Securing Consumer IoT in the Smart Home: Architecture, Challenges, and Countermeasures

Yan Meng, Wei Zhang, Haojin Zhu, and Xuemin (Sherman) Shen

Abstract

The consumer Internet of Things (IoT) platforms are gaining high popularity. However, due to the open nature of wireless communications, smart home platforms are facing many new challenges, especially in the aspect of security and privacy. In this article, we first introduce the architecture of current popular smart home platforms and elaborate the functions of each component. Then we discuss the security and privacy challenges arising from these platforms and review the state of the art of the proposed countermeasures. We give a comprehensive survey on several new attacks on the voice interface of smart home platforms, which aim to gain unauthorized access and execute over-privileged behaviors to compromise the user's privacy. To thwart these attacks, we propose a novel voice liveness detection system, which analyzes the wireless signals generated by IoT devices and the received voice samples to perform user authentication. We implement a real-world testbed on Samsung's SmartThings platform to evaluate the performance of the proposed system, and demonstrate its effectiveness.

Introduction

Internet of Things (IoT) systems are playing a vital role in the emerging smart home environment. According to the latest research report, the smart home market is expected to be valued at US$137.91 billion by 2023, growing at a compound annual growth rate (CAGR) of 13 percent between 2017 and 2023. Different from the traditional IoT architecture where devices have only wired connections, this new IoT paradigm allows smart devices (e.g., lighting control, access control, home healthcare, smart kitchen, and home appliances) to connect with each other via wireless communications. To foster inter-vendor compatibility and encourage community-based software development ecosystems, some major players in the market have developed a few smart home platforms to encourage manufacturers to produce compatible devices and software developers to develop applications with a uniform abstraction of smart devices. Prominent examples of these platforms include Samsung SmartThings, Apple HomeKit, Google Home, Amazon Echo, and AllSeen AllJoyn.

In these IoT systems, a wide range of wireless protocols including WiFi, ZigBee, Z-Wave, and NB-IoT are supported, which allow different smart devices to communicate with each other as well as a local gateway (e.g., a hub or a base station). For example, the Google Home platform can work with over 1500 smart devices from more than 200 brands. Further, the smart devices in the platform can be remotely controlled by applications running in the cloud back-end. This feature reduces the hardware complexity and energy consumption of smart devices. It also enables third-party developers to develop customized applications to achieve their special requirements, which stimulates a large number of developers to participate in the consumer IoT ecosystems.

Along with the popularity of consumer IoT is increasing concern on security and privacy breaches in smart homes. The consumer IoT typically comprises a wide range of resource-constrained smart devices, which are particularly vulnerable to a series of passive/active attacks. First, since most intra-network data are delivered over wireless channels, the transmissions between smart devices and the hub can be interfered easily by a jamming attack. Further, some IoT devices do not employ encryption in order to save power consumption, and the communication patterns between the hub and the smart devices show limited variety: it only sends a few different types of messages to indicate specific events and expects a limited number of commands from the hub (or the cloud). These make it easy to be eavesdropped and analyzed, which may potentially compromise a user's privacy. From the system and mobile application security perspective, the attacker may lure a user to install malicious applications on his/her platform, which can be leveraged to perform over-privileged executions to steal the user's privacy information. In this article, we give a comprehensive survey on the potential security challenges in consumer IoT as well as the state of the art in countermeasure proposals.

In addition to the above physical/network layer and mobile device security challenges, the security issues on the voice control interface of the smart home platforms are gaining increasing attention. Voice control is one of the most important user interfaces in IoT platforms (e.g., Amazon Echo, Google Home). However, recent studies show that attackers can forge unauthorized voice commands to take over the IoT platform or violate the user's privacy by forging a voice command that is even inaudible to the human. To thwart these attacks, we propose a novel two-factor authentication framework to validate the

Yan Meng, Wei Zhang, and Haojin Zhu are with Shanghai Jiao Tong University; Xuemin (Sherman) Shen is with the University of Waterloo.

Digital Object Identifier: 10.1109/MWC.2017.1800100

Smart devices are the key building blocks of the entire infrastructure. They are manufactured to provide diversified functionalities for human-centered environment. Different from traditional electronic devices, smart devices support network communication via different wireless protocols such as Bluetooth, Wi-Fi, ZigBee, Z-Wave and NB-IoT, which could carry forward the advantages of IoT.

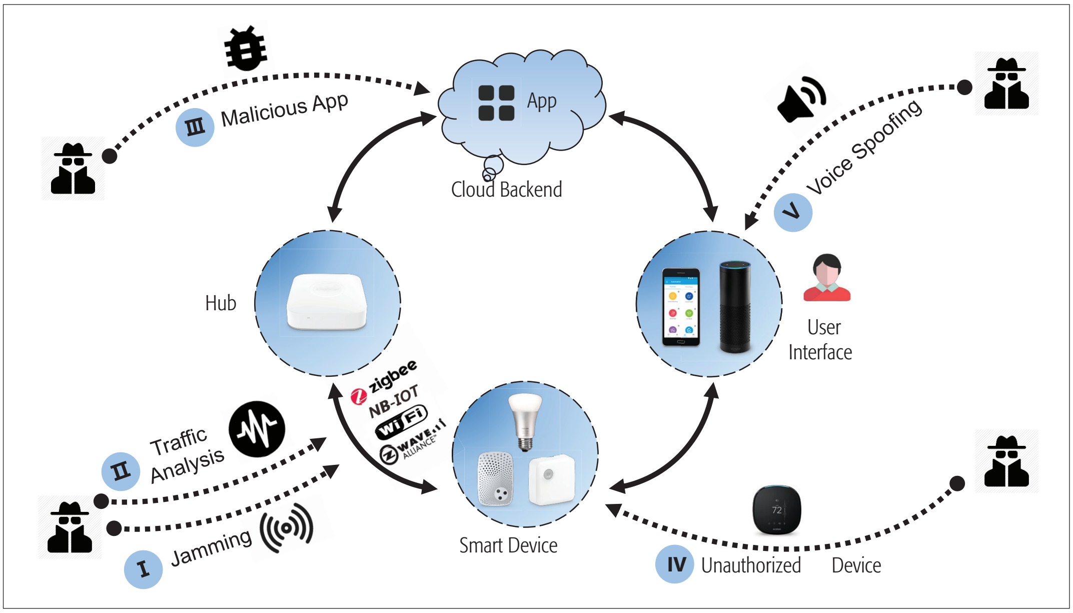


FIGURE 1. Architecture and emerging challenges of consumer IoT.

veracity of the voice command. This framework exploits the persuasive wireless signals generated from the existing IoT devices to perform authentication. We design and implement a testbed on Samsung's SmartThings platform to evaluate the proposed framework. The experimental results demonstrate its effectiveness well.

The remainder of this article is organized as follows. We first describe the architecture of some popular smart home platforms. Then we introduce the security and privacy challenges of smart home platforms and the existing solutions for securing these platforms. Furthermore, to thwart new attacks on the voice interface of smart homes, we propose a novel framework and evaluate its performance. Finally, we conclude the whole article by pointing out some future works.

Architecture of a Smart Home Platform

In this section, we present a typical network architecture of smart home IoT systems. As shown in Fig. 1, there are four kinds of major components in smart home IoT: *smart device*, *hub*, *cloud back-end*, and *user interface*.

Smart Device: Smart devices are the key building blocks of the entire infrastructure. They are manufactured to provide diversified functionalities for the human-centered environment. Different from traditional electronic devices, smart devices support network communication via different wireless protocols such as Bluetooth, Wi-Fi, ZigBee, Z-Wave, and NB-IoT, which could carry forward the advantages of IoT. To exhibit context awareness of the environment, it is fairly common for these devices to sense or change the environment. The *sensor* is a type of smart device that senses changes in the environment (e.g., temperature or motion). It collects information and delivers it to other devices through a network for further processing. Another type of smart device, the *actuator*, is designed for changing the environment. It is usually backed by a cloud for computation. After the computation is performed, the final command is delivered to the actuators, which then change their state and the environment (e.g., unlocking a door).

Hub: In a consumer IoT environment, the hub acts as the brain of the wireless network. A hub usually supports a variety of communication standards because heterogeneous devices are connected to it. Also, it serves as the gateway to the back-end cloud, which holds the aggregation logic for various devices. Many hubs are in the shape of home routers and placed in proximity to smart devices. In general, the hub receives the wireless packets from the sensors, sends the information to the cloud, and controls other connected devices according to the message returned by the cloud.

Cloud Back-End: In a smart home ecosystem, the cloud back-end takes charge of most of the computational tasks. Specifically, the cloud back-end is responsible for performing the control logic of the various devices in order to reduce the power consumption of the resource-constrained smart devices. For example, in Samsung SmartThings, users could install customized *SmartApps* in the cloud for device automation. These SmartApps are capable of manipulating multiple devices through the hub by following fixed rules (e.g., turning a light on when someone is coming). On the other hand, the developers in a community could build various SmartApps to make these devices more intelligent.

User Interface: The user interface serves as the interface between the user and smart devices. By installing an application from the app market, users can control devices via the interface. Recently, the voice user interface (VUI) has become the most popular interface due to its great convenience. By removing the need for operating by hand, consumers can easily control the devices by voice command. In this type of IoT platform, the voice control system plays the key role. In summary, the smart home represents a novel network architecture. In the next section, we present some emerging attacks on this new architecture.

Security and Privacy Challenges for the Smart Home Platform

In this section, we summarize the potential security challenges for consumer IoT in smart homes as shown in Fig. 1. According to the different attacking interfaces (e.g., physical layer, network layer, mobile applications, access control, and voice user interface), these challenges can be classified into the following five types.

Attack I: Jamming Attacks on the Physical Layer: The main security threat in the physical layer of a smart home network is a jamming attack. The aim of a jamming attack is to disturb the communications between the smart device and the hub. To achieve this goal, the attacker exploits a high-power radio source (jammer) to emit wireless signals with the same working frequency as the network [1]. In the worst case jamming attack (e.g., a jammer emits signals with very high power), the whole communications of the IoT system will be paralyzed. More seriously, the smart devices suffering jamming attacks have to send wireless packets due to the unstable communication environment; therefore, the batteries of smart devices will be drained drastically.

Attack II: Wireless Traffic Analysis on the Network Layer: In smart home networks, smart devices always communicate with each other through local wireless network (e.g., Wi-Fi, ZigBee, Z-Wave network). An eavesdropping adversary can sniff the wireless channel, analyze the semantics of network traffic, reverse-engineer the communication protocols, or even perform spoofing attacks on IoT systems. In [2], an automatic spoofer is proposed to analyze and reconstruct the wireless customized protocols over IEEE 802.15.4. Real-world systems can be spoofed successfully by the automatically generated packets. More seriously, the attacker can infer the device execution states from wireless traffic, which may potentially compromise the user's privacy since most IoT events are highly correlated with human activities [3].

Attack III: Over-Privileged Attacks on Mobile Applications: Most smart home platforms support third-party apps to support more IoT services. By running on the cloud back-end, these apps provide methods to remotely manage the working logic of smart devices or service through the central hub and wireless network. Since these apps can process the data flow among the devices after being authorized, imperfect access control systems may become a potential threat. For example, Samsung SmartThings applies a capability-based model to regulate the behavior of apps. However, the coarse granularity is pointed out as a design flaw in this authorization model [4]: many apps can potentially gain access to privileges that were not explicitly requested. For example, an app that only needs partial right of a capability (e.g., *on*() command of capability *lock*) will always be granted the whole capability. In this way, other functions (e.g., *off*() command) can be potentially abused. Therefore, malicious apps can exploit this flaw to launch various attacks, including fake alarm and backdoor pin code injection.

Attack IV: Unauthorized Device Access: Authorized device access is vital for smart home security. However, in real-world smart home IoT systems, due to poor implementation of security protocols or lack of appropriate authentication mechanisms, intruders can establish malicious connection with devices and gain access to sensitive data. For example, some ZigBee devices may use the default link key during network joining, and an attacker can illegally hijack the connection and take control of devices. Besides, a nearby adversary can exploit a weak authentication mechanism in Bluetooth to attack wearables or personal devices. On the application layer, many vendors also fail to carefully implement the authentication mechanisms. A remote adversary can launch an attack toward smart home devices with weak protection over the Internet. According to a recent report released by WikiLeaks,[1] dozens of IoT smart devices, such as Samsung smart TVs, are turned into silent listening and monitoring devices by the Central Intelligence Agency.

Authorized device access is vital for smart home security. However, in real-world smart home IoT systems, due to poor implementation of security protocols or lack of appropriate authentication mechanisms, the intruders can establish malicious connection with devices and gain access to the sensitive data.

Attack V: Spoofing Attacks on Voice User Interface: As the primary user interface in most smart home platforms, the VUI is becoming the new target of spoofing attacks, which can be classified into *replay attacks*, *hidden command attacks*, and *inaudible command attacks*. In a replay attack, an adversary tries to fool the VUI by using the pre-recorded voice of a legitimate user. Hidden command attacks use a falsified speech signal as the system input [5]. As an extreme case of spoofing attack, the latest research [6] shows that it is possible to inject some hidden or even inaudible voice commands that cannot be understood/heard by the human but can still be understood by the VUI. This kind of spoofing attack opens a new door for the adversary to query a user's sensitive information and perform undesirable operations, which poses a serious threat to the security of smart home systems.

In summary, there are some new security challenges for smart home IoT. In the next section, we introduce the state-of-the-art countermeasures.

Existing Proposals for Securing Smart Home Platforms

In this section, we summarize the existing works related to the security threats of smart home platforms.

Thwarting a Jamming Attack: When the jammer is outside the smart home platform, an intuitive approach to thwart a jamming attack is changing the frequency of the communication channel. The hub and devices in a smart home can change the channel frequency periodically with pre-shared channel sequence numbers. For the case in which the jammer is the smart device of the IoT platform, [1] proposes a jammer inference framework to infer the likelihood of a device being a jammer. This framework detects a jammer based on the observed jamming events, and can work well in multiple-jammer scenarios.

Thwarting Traffic Analysis Attacks: Traffic attacks typically arise due to the exposed and unprotected wireless network environment. An effective approach to thwart an eavesdropping attack is traffic encryption. Currently, the popular wireless standards (e.g., ZigBee, Z-Wave) have adopted strong encryption mechanisms on the transmitted messages. ZigBee supports 128-bit AES-CCM* encryption mode in the network layer,

[1] https://www.wired.com/2017/03/cia-can-hack-phone-pc-tv-says-wikileaks/

WiVo consists of the following four steps. First, WiVo collects the voice samples and their corresponding CSI data. Second, WiVo removes the noises in CSI and segments the syllables from the voice samples. Third, WiVo selects appropriate features from different levels. Finally, WiVo determines whether the received voice command is an authentic one or is suffering from spoofing attacks.

while Z-Wave provides the S2 security solution, which implements the 128-bit AES encryption method. For the encrypted traffic, a potential attack strategy is leveraging machine-learning-based meta data analysis, which can be thwarted via randomly injecting some dummy traffic.

Thwarting App Over-Privilege: This security problem typically results from the coarse granularity of the privilege mechanisms of developer-friendly platforms. A study [4] performed an empirical security evaluation on the SmartThings platform and identified several design flaws that may cause app over-privilege. There are some existing approaches [7–9] to improve the granularity of access control systems. In [4], a system named *FlowFence* is presented to require consumers of sensitive data to declare their intended data flow patterns. Also, [7] proposes a context-based permission system to provide IoT platforms with contextual integrity by supporting fine-grained context identification for sensitive actions. Another study [8] uses code analysis and natural language processing to automatically collect security-relevant information from app code and description, and then generates an authorization user interface for app-sensitive operations.

Thwarting Unauthorized Devices: As unauthorized device threats are mainly caused by the weak authentication mechanism, there are two types of approaches to tackle this problem. The first way is to add authentication to the applications. For example, in [10], cryptographic secret handshakes between mobile devices on top of Bluetooth Low Energy are introduced to enhance the authentication. On the other side, a large body of work addresses the weak authentication problem [11, 12] by introducing the authentication enhancement between IoT devices. Specifically, identity authentication and a capability-based access control model are proposed for the smart home environment. This mechanism is often implemented on communication protocols to protect the device-level authentication.

Thwarting Voice Spoofing Attacks: To enhance the security of the voice interface against voice spoofing attacks, the key issue is distinguishing the voice samples generated by the legitimate user and the attacker. Existing articles propose two-factor-based voice authentication schemes. An acceleration-based scheme is proposed in [13], which collects acceleration data from the user's wearable devices. The insight of this scheme is that when a real human speaks a voice command, it will generate unique vibration on his/her skin. Therefore, this scheme monitors the vibration of human skin using an accelerator, and leverages it as a key differentiating factor between a human speaker and a machine (e.g., loudspeaker). A proposal in [14] utilizes the Doppler effect of ultrasonics generated from the loudspeaker of a smartphone to perform voice authentication. When a real user speaks a voice command, his/her mouth motion will introduce special Doppler frequency bias of the reflected ultrasonics, and thus can be used to determine whether the received voice command is generated from a real human or a machine.

These aforementioned proposals require the user to carry specialized sensing devices to collect the liveness information, which may potentially limit them being adopted in practice. In the next section, we introduce a novel device-free liveness detection framework by leveraging the prevalent wireless signals in the IoT environment.

Wireless-Signal-Based Liveness Detection Framework

In this section, we propose a two-factor authentication framework named WiVo to thwart voice spoofing attacks. Specifically, WiVo leverages the prevalent wireless signals to differentiate a legitimate voice command and a spoofing one, and thus does not require users to carry any additional devices or sensors [15]. We elaborate the basic idea and system design of WiVo as below.

The Basic Idea of WiVo

In a spoofing attack, the adversary aims to spoof the VUI by using a pre-recorded user's voice or synthesized inaudible voice commands. The basic insight of WiVo is that, different from a fake voice command, authentic voice commands should have corresponding mouth motions. This observation motivates us to distinguish a fake voice command from authentic ones by checking if their mouth motions are consistent by leveraging channel state information (CSI)-based device-free sensing technologies. In particular, WiVo aims to build the correlation between the CSI change and the mouth motion, and leverages this correlation to verify the liveness of a voice command.

We perform an experiment to validate our insight. As shown in Fig. 2a, when a user speaks a voice command, WiVo exploits a pair of antennas on the IoT devices to collect the CSI of wireless signals, and the microphone starts recording the voice samples simultaneously. As shown in Fig. 2b, the dramatic fluctuations of CSI waveforms happen with the occurrence of a human voice. Therefore, it is feasible to leverage the consistency of fluctuations between collected voice samples and CSI data to detect spoofing attacks.

The System Design of WiVo

WiVo consists of the following four steps. First, WiVo collects the voice samples and their corresponding CSI data. Second, WiVo removes the noises in CSI and segments the syllables from the voice samples. Third, WiVo selects appropriate features from different levels. Finally, WiVo determines whether the received voice command is an authentic one or is suffering from spoofing attacks.

Voice and CSI Collection: It is technically feasible to collect the voice samples and their corresponding CSI data simultaneously in a smart home. For most VUIs, it is required for the user to speak a predefined word, which can be utilized as a trigger (e.g., "Alexa" in Amazon Echo). WiVo only starts when the voice trigger is recognized by the VUI. After WiVo has been activated, WiVo utilizes two antennas to collect CSI data. These antennas can be equipped by different devices in the smart home. WiVo allows a transmit antenna to continuously send wireless packets (e.g., broadcast packets) and another antenna to receive packets, and extracts CSI data from these packets.

Data Preprocessing: For the collected CSI data, WiVo leverages the wavelet de-noising to

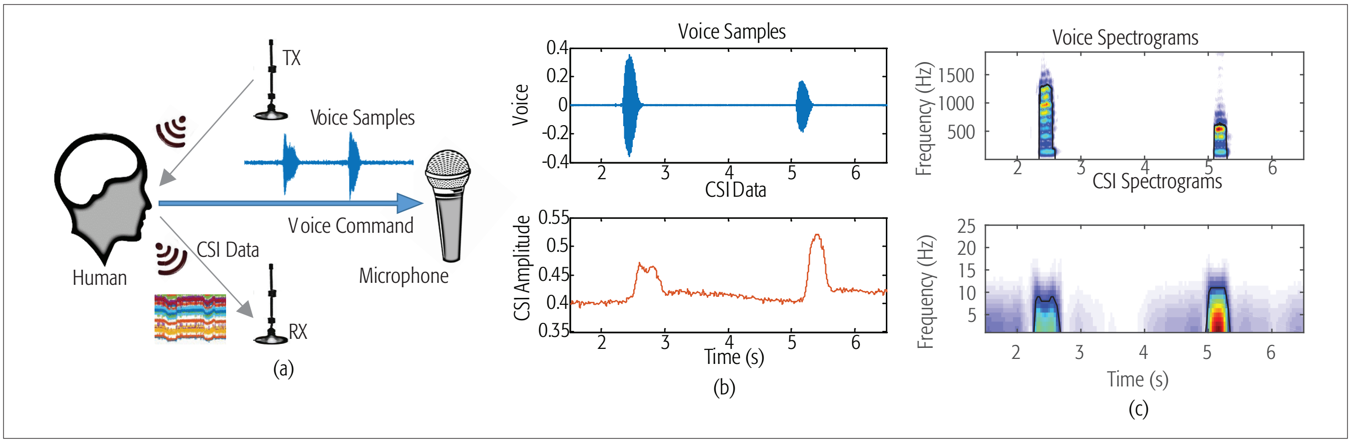


FIGURE 2. Illustration of WiVo: a) framework scenario; b) voice and CSI waveforms; c) voice and CSI spectrograms.

eliminate high-frequency noises. For the voice samples, WiVo performs syllable detection on the voice samples. To detect the start and end points of a syllable, WiVo utilizes the Munich Automatic Segmentation System (MAUS), a widely adopted phonetic segmentation system.[2] Then WiVo extracts the CSI syllable data according to the timestamps of corresponding voice samples.

Feature Selection: To calculate the consistency between the voice samples and the CSI data, WiVo needs to extract appropriate features. The first feature was based on the presence of voice. As shown in Fig. 2b, CSI variation occurs along with human pronunciation. Thus, WiVo first performs short time Fourier transform (STFT) on the CSI data and voice samples to obtain their two-dimensional frequency spectrograms. Then, as shown in Fig. 2c, WiVo resizes the CSI spectrogram with frequency from 0 to 30 Hz into an $M \times n$ matrix $M_{CSI(i,h)}$, and chooses a pre-defined *threshold* to get the contour $C_{CSI(i)}$, where $i = 1, \ldots, n$. $C_{CSI(i)}$ is the maximum value j which satisfies that $M_{CSI(i,j)} \geq threshold$. Calculating contours $C_{Voice(i)}$ for the voice spectrograms is similar to calculating $C_{CSI(i)}$. Figure 2c shows the spectrograms and marks contours for both voice samples and CSI data. In a non-attack scenario, the contours of CSI and voice samples have similar variations due to the presence of mouth motions. Thus, we measure the correlation *Corr* between these two contours by adopting a Pearson correlation coefficient. $|Corr|$ ranges from 0 to +1, where a higher value of $|Corr|$ represents a higher level of similarity. We utilize $|Corr|$ as the first feature.

WiVo further extracts features from the CSI syllable data. As shown in Fig. 3, it is observed that similar mouth motions can cause similar CSI vibrations. For instance, the CSI subcarrier waveforms of syllable /*a*:/ and /*la*:/ have similar shapes and amplitude vibrations, but the waveforms of /*a*:/ and /*u*:/ are quite different. Thus, we can extract the shapes from the CSI waveforms as their time domain features, and the contours from the CSI frequency spectrograms as the frequency domain features. Before launching WiVo, for each syllable, we pre-collect its corresponding CSI data. Then, for each collected CSI data, WiVo finds its corresponding N_s syllables by analyzing the collected voice samples. To reduce the computation, we divide the syllables into four categories according to their corresponding mouth motions. The four mouth motion types are hiant (e.g., heard like "bar" and "ha"), grin (e.g., heard like "a" and "bay"), round (e.g., heard like "law" and "saw"), and pout (e.g., heard like "root" and shoe"). After such division, different types of syllables can be correlated with different CSI features.

WiVo utilizes the Dynamic Time Wrapping (DTW) method to calculate the similarity $S_{syll(i)}$ between the *i*th syllable and its pre-collected CSI profile. Finally, WiVo generates the syllable-level feature $S_{syll} = \prod_{i=1}^{N_s} S_{syll(i)}$, where $i = 1, \ldots, N_s$.

Liveness Detection: After WiVo extracts features from collected signals, we can calculate the final decision score of the input, which is calculated as $Score = |Corr| \times S_{syll}$. WiVo utilizes a threshold-based mechanism to perform human liveness detection in this article. For the given voice command input, if its *Score* is larger than the pre-defined threshold, WiVo regards it as an authentic voice command. Otherwise, WiVo judges it as a fake command and refuses to execute it. We give a detailed experimental evaluation in the following section.

Experiment and Evaluation

Experiment Setup: WiVo consists of two unites of hardware:

1. A universal software radio peripheral (USRP) N210 device that connects two commercial WiFi antennas
2. A microphone responsible for collecting voice samples

The USRP N210 acts as an IoT device in a smart home, and collects CSI data at the rate of 1000 pkts/s. WiVo is incorporated with the Samsung SmartThings platform, which is compatible with Amazon Alexa, a popular VUI around the world. As shown in Fig. 4, we develop a SmartApp in the SmartThings platform to implement the function of WiVo. When Alexa receives the human voice command "let there be light," it will send the corresponding command to the hub, and at the same time, WiVo performs liveness detection by analyzing the collected CSI and voice samples. SmartApp will open the smart light if and only if the liveness detection of WiVo is successful. Otherwise, SmartApp regards the voice command as an inauthentic one and does not execute it.

Five volunteers were recruited in the experiment, and before performing voice commands,

[2] http://www.bas.uni-muenchen.de/Bas/BasMAUS.html

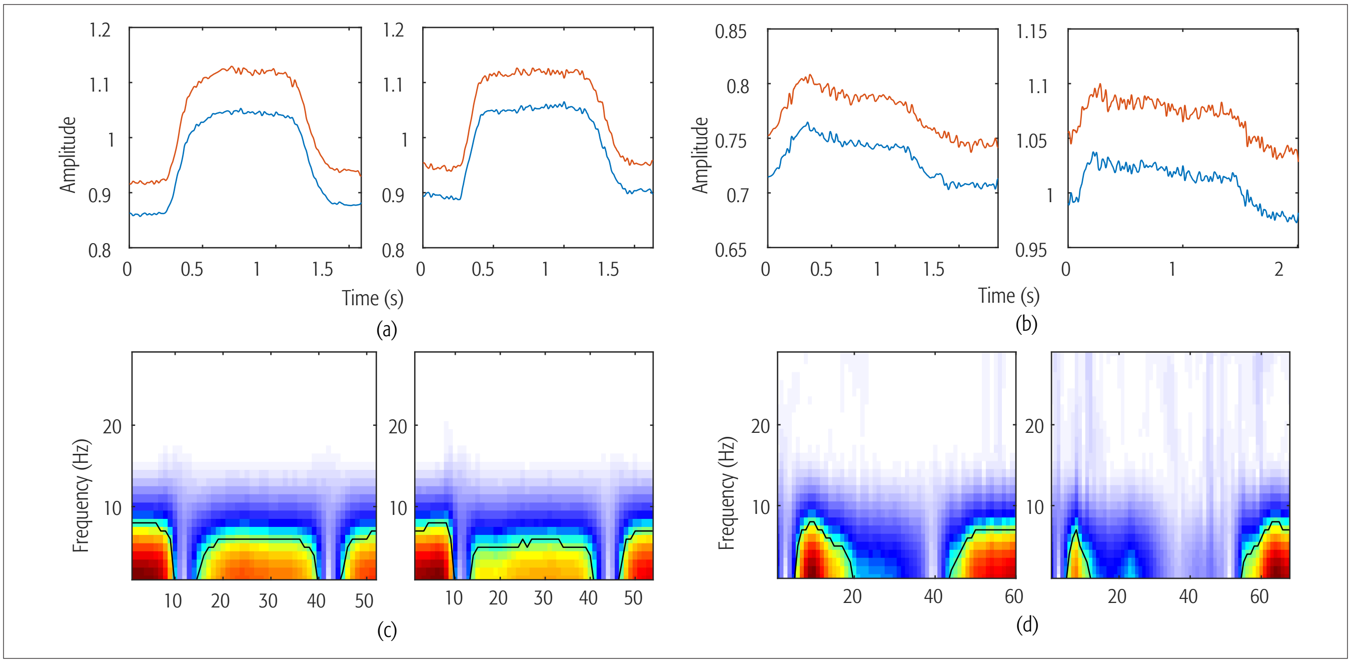
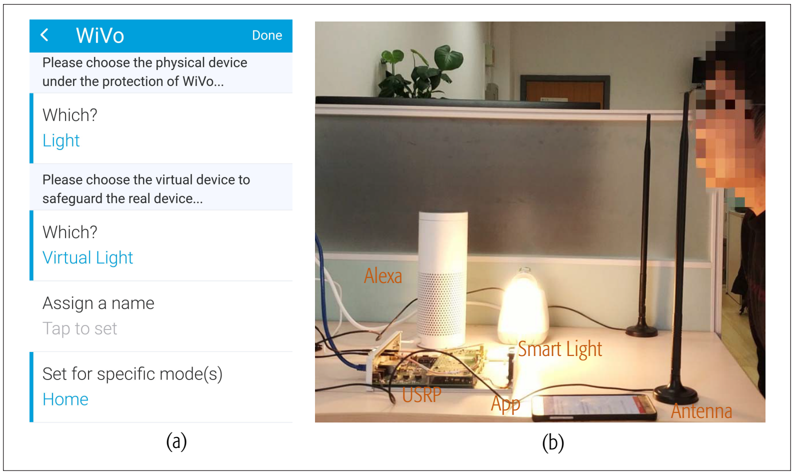


FIGURE 3. Different syllables: a) two CSI subcarriers waveforms of /a:/ and /la:/; b) two CSI subcarriers waveforms of /u:/ and /gu:/; c) CSI spectrograms of /a:/ and /la:/; d) CSI spectrograms of /u:/ and /gu:/.


WiVo Done

Please choose the physical device under the protection of WiVo...

Which?
Light

Please choose the virtual device to safeguard the real device...

Which?
Virtual Light

Assign a name
Tap to set

Set for specific mode(s)
Home

FIGURE 4. Testbed: a) SmartApp user interface; b) real case scenario.

each of them was required to perform the four categories of mouth motions (i.e., syllables) 10 times as pre-collected syllable profiles. To assess the performance of WiVo, we choose the false accept rate (FAR) and true accept rate (TAR) as metrics. Both FAR and TAR are influenced by adjusting the verification threshold, and their relationships are shown in a receiver operating characteristic (ROC) curve.

Detection Accuracy: We first evaluate the effectiveness of WiVo to defend against spoofing attacks. To perform legitimate voice commands, each volunteer is required to speak 150 voice commands. After that, we perform spoofing attacks for each user's syllable profiles 750 times. For a total of 4500 voice commands, the lengths of those syllables range from 4 to 8. Figure 5 depicts the ROC curve of WiVo, and we observe that with 1 percent FAR, the detection rate is as high as 99 percent. The average time delay of performing per liveness detection is 0.32 s, which is acceptable in practice. In summary, our experimental results well validate the effectiveness of WiVo in defending against spoofing attacks.

Performance in a Multiple-User Scenario: In the ideal case, for each user, WiVo performs liveness detection based on his/her CSI syllable profiles. However, in some smart home environments with multiple users, it is inconvenient to collect each user's syllable profiles. A more desirable design is to collect once but work for multiple users. In our experiment, we first recruit a volunteer to provide WiVo with his/her syllable profiles, and then recruit another volunteer to perform voice commands 450 times. After that, we implement spoofing attacks 450 times. Figure 5 shows the evaluation result of WiVo, where WiVo achieves 96 percent TAR with 1 percent FAR. The detection rate is smaller than that in the single-user scenario, since the mouth motion of another volunteer is not the same as the user that provides the pre-collected syllable profiles. However, the detection accuracy is high enough for WiVo to perform liveness detection in the multiple-user scenario.

Future Work: There are some limitations that may degrade the performance of WiVo. In the experiment, the user performs voice commands in a stable environment. However, the collected CSI data may be interfered by the movement of surrounding objects. A potential countermeasure is utilizing a sophisticated method such as multiple-input multiple-output (MIMO) beamforming to improve the wireless sensing capability. In addition, when the distance between the user and the antennas of WiVo is too long, the collected CSI cannot reflect the mouth motion components. Increasing the density of IoT devices to make sure that the user is located in the effective range of WiVo is a practical solution to address this limitation.

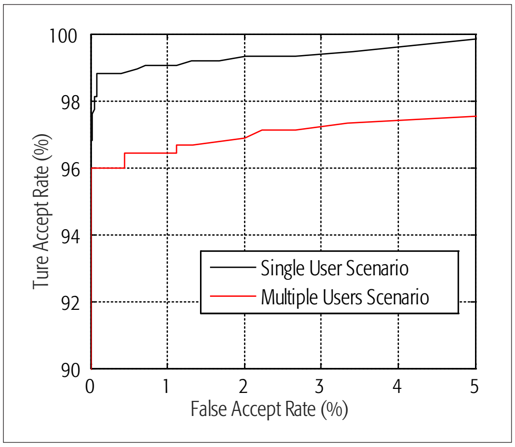


FIGURE 5. The performance of WiVo on thwarting spoofing attacks.

Conclusion

In this article, we first survey the emerging security and privacy challenges for the consumer IoT in smart homes, and discuss the existing proposed countermeasures. To thwart the attacks on the VUI of smart homes, we propose a novel wireless-signal-based liveness detection framework named WiVo. WiVo utilizes the prevalent wireless signals in IoT environments to sense human mouth motion, and then verifies the liveness of voice commands according to the consistency between voice samples and CSI data. We implement WiVo on a SmartThings platform to demonstrate its feasibility and effectiveness.

To thwart the attacks on the VUI of smart homes, we propose a novel wireless-signal-based liveness detection framework named WiVo. WiVo utilizes the prevalent wireless signals in IoT environments to sense human mouth motion, and then verifies the liveness of voice commands according to the consistency between voice samples and CSI data.

Acknowledgment

This work is supported by the National Science Foundation of China (No. U1401253, 61672350).

References

[1] H. Zhu *et al.*, "You Can Jam But You Cannot Hide: Defending Against Jamming Attacks for Geo-Location Database Driven Spectrum Sharing," *IEEE JSAC*, vol. 34, no. 10, 2016, pp. 2723–37.

[2] K. Choi *et al.*, "Dissecting Customized Protocols: Automatic Analysis for Customized Protocols Based on IEEE 802.15.4," *Proc. ACM WiSec*, 2016, pp. 183–93.

[3] W. Zhang *et al.*, "HoMonit: Monitoring Smart Home Apps from Encrypted Traffic," *Proc. ACM CCS*, 2018.

[4] E. Fernandes *et al.*, "FlowFence: Practical Data Protection for Emerging IoT Application Frameworks," *Proc. USENIX Security*, 2016, pp. 531–48.

[5] N. Carlini *et al.*, "Hidden Voice Commands," *Proc. USENIX Security*, 2016, pp. 513–30.

[6] G. Zhang *et al.*, "DolphinAttack: Inaudible Voice Commands," *Proc. ACM CCS*, 2017, pp. 103–17.

[7] Y.J. Jia *et al.*, "ContexIoT: Towards Providing Contextual Integrity to Appified IoT Platforms," *Proc. NDSS*, 2017.

[8] Y. Tian *et al.*, "SmartAuth: User-Centered Authorization for the Internet of Things," *Proc. USENIX Security*, 2017, pp. 361–78.

[9] Y. Cheng *et al.*, "A Lightweight Live Memory Forensic Approach Based on Hardware Virtualization," *Info. Sciences*, vol. 379, 2017, pp. 23–41.

[10] Y. Michalevsky, S. Nath, and J. Liu, "MASHaBLE: Mobile Applications of Secret Handshakes over Bluetooth LE," *Proc. ACM MobiCom*, 2016, pp. 387–400.

[11] L. Wu, X. Du, and J. Wu, "Effective Defense Schemes for Phishing Attacks on Mobile Computing Platforms," *IEEE Trans. Vehic. Tech.*, vol. 65, no. 8, 2016, pp. 6678–91.

[12] Z. Guan *et al.*, "Achieving Efficient and Secure Data Acquisition for Cloud-Supported Internet of Things in Smart Grid," *IEEE Internet of Things J.*, vol. 4, no. 6, 2017, pp. 1934–44.

[13] H. Feng, K. Fawaz, and K. Shin, "Continuous Authentication for Voice Assistants," *Proc. ACM MobiCom*, 2017, pp. 343–55.

[14] L. Zhang, S. Tan, and J. Yang, "Hearing Your Voice Is Not Enough: An Articulatory Gesture Based Liveness Detection for Voice Authentication," *Proc. ACM CCS*, 2017, pp. 57–71.

[15] Y. Meng *et al.*, "WiVo: Enhancing the Security of Voice Control System via Wireless Signal in IoT Environment," *Proc. ACM MobiHoc*, 2018, pp. 81–90.

Biographies

Yan Meng (yan_meng@sjtu.edu.cn) is a Ph.D. candidate in the Department of Computer Science and Engineering, Shanghai Jiao Tong University, China. He received his B.S. degree in electronic and information engineering from Huazhong University of Science and Technology in 2016. His research interests include wireless network security and social network privacy.

Wei Zhang (zhang-wei@sjtu.edu.cn) is a graduate student working toward his M.Sc. degree in the Department of Computer Science and Engineering, Shanghai Jiao Tong University. He received his B.S. degree in computer science from the University of Science and Technology of China in 2016. His research interests include Internet of Things security and social network privacy.

Haojin Zhu [M'09, SM'16] (zhu-hj@cs.sjtu.edu.cn) received his B.Sc. degree (2002) from Wuhan University, China, and his M.Sc. (2005) degree from Shanghai Jiao Tong University, both in computer science, and his Ph.D. in electrical and computer engineering from the University of Waterloo, Canada, in 2009. Since 2017, he has been a full professor in the Computer Science Department of Shanghai Jiao Tong University. His current research interests include network security and privacy enhancing technologies. He has published more than 40 international journal papers, in publications including the *IEEE Journal on Selected Areas in Communications*, *TDSC*, *TPDS*, *TMC*, *TWC*, *and IEEE Transactions on Vehicular Technology*, and 60 international conference papers, at conferences including ACM CCS, ACM MOBICOM, ACM MOBIHOC, IEEE INFOCOM, and IEEE ICDCS. He has received a number of awards including: IEEE ComSoc Asia-Pacific Outstanding Young Researcher Award (2014), Top 100 Most Cited Chinese Papers Published in International Journals (2014), Supervisor of Shanghai Excellent Master Thesis Award (2014), Distinguished Member of the IEEE INFOCOM Technical Program Committee (2015), Outstanding Youth Post Expert Award for Shanghai Jiao Tong University (2014), and SMC Young Research Award of Shanghai Jiao Tong University (2011). He was a co-recipient of best paper awards at IEEE ICC 2007 and Chinacom 2008, IEEE GLOBECOM Best Paper Nomination (2014), and WASA Best Paper Runner-up Award (2017). He received the Young Scholar Award of the Changjiang Scholar Program by the Ministry of Education of P.R. China in 2016.

Xuemin (Sherman) Shen [M'97, SM'02, F'09] (xshen@bbcr.uwaterloo.ca) is a university professor, Department of Electrical and Computer Engineering, University of Waterloo. He was also the associate chair for graduate studies. His research focuses on resource management, wireless network security, social networks, smart grid, and vehicular ad hoc and sensor networks. He was an elected member of the IEEE ComSoc Board of Governors and the Chair of the Distinguished Lecturers Selection Committee. He served as the Technical Program Committee Chair/Co-Chair for IEEE GLOBECOM '16, IEEE INFOCOM '14, IEEE VTC-Fall '10, and IEEE GLOBECOM '07. He received the Excellent Graduate Supervision Award in 2006, and the Outstanding Performance Award in 2004, 2007, 2010, and 2014 from the University of Waterloo. He is a registered professional engineer of Ontario, Canada, an Engineering Institute of Canada Fellow, a Canadian Academy of Engineering Fellow, and a Royal Society of Canada Fellow, and was a Distinguished Lecturer of the IEEE Vehicular Technology Society and the IEEE Communications Society.