classified considering the attack vectors as well as the time and knowledge required to implement a control. Once the controls are identified, we develop an optimisation model based on a cost-benefit analysis of the controls to compute the set of controls which effectively minimise risks in smart-homes. Our results highlight the impact of these metrics on the cybersecurity posture of a smart-home eventually impacting the privacy and security of the occupants.

This paper is organised as follows. Section II reviews and presents related work and application of AI in a domestic environment. The most common threats in smart-homes are presented in Section III. Security controls to protect occupants from cyberattacks and the optimisation model are presented in Section IV. The suggested recommendations and the numerical results are presented in Section V. Finally, Section VI concludes this paper.

II. AI-enabled domestic Applications

This section presents currently available domestic AI devices and applications deemed relatively common during 2021.

1) Smart Speakers and Voice Assistants: Smart speakers currently dominate the market for AI for home applications. Amazon captures approximately 70% of the US smart speaker market share[1]. The Google Nest family is the other set of smart speakers that function similar to the Amazon Echos which utilises speech recognition. Other examples of speaking assistive devices include Microsoft Cortana, Ivee Sleek, Jibo, Athom Homey, Apple HomePod and Josh Micro [10]. Engagement of such devices with users and access to their environment makes them an attractive target for malicious entities [11], [12].

2) Home Energy Management Systems (HEMS): AI has only recently been used to provide services for HEMS. EON Electric[2] and Romatech[3] are pushing forward with plans to improve energy management services to their customers.

[1] https://www.theverge.com/2020/2/10/21131988/amazon-alexa-echo-google-apple-smart-home-speaker.

[2] http://www.eonelectric.com/.

[3] https://www.romatech.co.uk/.

3) Medical IoT: The application of AI to enable care at home from a distance has been in progress since 2012. To assist elderly people, Essence[4] developed a package of services that use AI to alert caregivers of any unusual activity. It uses activity recognition to distinguish between normal and abnormal activities of users to identify adverse situations such as falls.

4) Smart-Home Managers: Smart-home management was originally a form of consumer-focused device management, providing a platform to monitor and manage other consumer IoT devices including thermostats, lights, security systems, and appliances. Veego[5] developed smart-home managers for internet service providers (ISP) rather than homes. They use machine learning along with a global database of devices that use Veego products to provide tailored services.

5) Chatbots: Chatbots can be rule-based to provide a relatively restrictive set of assistance or use AI to generate more flexible interactions and responses. Chatbots are increasingly being introduced into social-housing scenarios to provide information and assistive services. Users can access the chatbot functionalities through SMS, a specific website or social media platforms such as Facebook. They are popular for providing out of hours support, aiding with filling in forms and sending user reminders. Although chatbots can be incredibly useful, they provide a vast attack surface for adversaries due to their unregulated nature.

6) Smart Thermostats: Smart thermostats use machine learning techniques to learn from the changes occupants make, sense when occupants are in the home and away, and create a heating schedule that suits the occupants best [13]. For example, the Nest Learning Thermostat and Hive use AI to learn how best to heat and cool homes.

7) Smart-Home Physical Security: Smart-home physical security includes a range of products and presents a real opportunity for adversaries. AI is used in many ways to support physical home security such as the use of facial recognition to identify people through installed video cameras (e.g Google Nest Cam). Another application of AI in security systems is the use of spoken interaction with the system. AI can also be used for data analysis to detect unusual patterns of behaviour and intrinsic relationships [10]. August Smart Lock + Connect[6] and Google Nest Protect provide such features. Smart locks which use a combination of sensors and AI to continually monitor also play a key role in physical security systems.

8) Smart Vacuums: Smart vacuums are growing in popularity and their efficacy has increased significantly in recent years. Although some only use sensors in vacuum cleaners, iRobot[7] uses AI to support voice interactions with users. The Bosch Roxxter[8] also uses AI to provide interactive maps of the room.

III. Threats

The acclimatisation of IoT technologies, AI and cloud computing with advanced sensing and actuation capabilities has led to more convenient smart-homes but also has attracted significant adversarial attention. Securing and preventing threats requires identifying vulnerabilities and implementing appropriate controls to mitigate or eliminate the effects of a threat. Heartfield et al. [14] classified cyber threats considering the attack vectors and potential impact on occupants and their domestic life. From a system security perspective, [15] highlighted threat scenarios involving interaction between entities in smart-home and smart grid environments and evaluated their impact on overall system security. A categorisation of security threats for smart-home appliances with attack surfaces and vulnerabilities is presented in [16]. Honeypot-based solutions have been proposed to collect threat intelligence and support security posture of IoT-enabled environments [17]–[19].

Based on the list of smart-home applications discussed and the threats to smart-homes identified in reports such as [9][10], we recognise the common cyber threats to smart-homes are: (i) Physical Intrusion, (ii) Credit Card Data Theft, (iii) Ransomware; and (iv) Eavesdropping. The scenarios highlighted in this section are considered for the evaluation of our model and the outcome of the discussion presented in the later sections of this paper.

1) Physical Intrusion: Physical intrusion is a major threat to home users. Smart locks might be a convenient solution to the traditional approach, but they increase the attack surface of intrusion. Attackers can potentially unlock the door via a Wi-Fi connected device, allowing entry to the home without having to “break into” it. Nevertheless, an attack against one of the AI applications, we discussed in the previous section, can take place if the home user owns both a smart lock and a smart speaker which is also connected to the smart lock. Smart speakers can be set up so that they unlock a smart lock when the user commands them to and correctly identify a PIN. This PIN could be as short as four digits, and thus, if the user has little insight into the risks that it poses, they may choose something obvious such as 1234. The most recent published research into PIN setting was carried out by[11], a considerable time ago, however, their findings were revealing. 1234 accounted for 10.7% of all PINs, and 1111 along with 0000 accounted for another 9.9% in total. Many attacks against smart speakers (e.g., [20],) have been identified in the literature that can be applied to this step of this broader intrusion attack. These attacks can be grouped under the title of unauthorised voice commands, however, their methods vary within the group.

[21], [22] present attacks that work similarly, transmitting a voice command through an unauthorised means, however, their

[4] https://www.essencesmartcare.com/.

[5] https://veego.io/.

[6] https://august.com/.

[7] https://www.irobot.co.uk/.

[8] https://www.bosch.com/stories/smart-robot-vacuum/.

[9] https://www.forbes.com/sites/forbestechcouncil/2021/11/09/top-five-cybersecurity-threats-and-how-to-avoid-them/

[10] https://www.verizon.com/business/resources/reports/dbir/2021/masters-guide/

[11] https://www.theguardian.com/money/blog/2012/sep/28/debit-cards-currentaccount.

commands are distorted so that if somebody is nearby, they will not know what the transmitted command said. Another attack that fits into this group of unauthorised attacks, requiring access to a Bluetooth speaker is called the "DolphinAttack" [23]. Again, it first requires the adversary to record or produce an audio command, but they then go on to modulate this audio on ultrasonic carriers, before using the audio in the same way discussed above.

2) Credit Card Fraud: The next threat that we consider is credit card fraud, which can be achieved by an attack aiming at domestic AI applications related to chatbots and smart speakers both included in our list of domestic AI applications. If credit card information is processed appropriately, it may end up visible in system logs and be disclosed to external services that the user does not necessarily trust [24]. An adversary can develop a malicious chatbot or smart speaker application. These applications must imitate one that the user already uses, or must offer a service that the user thinks would be helpful, therefore encouraging them to engage with it. The purpose of the application must also at some point require the user to input their bank card details, for example, if the user is paying the chatbot for a service that they are providing or buying something through the smart speaker application. Next, the user must discover the malicious chatbot or download the smart speaker application, without realising that it is malicious. Finally, after interacting with the chatbot or the application, they would ask the user for their bank card details and the user would be fooled into giving them. More details on attacks against chatbots can be found here [25].

When the vulnerable AI application is a smart speaker, two attack methods have been identified by two separate teams, who labelled it "Skill squatting" [11] and "voice squatting" [26]. They utilise the fact that Alexa only correctly interprets words an average of 68.9% of the time [11] and that many of these inaccurate interpretations are predictable. When choosing the malicious smart speaker application to develop, the adversary must first identify a known application with a vulnerable name (the "squattable skill"). They can then develop a malicious skill with a name similar enough to the squattable skill that could be misidentified as the intent by the Smart Speaker. It must be similar to the squattable skill, fooling the user into thinking they are using the skill that they desired, and must then carry out the functions that the adversary desires, such as gaining the credit card details of the home user.

3) Ransomware: Ransomware is a growing threat that has become very common over the past years, with SonicWall Capture Labs recording 121.2 million attacks in just the first half of 2020[12], an increase of 20%. Although ransomware attacks have previously been aimed at organisations, the Covid-19 pandemic has meant more people are working from home, using the same devices that they complete other activities on, simply connecting through a virtual network. Ransomware attack targets are not solely organisations, however, and all home users are vulnerable.

For example, a ransomware attack can target a Chatbot. The first step of the attack is for the adversary to develop a malicious Chatbot that imitates one that the user already uses or offers a related service that the user is likely to engage with. Then, the malicious Chatbot must either acquire the user's email address and perform a social engineering attack, e.g., send an email infected with a link that will install the ransomware. upon successful exploitation, the adversary will have control of their device, and it will result in the user being made to pay a ransom to get back access to their device or data on their device.

4) Eavesdropping: Much effort has been put into testing Smart Vacuum's vulnerabilities and attempting to hack into them[13]. They are seemingly secure devices, however, [27] developed a method to reverse engineer the ARM Cortex-M based firmware of the robot, and thus install malicious firmware, allowing an adversary to gain root access through the Dustcloud software stack. A proof of concept attack, named LidarPhone [28] uses these vulnerabilities to go on to override the usual connection to the Xiaomi cloud ecosystem with the Valetudo software stack on the rooted device, therefore controlling the robot over a local network.

In the proof of concept attack discussed by [28], the adversary targets private and potentially sensitive information from speech that is emitted by the computer speakers as the victim engages in teleconferencing. This is particularly pertinent with the increased frequency of working from home on the Corona Virus pandemic.

IV. Security Controls

Risk modification is the process of reducing the risk by applying appropriate controls. Controls come with costs, both financial and indirect, and thus applying all available controls to all known risks may not be the best course of action. This paper focuses on identifying controls that minimise residual risk using optimisation. This requires the identification of a repository of controls to be used against known risks. The end goal of this is to provide the most suitable, or optimal, controls to mitigate the risks. This section details the process of selection for our final set of controls suitable for smart-homes from three sources of controls: CHI, CIS and Cyber Essentials used in the literature.

A. Control Repository for Smart-home

As expressed, "cyber hygiene" is the descriptive name for a group of cybersecurity controls appropriate for home use. Several reputable sources were identified, and each of these sources offers an important and unique contribution. Due to the specific and novel nature of the field being investigated in this paper, the identified controls are combined to provide coverage against threats in smart-homes. We begin by identifying a list of cybersecurity controls that could be used against threats in

[12] https://www.itpro.co.uk/security/ransomware/356567/1212-million-ransomware-attacks-in-the-first-half-of-2020.

[13] https://www.kaspersky.com/blog/xiaomi-mi-robot-hacked/20632/.

a smart-home. A control could be implemented at different levels (e.g., high, medium, low) with each having a degree of efficacy against a threat. A higher implementation level fetches higher costs compared to the lower levels. Groupings of varying implementation levels of controls were considered to identify the set of controls that optimally mitigated the risk. Table I presents a list of CHI controls, initially identified in [29], that could be implemented against threats identified in this paper. We further group some of the controls (e.g, Secure Password Behaviour) as multiple CHI controls could be associated with a control group. However, the controls grouped into one are considered to be of the same implementation level requiring similar time and knowledge costs and having similar efficacy ratings.

CHI Measures	**Application to Identified Threats**
Checking the quality of SSL certificates when doing online financial transaction	Yes
Enabling firewalls on your computing devices	Yes
Running virus scan on any new USB or external storage devices	No
Checking an incoming email's header	Yes
Checking a sender's email domain name	Yes
Checking to see if email requests have grammatical or typographical errors	Yes
Monitoring different processes such as CPU, power, or network usage on your device	Yes
Changing default username from administrator to something unique on all Internet enabled devices	Yes
Changing default passwords on all internet enabled devices	Yes
Keeping virus protection updated	Yes
Managing how your browser stores passwords	No
Creating new/unique logins and passwords for all your online sign-ins	Yes
Storing logins and passwords on encrypted online password vaults	Yes
Placing online alerts from your name or personal information	No
Assessing the authenticity of social media friend/information requests	No
Knowing who you are connected to on social media	No
Reassessing social media friends/connections	No
Ensuring the location information is not leaked in posts	Yes

TABLE I: CHI Security Controls

Vishwanath et al. [29] presented a conceptual model for cyber hygiene, empirically identified its sub-dimensions and have developed an inventory of 18 cybersecurity controls that could be applied to a range of people with different levels of cybersecurity needs. Such et al. [30] studied basic cyber hygiene and its efficacy for Small and Medium Enterprises (SMEs) and identified cybersecurity controls that are cheaper to implement, yet are effective against remotely exploitable commodity-level vulnerabilities. Such et al. discovered that 69.3% of the vulnerabilities explored were fully mitigated by applying the controls identified to meet UK Government's Cyber Essentials Scheme[14], 29.2% were partially mitigated, and only 1.5% were not mitigated at all. Therefore, it is evident that Cyber Essentials controls are fairly effective against commodity attacks and thus must be considered while developing cybersecurity guidelines.

Besides the CHI controls, the Cyber Essentials Scheme also provides a list of security controls. This list includes additional controls such as "Malware Protection", "Patch management" and "Secure Configuration". These controls are essential to protect against many of the identified threats and are not included in the CHI controls list. Table II presents the Cyber Essentials controls.

Cyber Essentials controls	**Application to Identified Threats**
Firewalls and Gateways	Yes
Secure Configurations	Yes
User Access Control	Yes
Malware Protection	Yes
Patch Management	Yes

TABLE II: Cyber Essentials Controls

Similarly, controls were identified from the first six control family of CIS Controls that apply to a domestic environment. For example, ensuring that software is still supported by the vendor is deemed to be a control with a higher level of implementation as it required higher knowledge and time. Table III presents the CIS controls applicable to a domestic environment. These CIS controls are carefully selected as most of the CIS controls are applicable to businesses with broader network architecture than smart-homes. The CIS has a different set of controls to manage hardware and software assets, however as the range of software and hardware in a smart-home is significantly lower than that of an organisation, they have been grouped as one in this paper. Maintaining an inventory of assets, and addressing unauthorised assets can be logically seen as similar activities where the efficacy to mitigate threat is a combined result of both the behaviour. Ensuring that software is still supported by the vendor is considered to be a control that requires an increased level of knowledge and time, demanding additional effort from the user. Thus, it has a higher value for knowledge and time than other controls of the same group.

The next control group, "Continuous vulnerability management" provides two controls linked to deploying automated patch management tools. As Cyber Essentials also name Patch Management to be one of their five controls, these are put in the same group. However, these controls are specifically aimed at organisations. Automated patch management tools, both for software and operating systems, are not available to all home users and require an additional degree of knowledge. Thus, the control "Update software and operating system regularly" has been added within the same group of Patch Management to support home users. CIS Control 4.2 is linked closely to the "Secure Password Behaviour" group from the CHI controls which are covered by different levels within the group, and thus will not be added separately. Lastly, as with Cyber Essential's controls, CIS specify establishing secure

[14] https://www.ncsc.gov.uk/cyberessentials/overview.

CIS Controls	**Applicable to Identified Threats**
Inventory and control of hardware assets	
1.4 Maintain detailed asset inventory	Yes
1.6 Address unauthorised assets	Yes
Inventory and control of software assets	
2.1 Maintain inventory of software assets	Yes
2.2 Ensure software is supported by vendor	Yes
2.6 Address unapproved software	Yes
Continuous vulnerability management	
3.4 Deployed automated os patch management tool	Yes
3.5 Deploy automated software patch management tool	Yes
Control of administrative privileges	
4.2 Change default password	Yes
4.3 Ensure the use of dedicated administrative accounts	No
Secure configuration of hardware and software	
5.1 Establish secure configurations	Yes
Maintenance, monitoring and analysis of audit logs	
6.2 Activate audit logging	No

TABLE III: CIS controls applicable to domestic environment

configurations. Again, this has already been added to the final list, so can be dismissed.

Finally, Table IV presents a final set of 14 controls to address threats in a smart-home. Seven of these controls are grouped into three groups of varying levels of implementation. The table also presents the cost, knowledge and time required and the flow reduction value considered in this paper. More detail on these values is covered in the following section.

Cyber Hygiene Measures	**Source**	**Time**	**Knowledge**	**Flow Reduction**
Check the quality of the SSL certificate when doing online financial transactions	CHI	1	2	0.01
Enable firewalls on your computing devices	CHI/CE	1	2	0.9
Secure email and messaging behaviour	CHI	1	2	0.5
Monitor different processes such as CPU, power, or network usage on your device	CHI	3	3	0.2
Secure Password Behaviour	CHI			
1. Change default passwords on all internet enabled devices	CHI	1	1	0.6
2. Create unique log ins and passwords for all online sign ins	CHI	2	1	0.5
3. Storing logins and passwords on encrypted online password vaults	CHI	3	3	0.4
Keep virus and malware protection updated	CHI/CE	2	2	0.5
Ensuring location is not leaked on social media	CHI	1	1	0.9
Establish Secure Configurations	CE/CIS	2	3	0.0001
Patch Management	CE			
1. Update software and operating systems regularly	CIS	1	1	0.8
2. Deploy automated OS and software patch management system	CIS	1	3	0.7
Monitor and manage software and hardware in your home	CIS			
1. Maintain an inventory, checking for any unrecognised software or hardware, removing them	CIS	2	1	0.6
2. Check software is still supported by vendor	CIS	2	2	0.5

TABLE IV: Cyber Hygiene Repository for smart-home

B. The Optimisation Model

Due to the cost and benefit associated with the implementation of security control and the availability of a budget that could be spent, the selection of controls can be formulated as an optimisation problem. The aim is to identify the best set of controls that optimally mitigate the risk and is within an investment budget. We associate the direct control cost to be the time required to implement a control and the indirect cost to be the knowledge required to implement, practice and maintain this control. The costs (knowledge and time) are further categorised into three levels: Low, Medium and High represented as $\{1, 2, 3\}$, respectively. The categorical values are speculated based on the understanding of the effort required to implement a control where 1 represents the least effort.

Similar to [31], we also consider assigning each control a flow reduction value which expresses how effective control is against an attack. We define as "edge flow" the likelihood of success of an attack for that step given the implementation of security control. This likelihood is calculated using the edge flow value of the preceding step and the efficacy of the implemented control against the attack. Similarly, the "default flow value" is defined as the likelihood of success of an attack when no security controls are implemented in that step. The default flow value is selected to be either 1 or 0.5. If the attack step is likely to be successful under any circumstances then we set the default edge value to 1, otherwise is set to 0.5. For example, a deauthentication attack in comparison to the KRACK attack will always be successful if the adversary has gained access to the network leading to a default flow value of 1. KRACK attack, on the other hand, relies on a router using the WPA-2 protocol which might be unlikely for every case leading to the default flow value of 0.5.

We identified security controls for all seven attack scenarios (presented in section III) considering all possible combinations of levels for knowledge and time. These controls are used to determine the likelihood of success of an attack. Figure 1 illustrates how this likelihood is determined for the credit card data theft through a malicious software attack scenario considering low knowledge and low time.

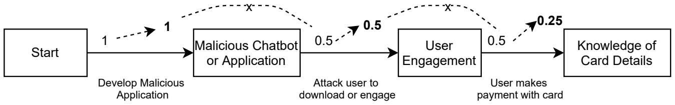
Start
1
1
x
Malicious Chatbot or Application
0.5
0.5
x
User Engagement
0.5
0.25
Knowledge of Card Details
Develop Malicious Application
Attack user to download or engage
User makes payment with card

Fig. 1: Low Knowledge and Low Time

When low knowledge and low time is considered, there were no controls applicable to any of the attack steps leading to an edge flow value of $1 \times 0.5 \times 0.5 = 0.25$. Raising the time by one level (i.e., low knowledge and medium time), "Maintain an inventory, checking for any unrecognised software or hardware, removing them" which has a time score of 2 could apply to the second step of the attack graph. This control has an efficacy of 0.6 against malicious software attacks and thus resulting to an edge flow of $1 \times (0.5 \times 0.6) \times 0.5 = 0.15$. Implementation of this control reduces the success probability of the attack by 0.1. Similarly, more advanced controls based on the level of knowledge and time (e.g., "Keep virus and malware protection updated" control which requires medium knowledge and medium time) could be implemented to further influence the success probability of an attack.

Once the likelihood of success of an attack, the probable loss from an attack and the cybersecurity controls that could be implemented to mitigate the expected loss have been identified,

	Low Knowledge			**Medium Knowledge**			**High Knowledge**		
	Low **Time** (LL)	**Medium** **Time** (LM)	**High** **Time** (LH)	**Low** **Time** (ML)	**Medium** **Time** (MM)	**High** **Time** (MH)	**Low** **Time** (HL)	**Medium** **Time** (HM)	**High** **Time** (HH)
Intrusion via smart lock	0.16	0.144	0.11025	0.16	0.114	0.000007875	0.16	0.114	0.00000126
Intrusion via smart lock and smart speaker	0.216	0.216	0.189	0.18	0.09	0.07875	0.18	0.09	0.063
Credit card data theft through network	0.08	0.0008	0.0006125	0.08	0.0008	0.00000004375	0.08	0.0008	0.00000000875
Credit card data theft through malicious software	0.25	0.00125	0.00125	0.15	0.0003125	0.0003125	0.15	0.0003125	0.0003125
Ransomware attack	0.2	0.045	0.039375	0.2	0.0225	0.019688	0.2	0.0225	0.019688
Eavesdropping attack through smart vacuum	0.24	0.24	0.21	0.2	0.1	0.0875	0.2	0.1	0.014
Eavesdropping attack through smart speaker	0.2	0.2	0.2	0.12	0.1	0.1	0.12	0.1	0.02

TABLE V: Likelihood of success of an attack

the expected risk for each attack method $i = \{1, 2, \cdots, N\}$ can be defined as

$$R_i = A_i \times S_i \times L_i, \tag{1}$$

where A_i is the likelihood of being attacked, S_i is the probability of success of the attack and L_i is the probable loss [32]. A_i and L_i vary based on the type of attack, rather than the attack vector or device and the risk is associated with the owner of the devices rather than all occupants of the smart-home. Using a multi-factor optimisation, we aim to identify the optimal set of controls that minimise the expected risk within the investment budget. A similar approach has been used in [31], however, it is seen that this approach does not minimise the expected risk. For example, when a lower budget is selected their method recommended controls fitting the budget regardless of the required level of knowledge and time to implement it. Such a recommendation would not be appropriate for occupants with limited cybersecurity knowledge. To overcome this challenge, we adapted our optimisation to consider only those combinations of time and knowledge costs that are within the budget as suggested in [33]–[35].

We define the overall risk improvement as the sum of residual risk for all attack methods and express it as:

$$\sum_{i=1}^{N} R_i - R_i \times \text{Efficacy of control against attack} \tag{2}$$

Table V presents the likelihood of success of various attack scenarios considered in our paper.

V. Evaluation

In this section, we develop and evaluate the proposed model using Python and execute 10,000 simulation runs to generate the results. Our implementation identifies the optimal cybersecurity controls to reduce the overall risk of the smart-home.

1) Intrusion: In 2017, the Office for Nation Statistics reported that two in every 100 households were victims of burglary each year[15]. This number was nine in every 100 houses in 1995. Assuming this reduction is linear, we consider one in every 100 houses to be a victim of burglary for the year 2021. We selected a borough in the UK of 47.35 square km in size, which has a rate of 32 burglaries per square km leading to 1514 burglaries per year. In 2019, on average 2.54 people occupied each dwelling[16], and approximately 290,000 people are living in this borough, indicating that there are about 112,671 dwellings. This results in 1.3% of dwellings on average being burgled each year. This is for intrusion in general rather than intrusion through the smart lock in particular. Thus, we label all attacks under the intrusion category with an A_i of 0.013. The average loss of a burglary L_i is deemed to be £3,030[17].

2) Credit Card Fraud: In 2019, 21% of UK adults have cancelled and replaced their credit cards due to attempted fraud and in 51% of these cases money was stolen resulting in a theft of £846 on average[18]. Considering this, we set $A_i = 0.21 \times 0.51 = 0.1071$ and $L_i = £846$.

3) Ransomware: In 2020, the Ransomware attack rose by 80% in the third quarter costing the UK around £365 million[19]. A majority of home Ransomware attacks are aimed at organisations rather than home users. So, we consider a relatively low $A_i = 0.001$. While the most common demanded ransom is $L_i = £10{,}000$ in the UK[20].

4) Eavesdropping: Many proof-of-concept attacks through smart vacuum cleaners have been discussed in [28]. As there is no evidence that such attacks have happened, we consider a low $A_i = 0.001$. [26] discussed Eavesdropping attacks through the smart speaker but then no actual cases have been recorded yet leading to $A_i = 0.001$. The L_i of both eavesdropping attacks is considered £100 due to a slim chance that credit card information may be spoken out loud.

Considering these values, we calculate over 10,000 simulation runs (i) the average cyber risk using our optimisation model; and (ii) the average cyber risk considering the same number of controls chosen at random. The benefit demonstrated and evaluated here is specifically a reduction in risk between using our model and the user randomly choosing the same number of cybersecurity controls from a predefined list

[15] https://www.ons.gov.uk/peoplepopulationandcommunity/crimeandjustice/.

[16] https://www.london.gov.uk/sites/default/files/housinginlondon2019.pdf.

[17] ttps://www.theecoexperts.co.uk/home-security.

[18] https://www.comparethemarket.com/credit-cards/content/latest-credit-card-fraud-scam-figures/

[19] https://www.itpro.co.uk/security/ransomware/357353/uk-ransomware-attacks-increased-by-80-in-past-quarter

[20] https://www.itpro.co.uk/security/ransomware/359364/cost-of-ransomware-doubles-in-a-year.

	LL	LM	LH	ML	MM	MH	HL	HM	HH
Random	59.9469	52.7739	52.5738	38.9987	24.5598	24.5335	32.0107	10.6613	4.9119
Our Model	44.9548	34.4641	34.4641	18.1987	9.3606	5.2719	12.0369	3.1690	2.5330
Improvement	25.00%	34.69%	34.44%	53.33%	61.89%	78.51%	62.40%	70.28%	48.43%

TABLE VI: Cyber risk comparison using our model vs random cybersecurity controls selection.

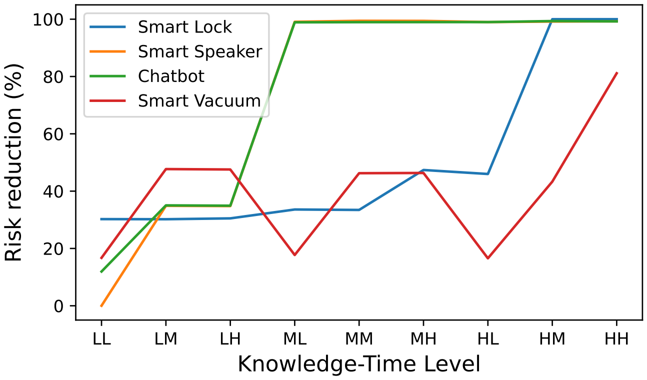

Fig. 2: Cyber risk improvement for a single AI-powered device.

and implementing the ones that they can. Table VI presents the cyber risk for various combinations of knowledge and time with our model and with the random selection of cybersecurity controls. The results also present the percentage improvement in risk reduction using our model.

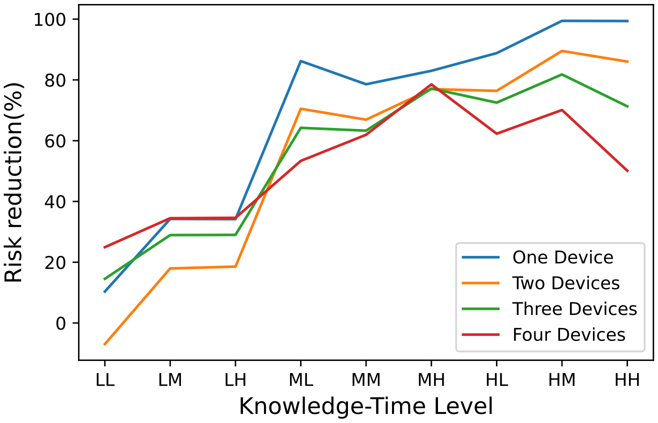

Fig. 3: Cyber risk improvement for different number of AI-powered devices.

Next, we calculate the risk associated with each device in a smart-home. Figure 2 presents the device-specific risk improvement considering our model against a random selection of the same number of controls. A similar comparison considering multiple devices is presented in Figure 3. In these figures, the x-axis ticks represent a similar combination of knowledge and time used in Table VI. It is evident from the results that the risk to occupants increases with an increase in several smart devices in the home. Such insights on the level of risk could assist cyber insurers inappropriately profiling home occupants and designing custom insurance policy packages [36]–[38] with premium discounts [39].

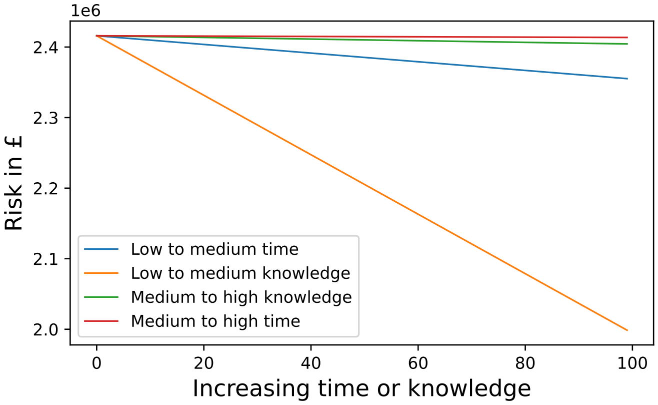

Fig. 4: Risk vs Improvement in level of time or Knowledge.

To understand the impact of knowledge and time required to implement controls on a user group, we extend our analysis by considering the budget required to improve a level of time or knowledge i.e, from low to medium and medium to a high level. Figure 4 presents the risk for these transitions. The results illustrate that the most effective use of resources would be to focus on improving the cybersecurity knowledge of occupants as seen in the transition from low to medium level of knowledge.

VI. Conclusion

A security challenge for smart-home occupants is to identify mechanisms for preventing attacks and their potential impact. We have highlighted potential threats in an AI-powered smart-home and identified a list of cybersecurity controls to mitigate these threats, considering the attack vectors, as well as the time and knowledge required to implement a control. Taking these metrics into consideration, we have developed an optimisation model to identify controls that optimally reduce the risk exposure. In doing so, we have also highlighted the impact of time and knowledge required to implement control on the risk and consequently impact on the occupants of a smart-home. A direction for continued research is to consider attack graphs for specific AI applications (e.g. Alexa, smart radiator valve) and services and optimise user security recommendations. A richer optimisation model with fine-grained quantification of key parameters could be developed to obtain better insights into cybersecurity investments. Future research could also help in establishing guidelines for secure smart-home environments and formalising security practices based on abnormal behaviour that are particularly critical to ensure a safer and resilient home.

References

[1] J. Cahn, "Chatbot: Architecture, design, & development," *University of Pennsylvania School of Engineering and Applied Science Department of Computer and Information Science*, 2017.

[2] M. Dredze, H. M. Wallach, D. Puller, T. Brooks, J. Carroll, J. Magarick, J. Blitzer, F. Pereira *et al.*, "Intelligent email: Aiding users with ai." in *AAAI*, 2008, pp. 1524–1527.

[3] S. Kwon, J. Kim, and K. R. Ryu, "Performance comparison of situation-aware models for activating robot vacuum cleaner in a smart home," *International Journal of Computer and Information Engineering*, vol. 10, no. 2, pp. 312–316, 2016.

[4] C. B. Kobus, E. A. Klaassen, R. Mugge, and J. P. Schoormans, "A real-life assessment on the effect of smart appliances for shifting households' electricity demand," *Applied Energy*, vol. 147, pp. 335–343, 2015.

[5] W. S. Lima, E. Souto, T. Rocha, R. W. Pazzi, and F. Pramudianto, "User activity recognition for energy saving in smart home environment," in *2015 IEEE Symposium on Computers and Communication (ISCC)*. IEEE, 2015, pp. 751–757.

[6] K. Sha, W. Wei, T. A. Yang, Z. Wang, and W. Shi, "On security challenges and open issues in internet of things," *Future Generation Computer Systems*, vol. 83, pp. 326–337, 2018.

[7] G. Mantas, D. Lymberopoulos, and N. Komninos, "Security in smart home environment," in *Wireless Technologies for Ambient Assisted Living and Healthcare: Systems and Applications*. IGI global, 2011, pp. 170–191.

[8] U. Saxena, J. Sodhi, and Y. Singh, "Analysis of security attacks in a smart home networks," in *2017 7th International Conference on Cloud Computing, Data Science & Engineering-Confluence*. IEEE, 2017, pp. 431–436.

[9] R. Heartfield, G. Loukas, A. Bezemskij, and E. Panaousis, "Self-configurable cyber-physical intrusion detection for smart homes using reinforcement learning," *IEEE Transactions on Information Forensics and Security*, vol. 16, pp. 1720–1735, 2020.

[10] X. Guo, Z. Shen, Y. Zhang, and T. Wu, "Review on the application of artificial intelligence in smart homes," *Smart Cities*, vol. 2, no. 3, pp. 402–420, 2019.

[11] D. Kumar, R. Paccagnella, P. Murley, E. Hennenfent, J. Mason, A. Bates, and M. Bailey, "Skill squatting attacks on amazon alexa," in *27th {USENIX} Security Symposium ({USENIX} Security 18)*, 2018, pp. 33–47.

[12] Y. Chen, X. Yuan, J. Zhang, Y. Zhao, S. Zhang, K. Chen, and X. Wang, "Devil's whisper: A general approach for physical adversarial attacks against commercial black-box speech recognition devices," in *29th {USENIX} Security Symposium ({USENIX} Security 20)*, 2020, pp. 2667–2684.

[13] G. Hernandez, O. Arias, D. Buentello, and Y. Jin, "Smart nest thermostat: A smart spy in your home," *Black Hat USA*, no. 2015, 2014.

[14] R. Heartfield, G. Loukas, S. Budimir, A. Bezemskij, J. R. Fontaine, A. Filippoupolitis, and E. Roesch, "A taxonomy of cyber-physical threats and impact in the smart home," *Computers & Security*, vol. 78, pp. 398–428, 2018.

[15] N. Komninos, E. Philippou, and A. Pitsillides, "Survey in smart grid and smart home security: Issues, challenges and countermeasures," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 1933–1954, 2014.

[16] A. Gai, S. Azam, B. Shanmugam, M. Jonkman, and F. De Boer, "Categorisation of security threats for smart home appliances," in *2018 International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2018, pp. 1–5.

[17] A. Tambe, Y. L. Aung, R. Sridharan, M. Ochoa, N. O. Tippenhauer, A. Shabtai, and Y. Elovici, "Detection of threats to iot devices using scalable vpn-forwarded honeypots," in *Proceedings of the Ninth ACM Conference on Data and Application Security and Privacy*, 2019, pp. 85–96.

[19] S. Panda, S. Rass, S. Moschoyiannis, K. Liang, G. Loukas, and E. Panaousis, "Honeycar: A framework to configure honeypotvulnerabilities on the internet of vehicles," *arXiv preprint arXiv:2111.02364*, 2021.

[18] N. Boumkheld, S. Panda, S. Rass, and E. Panaousis, "Honeypot type selection games for smart grid networks," in *International Conference on Decision and Game Theory for Security*. Springer, 2019, pp. 85–96.

[20] X. Lei, G.-H. Tu, A. X. Liu, C.-Y. Li, and T. Xie, "The insecurity of home digital voice assistants-vulnerabilities, attacks and countermeasures," in *2018 IEEE Conference on Communications and Network Security (CNS)*. IEEE, 2018, pp. 1–9.

[21] N. Carlini, P. Mishra, T. Vaidya, Y. Zhang, M. Sherr, C. Shields, D. Wagner, and W. Zhou, "Hidden voice commands," in *25th {USENIX} Security Symposium ({USENIX} Security 16)*, 2016, pp. 513–530.

[22] T. Vaidya, Y. Zhang, M. Sherr, and C. Shields, "Cocaine noodles: exploiting the gap between human and machine speech recognition," in *9th {USENIX} Workshop on Offensive Technologies ({WOOT} 15)*, 2015.

[23] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang, and W. Xu, "Dolphinattack: Inaudible voice commands," in *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, 2017, pp. 103–117.

[24] G. Baudart, J. Dolby, E. Duesterwald, M. Hirzel, and A. Shinnar, "Protecting chatbots from toxic content," in *Proceedings of the 2018 ACM SIGPLAN International Symposium on New Ideas, New Paradigms, and Reflections on Programming and Software*, 2018, pp. 99–110.

[25] W. Ye and Q. Li, "Chatbot security and privacy in the age of personal assistants," in *2020 IEEE/ACM Symposium on Edge Computing (SEC)*. IEEE, 2020, pp. 388–393.

[26] N. Zhang, X. Mi, X. Feng, X. Wang, Y. Tian, and F. Qian, "Dangerous skills: Understanding and mitigating security risks of voice-controlled third-party functions on virtual personal assistant systems," in *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 2019, pp. 1381–1396.

[27] D. Giese, "Having fun with iot: reverse engineering and hacking of xiaomi iot devices," 2018.

[28] S. Sami, Y. Dai, S. R. X. Tan, N. Roy, and J. Han, "Spying with your robot vacuum cleaner: eavesdropping via lidar sensors," in *Proceedings of the 18th Conference on Embedded Networked Sensor Systems*, 2020, pp. 354–367.

[29] A. Vishwanath, L. S. Neo, P. Goh, S. Lee, M. Khader, G. Ong, and J. Chin, "Cyber hygiene: The concept, its measure, and its initial tests," *Decision Support Systems*, vol. 128, p. 113160, 2020.

[30] J. M. Such, P. Ciholas, A. Rashid, J. Vidler, and T. Seabrook, "Basic cyber hygiene: Does it work?" *Computer*, vol. 52, no. 4, pp. 21–31, 2019.

[31] M. Khouzani, Z. Liu, and P. Malacaria, "Scalable min-max multi-objective cyber-security optimisation over probabilistic attack graphs," *European Journal of Operational Research*, vol. 278, no. 3, pp. 894–903, 2019.

[32] M. E. Whitman and H. J. Mattord, *Principles of information security*. Cengage Learning, 2011.

[33] S. Panda, E. Panaousis, G. Loukas, and C. Laoudias, "Optimizing investments in cyber hygiene for protecting healthcare users," in *From Lambda Calculus to Cybersecurity Through Program Analysis*. Springer, 2020, pp. 268–291.

[34] A. Fielder, E. Panaousis, P. Malacaria, C. Hankin, and F. Smeraldi, "Decision support approaches for cyber security investment," *Decision support systems*, vol. 86, pp. 13–23, 2016.

[35] G. Gonzalez-Granadillo, S. A. Menesidou, D. Papamartzivanos, R. Romeu, D. Navarro-Llobet, C. Okoh, S. Nifakos, C. Xenakis, and E. Panaousis, "Automated cyber and privacy risk management toolkit," *Sensors*, vol. 21, no. 16, p. 5493, 2021.

[36] A. Marotta, F. Martinelli, S. Nanni, A. Orlando, and A. Yautsiukhin, "Cyber-insurance survey," *Computer Science Review*, vol. 24, pp. 35–61, 2017.

[37] R. Zhang and Q. Zhu, "FlipIn: A game-theoretic cyber insurance framework for incentive-compatible cyber risk management of internet of things," *IEEE Transactions on Information Forensics and Security*, vol. 15, pp. 2026–2041, 2019.

[38] A. Farao, S. Panda, S. A. Menesidou, E. Veliou *et al.*, "Secondo: A platform for cybersecurity investments and cyber insurance decisions," in *International Conference on Trust and Privacy in Digital Business*. Springer, 2020, pp. 65–74.

[39] S. Panda, D. W. Woods, A. Laszka, A. Fielder, and E. Panaousis, "Post-incident audits on cyber insurance discounts," *Computers & Security*, vol. 87, p. 101593, 2019.