APPLICATION OF LEARNING ALGORITHMS IN SMART HOME IOT SYSTEM SECURITY

JIAN MAO*, QIXIAO LIN AND JINGDONG BIAN

School of Electronic and Information Engineering
Beihang University, Beijing, China

(Communicated by Zhipeng Cai)

Abstract. With the rapid development of Internet of Things (IoT) technologies, smart home systems are getting more and more popular in our daily life. Besides providing convenient functionality and tangible benefits, smart home systems also expose users to security risks. To enhance the functionality and the security, machine learning algorithms play an important role in a smart home ecosystem, e.g., ensuring biotechnology-based authentication and authorization, anomalous detection, etc. On the other side, attackers also treat learning algorithms as a tool, as well as a target, to exploit the security vulnerabilities in smart home systems. In this paper, we unify the system architectures suggested by the mainstream service providers, e.g., Samsung, Google, Apple, etc. Based on our proposed overall smart home system model, we investigate the application of learning algorithms in smart home IoT system security. Our study includes two angles. First, we discussed the functionality and security enhancing methods based on learning mechanisms; second, we described the security threats exposed by employing learning techniques. We also explored the potential solutions that may address the aforementioned security problems.

1. Introduction. Due to the continuous improvement of telecommunication technology, the world has moved from the Internet-of-Computer era into the Internet-of-Things era. Nowadays, our homes, hospitals, factories, and cities are being enhanced with devices that have computational and networking capabilities [13]. This emerging network of connected Internet of Things (IoT), has evolved continuously. In 2005, International Telecommunication Union published a report about IoT, proposing a truly ubiquitous network anytime, anywhere, by anyone and anything [48]. Nowadays, Internet of Things literally means a world-wide network of interconnected objects uniquely addressable, based on standard communication protocols [6]. IoT is in the process of a rapid development, with a forecast of 20.8 billion connected devices in 2020 [15]. To provide convenience and efficiency, programming and intelligence technologies are widely applied to IoT systems, which transforms traditional IoT systems into smart IoT systems. Smart home is a typical implementation of such evolution.

2010 Mathematics Subject Classification. Primary: 58F15, 58F17; Secondary: 53C35.
Key words and phrases. Machine learning, smart home, security, internet of things.

The first author is supported by the National Natural Science Foundation of China (No.61402029, 61379002, 61370190, 61672083) and the Funding Project of Shanghai Key Laboratory of Integrated Administration Technologies for Information Security (No. AGK201708).

* Corresponding author: Jian Mao.
The early concept of smart home was proposed in 1992 and smart home is defined as the integration of different services within a home by using a common communication system [30]. Berlo et al. introduced automatic control to the definition of smart home [49]. Alam et al. defined smart home as an application of ubiquitous computing that is able to provide users context-aware automated or assistive services in the form of ambient intelligence, remote home control or home automation [2].

With the rapid development of IoT technologies, smart home breaks through the restriction caused by complicated device configurations and system setup procedures. Recently, cloud-based backend platforms are suggested by several smart home service providers, which provide programming frameworks to achieve easier system setup and convenient third-party application development. Such typical frameworks include Samsung’s SmartThings, Apple’s HomeKit, Vera Control’s Vera3, Google’s Weave/Brillo, and AllSeen Alliance’s AllJoyn, etc. [12].

Beyond the basic convenience functionalities in daily life, e.g., automatical light adjustment, door locker, etc., smart home systems also provide perceptible benefits. Such benefits include ensuring energy saving and efficient resource usage with the help of water flow sensing and smart meter optimization, achieving better home security control by connecting IP-based cameras, motion sensors, and door locks, etc.

Although smart home provides their customers/“inhabitants” such tangible benefits as comfort living, healthcare, and home security control [2], security vulnerabilities in smart home systems will lead to device invasion (e.g., remote control [1] [54]), privacy violation (e.g., eavesdropping [34] or surveillance [18], location or habit leakage [26][59]), monetary loss (e.g., break-in through unlocking door [19]), or even loss of life (e.g., inducing seizures in epileptic users [35]). Security becomes one of the most critical concerns in the development of smart home systems [51].

Machine learning evolved from the study of pattern recognition and computational learning theory in artificial intelligence. Machine learning algorithms are used to build a model from a training set of input observations in order to make data-driven predictions or decisions expressed as outputs without following explicitly static program instructions. Currently, learning mechanisms have been widely used in different areas, such as, biomedical informatics, computer vision, data mining, pattern recognition, search engine, and recommendation system, etc. Due to the distinguished properties and excellent performances in data mining and pattern recognition (especially, facial/speech recognition) applications, learning algorithms are often applied in smart home platforms to improve system functionalities. In addition, starting with the earliest application of naive Bayes classifiers for spam detection [3], machine learning algorithms have been increasingly applied to study complicated security problems. For example, Sommn et al. proposed an anomaly detection model using machine learning to identify network intrusions or other malicious activities [47].

In this paper, we unified the system architectures suggested by the mainstream service providers, e.g., Samsung, Google, Apple, etc. Based on our proposed overall smart home system model, we investigate the application of learning algorithms in smart home IoT system security. Our study consists of two aspects. First, we discussed the functionality and security enhancing methods based on learning mechanisms; second, we described the security threats exposed by employing learning
techniques. We also explored the potential solutions that may address the aforementioned security problems.

Contributions. In summary, we make following contributions in this paper.

- We presented a universal smart home system model, which is compatible with the frameworks suggested by the mainstream service providers. According to the proposed architecture, we discussed the functionality and security enhancing methods based on learning mechanisms.
- We systematically discussed the application of learning algorithms in smart home system privacy and side-channel analysis (i.e., treating learning algorithms as attack tools); we also presented the attacks to learning algorithms and the security threats exposed by employing learning techniques (i.e., treating learning algorithms as attack targets).
- We explored the potential solutions to address the aforementioned security problems, which include privacy preserving solutions against learning-based analysis and robust learning-based authentication/authorization mechanisms.

Paper Organization. The rest of this paper is organized as follows. Section 2 illustrates the background and proposes a universal model of smart home systems. Section 3 discusses functionality and security enhancements via learning; Section 4 describes attacks to smart home systems via learning mechanisms; Section 5 presents our suggestions and the potential defense solutions; and Section 6 concludes the paper.

2. Overview of smart home IoT system. In this section, we present the overall architecture of a smart home IoT system. As Samsung’s SmartThings is one of the most widely deployed smart home platforms and many other popular cloud-based systems, such as Google’s Weave/Brillo and Apple’s HomeKit, also use the similar design except for the slight differences in communication protocols [22], we unify the smart home system model according to the design of SmartThings and the universal architecture is illustrated in Figure 1.

As shown in Figure 1, a smart home system includes four parts, Home Space, Cloud backend Platform, External Systems and End Applications. Home space consists of the smart nodes/devices and a Hub/Gateway. The hub/gateway setup by a user enables several low-layer protocols, such as ZigBee, WiFi, Bluetooth and ZWave to communicate with the physical smart devices in the user’s home.

The cloud backend platform has three important components, SmartApp, Access Control Mechanism (Permission/Capability system in SmartThings), SmartDevice.
SmartApps are IoT apps executed by the cloud platform in a sandbox environment rather than running on smart devices. The SmartDevices are software wrappers for smart home devices. In a smart home system, instead of direct communication with a smart home/hardware device, a SmartApp communicates with a SmartDevice instance that encapsulates a physical device. A SmartDevices manages smart devices by using communication protocols, e.g., ZWave, WiFi, and ZigBee. In this way, physical smart devices are integrated into the smart home system. The access control mechanism is designed as a permission model and in Samsung SmartThings it is called the capability system. The access control module specifies the commands and attributes that smart devices can support, while SmartApps state the capabilities required by the devices. According to this mechanism, SmartDevices and SmartApps are combined during the app installation phase. SmartApps and SmartDevices run on the cloud backend and are published to the smart home app stores (e.g., SmartThings app store), which can be accessed by corresponding End Applications installed/run on users’ mobile phone, tablets, laptops or PCs. Users configure the hubs/gateways, associate smart devices, and install SmartApps from Appstores using the End applications (called SmartThings Mobile in SmartThings platform). The SmartApp can choose to expose web service endpoints to HTTP requests (e.g., GET, PUT, POST, and DELETE) from external application/systems (e.g., IFTTT), which is protected by OAuth-based authentication.

In the architecture of smart home systems, learning-based solutions can be used to enhance functionalities, as well as to attack components of the system, such as bio-technology-based authentication/authorization mechanisms. To facilitate the analysis of learning algorithm applications in smart home system security, we separate the overall smart home architecture into three layers, i.e., hardware layer, control layer, and processing layer, which are highlighted with different colors in Figure 1. Hardware layer consists of physical devices that interact with the hub/gateway. Control layer includes the hub/gateway, SmartDevices and Access Control module in the cloud backend platform. SmartApps, External Systems and End Applications comprise the Processing layer. In the hardware layer, physical devices perceive, collect the state information of the home environment and the devices themselves. All these state information is transmitted from the physical hardware to the hub/gateway (located in Control layer) and reaches cloud platforms for further processing; Processing layer analyzes the information collected by the Hardware layer, infers the state changes and the environment status and creates the corresponding responses/decisions/commands that will be assigned by the control layer (to hardware layer).

3. Functionality and security enhancements via learning. According to the proposed overall smart home architecture, in this section, we discuss the applications of learning algorithms in functionality and security enhancements. A brief summary is illustrated in Table 1.

3.1. Functionality ensurance by learning algorithms.

3.1.1. Data-mining-based service optimization. Recent studies [45], [24] show that the IoT devices should be intelligent and can perceive events, interact with humans, and make decisions on their own. How do we convert the data collected by IoT devices into human acceptable knowledge and create a more convenient environment? Machine-learning-based data mining technology can be seen as one of
the most promising technologies, because they have a powerful ability to find useful knowledge that is hidden in smart home IoT data. These knowledge can be used to improve system performances and create an intelligent environment with better quality of service. In a smart home system, learning algorithms are often applied to face recognition, activity recognition, behavior prediction and so on.

3.1.2. Image-speech-recognition-based system control. Machine learning has been studied extensively and become a powerful tool in recognition applications, including image recognition, speech recognition, etc. It has been a long time since researchers applied learning algorithms to image recognition, and in recent years deep learning is a specific type of feature learning that has shown great success in this application scenarios [33], including deep residual learning [17], deep convolutional networks [46] and so on. Moreover, as a branch of image recognition, face recognition technology has been through a growing process and now is put into practical production. Face recognition also uses learning algorithm to improve accuracy [38] [57].

In a smart home platform, image recognition using learning algorithm can be widely applied to meet user requirements. DeepFood [28] is an example that analyzes food images using learning algorithm to help users improve the accuracy of dietary assessment. Face recognition has been applied widely in smartphone platforms to verify user identities and could play a similar role in smart home platform.

As for speech recognition, machine learning is an important algorithm to identify voice as well [11]. In fact, existing smart home platforms all have a voice control function, like Apple Siri, Google Now or Amazon Echo. Recently, Apple even published Apple Machine Learning Journal [4] online, introducing how their engineers
use machine learning in different functions to help build innovative products. Istrate et al. [21] proposed a scheme that uses sound data collected from multiple microphones to identify sound events such as door opening, glass break sound, phone ring tones, footsteps, screams, and so on. The classifiers they create for detecting sound patterns use expectation maximization (EM) and k-means algorithms.

3.1.3. Incident-recognition-based home healthcare. Some other studies [60] focus on smart home health care, and they use different machine learning based detection techniques to predict sudden events such as sudden heart attacks. Unlike event detection using sound or images, this event detection requires the use of real-time data (such as human state information). Pang et al. [36] proposed a solution to solve the problem of incident handling in event detection. The scheme uses the adaptive neuro-fuzzy inference method to train the classifier for burst time detection. But how to choose a feature may have a significant impact on the accuracy of the classification results.

3.1.4. Energy saving. How to maximize the energy utilization of IoT systems has always been the focus of the researchers.

Shahriar et al. [44] designed a machine-learning-based smart home energy management system. They first collect data on solar photovoltaic panels and environmental sensors. Then they collect intelligent instrument data and motion sensor data. Finally, the machine learning algorithm is used to analyze the above two kinds of data. The energy income and consumption are predicted to design an optimal energy management system.

Ventura et al. [50] proposed a predictive model called ARIMA for energy saving of coffee machines in public places. They use the ARIMA model to derive energy usage patterns for coffee machines from historical energy usage data and predict the energy usage for the next week to achieve energy savings.

3.1.5. User preference. Another application of machine learning in smart home is to provide users with intelligent services. Choi et al. [10] described a context-aware home automation system middleware based on user preferences. They learn and predict user preferences for home appliances from six perspectives: pulse, body temperature, facial expression, room temperature, time and location. They use neural networks to learn user patterns and predict appropriate home services for their users. Rahmati et al. [39] introduced a privacy-respecting user preference collection framework that can be used for smart home. It extracts preferences which can be used by the recommended system from the user’s activity while protecting the privacy of the user. Privacy protection is achieved by precisely controlling the information the collector sends to the back end of the recommendation system.

3.2. Security enhancement by learning algorithms.

3.2.1. Anomalous behavior detection. Machine learning is increasingly being applied to IoT systems and network security, especially the use of machine learning to detect abnormal behavior of IoT systems. The common practice is to use normal system behavior and malicious attack behavior to form a training data set, followed by using the training set to train the classifier. The classifier can use the difference between normal behaviors and malicious behaviors to detect system anomalies.

Early IoT devices do not have the capability to update the system software, and even if some IoT devices have remote updates, there are still many challenges because the system software update may affect the operation of IoT devices. As a
result, many vulnerabilities will be exposed to attackers for an extended period of time. At the same time, many IoT devices use a weak password or default password, which will also cause a serious security risk. Based on the above discussion, IoT devices vulnerable to malicious software infected into botnets, causing DDoS attacks [54]. Livadas et al. [29] proposed a scheme for proactively detecting and identifying botnets. They use machine learning techniques to identify anomalous network traffic based on Internet Relay Chat (IRC), so as to achieve the purpose of detecting botnets. Zhang et al. [56] presented a semi-Markov-model-based solution to enhance the privacy checking in smart devices.

Bhide et al. [7] proposed a smart home scheme that can provide fault detection and correction for devices connected to the Smart Home frame. They use the naive Bayesian classifier algorithm to deal with the data generated when the smart home device is running to find the abnormal operation of the equipment, which makes the smart home system with self-learning ability.

3.2.2. Malware detection. To better understand the security threat in the smart IoT platform, Jia et al. [22] performed a survey of attacks reported on the smartphone platforms and studied the feasibility of their migration to the smart IoT platform. According to their investigation, there are many similarities between security strategies in smartphone platforms and smart IoT platforms. Malware could be installed on both smartphone and smart IoT devices. Thus, experience about defend against malware in smartphone platforms is also suitable for smart IoT platforms. Karbab et al. discussed a detection method based on a novel concept assuming that multiple similar Android apps with different authors are most likely to be malicious, and proposed Cypider, a framework to detect malware partly relying on learning-based patterns [23]. However, machine-learning-based malware detection remains a lot of critical challenges. The impacts of these challenges are systematically studied in [41].

For users without related technical background, it is difficult to decide whether an Android app is safe to use or not, which requires some reliable recommendation mechanism. Recently, Bahman et al. proposed an Android permission recommendation framework called DroidNet, where new apps are run under probation mode. Whether to accept or reject the permission requests is based on the decisions from peer expert users, who are identified by an expertise rating algorithm using transitional Bayesian inference model [40].

The cloud platform provides the execution environment and data storage for smart home system entities, where security audition is necessary to assure the transparency and accountability. Suryadipta et al. recently proposed a fully automated approach using learning-based techniques to extract dependency models from runtime events in order to facilitate the proactive security audition of cloud operations [32].

4. Attacks via learning algorithms. In this section, we discuss the security threats exposed by employing learning techniques. The learning-based attack vectors of smart home systems are marked in Figure 2.

4.1. To attack smart home system by exploiting vulnerabilities of learning algorithms. Existing research has extensively applied learning based techniques to satisfy the functionality and security requirements of smart home IoT systems,
Injecting anomalous input to manipulate learning algorithms result in authentication/authorization invasion.

Collecting sensing data packets and command packets to infer device states result in home/user’s privacy leakage.

**Figure 2.** Learning-based Attack Vectors in Smart Home Systems

Causative Attacks. Causative attacks change learner by changing the training data. The adversary has some control ability over learner’s training. IoT devices need to adapt to changing environments and online learning can change their predictions with input data. As a result, online learning is often used for IoT devices. However, online learning makes causative attacks easier. Adversary has the opportunity to change the training data to change learner and this attack is difficult to be detected.

At present, Voice Controllable System (VCS) and Image Recognition System have been widely used in IoT equipment, playing a user authentication and access control functions. Papernot et al. [37] tampered with the image transmitted to the automatic vehicle image recognition algorithm. The automatic vehicle-based image recognition algorithm identified the stop sign as a forward sign, causing a possible car accident. Zhang et al. [55] proposed an attack against a Voice Control System that designed an ultrasound that humans could not hear but contain voice control commands can. Due to the non-linearity of the microphone, ultrasound would produce the harmonics that could be learned by the machine. The harmonics including the control commands will be identified by the algorithm and control the device. Fogla et al. [14] proposed a method which uses the polymorphic blending attack to evade intrusion detection systems. The attacker disrupts the traffic pattern of the attack data into normal traffic mode, so that machine learning algorithm cannot identify the attack data.

Exploratory Attacks. Exploratory attacks do not change learner’s training data, but attempt to discover learner’s state information. Without any prior knowledge, the adversary can still infer the learner’s state. For example, adversary can continue to launch a probe attack on learners and collect learner feedback. Finally, the adversary can know the learner’s acceptable input data space. If the adversary has enough information, such as training data, learning algorithms, data classification
results, and learner status, the adversary will infer the learner’s decision boundary. For example, when an adversary knows a learner’s decision boundary, he can use an intrusion point that is misinterpreted by the learner to perform attack without being detected. The adversary also can use the normal point where the learner misclassifies the system to reject the normal data, or uses the normal point of the learner’s misclassification to make the system unavailable.

4.2. To attack smart home systems using learning-based techniques. With the rapid development of smart home technologies, various hardware devices equipped with crypto modules, are getting more popular. From the perspective of side-channel attacks, these new devices and applications provide them with extremely rich target devices. As the popularity of these devices and the control over them by attackers increase, the side-channels attacks are likely to be easier to implement. The core problem in side-channel attacks is classification and differentiation. As a classification algorithm, machine learning provides an effective analysis tool for side-channel attacks. Key-related side-channel information includes: power consumption, processing time, cache and etc.

Lerman et al. [25] and Hospodar et al. [20] measured power traces from cryptographic algorithm-enabled chips and used machine learning algorithms to discover the key information contained in power traces. They decompose the problem into several sub-problems and apply a machine learning algorithm for each sub-problem to reduce complexity. Lerman et al. [25] reduced the data dimension by using the learning algorithms such as Principal Component Analysis (PCA), the minimum redundancy maximum Relevance (mRMR), and Self Organizing Map (SOM). They also evaluated the analyzing performance of several machine learning algorithms that include SOM, Support Vector Machine (SVM) and Random Forest (RF) on 3DES encryption algorithm. Hospodar et al. [20] evaluated three feature selection methods: Pearson correlation coefficient approach for component selection, sum of squared pairwise t-differences (SOST) of the average signals and principal component analysis (PCA). They built an LS-SVM classifier to analyze the Advanced Encryption Standard (AES) algorithm. Template-based DPA [9] relies on parametric Gaussian estimation methods, and machine learning algorithms can overcome this disadvantage.

Physical unclonable functions (PUFs) is currently used in IoT to enhance its safety [53]. However, learning-based analysis approaches are introduced to attack the PUFs. Lim et al. [27] extracted the key from the integrated circuit based on SVM. There were solutions [42] [43] proposing a high-level modeling attack against PUFs and evaluating the performance of Logistic Regression, Evolution Strategies and SVM. Although machine learning is effective for PUF modeling, pure machine learning modeling attacks reach the limit [42]. In response to this situation, [31] propose a hybrid attack on PUF that combines power side-channel information with machine learning modeling techniques. Side-channel information can solve the problem of dramatic increase in computation. Zhang et al. [58] attack victim virtual machines running on the same physical computer through malicious virtual machines and build cache pattern classifiers based on SVM to extract fine-grained information.

Backes et al. [5] proposed a new attack that brings microphones closer to the printer and automatically recovers printed English text with machine learning, audio processing and speech recognition technique. If combined with contextual
Table 2. A taxonomy of learning-related attack in smart home

| Angles                        | Description                                      | Attack Vectors                                                                 | References |
|-------------------------------|--------------------------------------------------|-------------------------------------------------------------------------------|------------|
| Exploiting vulnerabilities of Learning | Automatic vehicle interference                   | Tampering with the image transmitted to the automatic vehicle image recognition algorithm | [37]       |
|                               | Controlling voice control system                  | Designing an ultrasound that contain voice control commands, but humans could not hear | [55]       |
|                               | Intrusion detection systems evasion               | Disguising traffic pattern of the malicious data                              | [14]       |
| Using Learning-Based Techniques | Attack cryptographic algorithm                    | Learning-based analysis of power traces to find secret key information         | [20], [25] |
|                               | PUF attack                                       | Learning-based modeling methods; Combining side-channel information with machine learning modeling techniques | [27], [31], [42], [49] |
|                               | Stealing information from cache                   | Building cache pattern classifier to extract information                      | [58]       |
|                               | Recovering printed text                          | Analyzing voice of printer via machine learning                              | [5]        |

information, the accuracy of the recovered text can be higher. The learning-algorithm-related works are summarized in summarized in Table 2.

5. Outlook. In this section, we explore the potential solutions that may address the aforementioned security problems.

5.1. Privacy preserving. We will discuss two kinds of approaches to protect privacy data against learning-based analysis.

Padding-Based Approaches. One solution to solve the side-channel attacks mentioned above is to add data redundancy to mislead the statistical-based privacy (e.g., living conditions, daily routines, habits, etc.) or secret (e.g., keys, passwords, etc.) inferences. An intuitive way is to fill all the packets to the same size, but this may result in additional communications and processing overhead. Random padding is suggested to improve the efficiency of redundancy insertion based solutions, which appends a random length of padding to a packet within a given interval to strike a balance between overhead and protecting user privacy. Cai et al. [8] proposed a data-sanitization method to prevent privacy information inference attacks in social applications. Fogla et al. [52] aimed at the privacy protection of web applications and designed a scheme to prevent web traffic analysis combining the techniques of Privacy Preserving Data Publishing (PPDP) with random padding.

Differential Privacy-Based Approaches. Differential privacy guarantees that the attacker can learn virtually nothing more about an individual than they would learn if that person’s records were absent from the dataset. Because their sensitive personal information is almost completely unrelated to the output of the system, users can be confident that organizations that process their data do not infringe their privacy.

Differential privacy algorithms are randomized algorithms adding noise at key points. It achieves a high efficiency in preventing privacy information from inference attacks. In June 2016, Apple announced it will begin collecting behavioral
statistics from the iPhone using the differential privacy algorithm [16]. Google has implemented a feature in Chrome that collects behavioral statistics from Google Chrome using a differential privacy random response algorithm. In a random response, random noise is added to the statistics before being submitted to the collector. For example, if the actual statistics is 0, the browser will replace 0 with a random choice of 0 or 1 with some probability.

5.2. Robust-learning-xsbased authentication mechanism. We discuss three aspects to improve the robustness of the learning algorithms employed in a smart home system, and consequently, secure the learning-based authentication mechanism.

Disinformation. The purpose of this method is to provide false information and make it impossible for an adversary to correctly estimate the learner’s status. As a result, the adversary cannot correctly guess the learner’s decision-making boundaries.

Randomization. Some of the exploratory attacks are only concerned with the classification of a small set of points. This attack is very sensitive to the decision boundary and the small changes in the decision boundary change the classification of a small set of points. This suggests that adding some randomization near the decision boundary may make it harder for the adversary to perform attacks.

Robustness evaluation. The ability of different algorithms to resist malicious training data is different, that is, the robustness is different. The measure of robustness can provide a reference for the researcher to improve the learning algorithm. We can use the performance of different learning algorithms against different attacks to represent the robustness. For example, if the training data is slightly modified while the performance of the learning algorithm will be greatly affected, indicating that the algorithm is vulnerable to attack. However, this method has a drawback: when the adversary found that the attack effect is not obvious, he will improve the intensity of the attack, resulting in the arms race between adversaries and learning algorithm designers.

6. Conclusion. To enhance the functionality and the security, machine learning algorithms play an important role in smart home systems. Meanwhile, attackers may also utilize learning algorithms as a tool or exploit learning mechanisms’ vulnerabilities to hack smart home systems. In this paper, we unify the smart home system architecture and investigate the application of learning algorithms in smart home IoT system security. We discussed the functionality and security enhancing methods based on learning mechanisms. We also described the security threats exposed by employing learning techniques. We explored the potential solutions to address the aforementioned security problems.

REFERENCES
[1] Researchers exploit zigbee security flaws that compromise security of smart homes, http://www.networkworld.com/article/2969402/microsoft-subnet/researchers-exploit-zigbee-securityflaws-that-compromise-security-of-smart-homes.html.
[2] M. R. Alam, M. B. I. Reaz and M. A. M. Ali, A review of smart homes–past, present, and future, IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 42 (2012), 1190–1203.
[3] I. Androustopoulos, J. Koutsias, K. V. Chandrinos, G. Paliouras and C. D. Spyropoulos, An evaluation of naive bayesian anti-spam filtering, arXiv preprint cs/0006013.
[4] Apple, Apple machine learning journal, https://machinelearning.apple.com/.
[5] M. Backes, M. Diemuth, S. Gerling, M. Pinkal and C. Sporleder, Acoustic side-channel attacks on printers, in *Proceedings of USENIX Security symposium*, (2010), 307–322.

[6] A. Bassi and G. Horn, Internet of things in 2020: A roadmap for the future, *European Commission: Information Society and Media*, 22 (2008), 97–114.

[7] V. H. Bhide and S. Wagh, 1-learning iot: An intelligent self learning system for home automation using IoT, in *Proceedings of 2015 International Conference on Communications and Signal Processing (ICCCSP)*, IEEE, 2015, 1763–1767.

[8] Z. Cai, Z. He, X. Guan and Y. Li, Collective data-sanitization for preventing sensitive information inference attacks in social networks, *IEEE Transactions on Dependable and Secure Computing*, PP (2017), 1–1.

[9] S. Chari, J. R. Rao and P. Rohatgi, Template attacks, in *Proceedings of International Workshop on Cryptographic Hardware and Embedded Systems*, Springer, 2002, 13–28.

[10] J. Choi, D. Shin and D. Shin, Research and implementation of the context-aware middleware for controlling home appliances, *IEEE Transactions on Consumer Electronics*, 51 (2005), 301–306.

[11] L. Deng and X. Li, Machine learning paradigms for speech recognition: An overview, *IEEE Transactions on Audio, Speech, and Language Processing*, 21 (2013), 1060–1089.

[12] E. Fernandes, J. Jung and A. Prakash, Security analysis of emerging smart home applications, in *Proceedings of 2016 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2016, 636–654.

[13] E. Fernandes, A. Rahmati, K. Eykholt and A. Prakash, Internet of things security research: A rehash of old ideas or new intellectual challenges?, *arXiv preprint, arXiv:1705.08522*.

[14] P. Fogla and W. Lee, Evading network anomaly detection systems: Formal reasoning and practical techniques, in *Proceedings of the 13th ACM conference on Computer and communications security*, ACM, 2006, 59–68.

[15] Gartner, Gartner says 6.4 billion connected “things” will be in use in 2016, up 30 percent from 2015, *http://www.gartner.com/newsroom/id/3165317*.

[16] A. Greenberg, Apple’s ‘differential privacy’ is about collecting your data—but not your data, *https://www.wired.com/2016/06/apples-differential-privacy-collecting-data/*.

[17] K. He, X. Zhang, S. Ren and J. Sun, Deep residual learning for image recognition, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, 770–778.

[18] A. Hesseldahl, A hackers-eye view of the internet of things, *http://recode.net/2015/04/07/a-hackers-eye-view-of-the-internet-of-things/*.

[19] G. Ho, D. Leung, P. Mishra, A. Hosseini, D. Song and D. Wagner, Smart locks: Lessons for securing commodity internet of things devices, in *Proceedings of the 11th ACM on Asia Conference on Computer and Communications Security*, ACM, 2016, 461–472.

[20] G. Hospodar, B. Gierlich, E. De Mulder, I. Verbauwhede and J. Vandewalle, Machine learning in side-channel analysis: A first study, *Journal of Cryptographic Engineering*, 1 (2011), 293–302.

[21] D. Istrate, M. Vacher, E. Castelli and C.-P. Nguyen, Sound processing for health smart home, in *Proceedings of International Conference on Smart homes and health Informatics*, 2004, 41–48.

[22] Y. J. Jia, Q. A. Chen, S. Wang, A. Rahmati, E. Fernandes, Z. M. Mao, A. Prakash and S. J. University, Contexti: Towards providing contextual integrity to appified iot platforms, in *Proceedings of the 21st Network and Distributed System Security Symposium (NDSS’17)*, 2017.

[23] E. B. Karbab, M. Debbabi, A. Derhab and D. Mouheb, Cypider: Building community-based cyber-defense infrastructure for android malware detection, in *Proceedings of the 32nd Annual Conference on Computer Security Applications*, ACM, 2016, 348–362.

[24] G. Kortuem, F. Kawser, V. Sundramoorthy and D. Fitton, Smart objects as building blocks for the internet of things, *IEEE Internet Computing*, 14 (2010), 44–51.

[25] L. Lerman, G. Bontempi and O. Markowitch, Side channel attack: An approach based on machine learning, *Center for Advanced Security Research Darmstadt*, 29–41.

[26] Y. Liang, Z. Cai, Q. Han and Y. Li, Location privacy leakage through sensory data, *Security and Communication Networks*, 2017 (2017), 12pp.

[27] D. Lim, J. W. Lee, B. Gassend, G. E. Suh, M. Van Dijk and S. Devadas, Extracting secret keys from integrated circuits, *IEEE Transactions on Very Large Scale Integration (VLSI) Systems*, 13 (2005), 1200–1205.
[28] C. Liu, Y. Cao, Y. Luo, G. Chen, V. Vokkarane and Y. Ma, Deepfood: Deep learning-based food image recognition for computer-aided dietary assessment, in *Proceedings of International Conference on Smart Homes and Health Telematics*, Springer, 2016, 37–48.

[29] C. Livadas, R. Walsh, D. Lapsley and W. T. Strayer, Using machine learning techniques to identify botnet traffic, in *Proceedings of 2006 31st IEEE Conference on Local Computer Networks*, IEEE, 2006, 967–974.

[30] R. Lutolf, Smart home concept and the integration of energy meters into a home based system, in *Proceedings of Seventh International Conference on Metering Apparatus and Tariffs for Electricity Supply*, IET, 1992, 277–278.

[31] A. Mahmoud, U. Rührmair, M. Majzoobi and F. Koushanfar, Combined modeling and side channel attacks on strong puf's, *IACR Cryptology ePrint Archive*, 2013 (2013), p632.

[32] S. Majundar, Y. Jarrraya, M. Ogaily, A. Alimohammadifar, M. Pourzandi, L. Wang and M. Debbabi, Leaps: Learning-based proactive security auditing for clouds, in *Proceedings of European Symposium on Research in Computer Security*, Springer, 2017, 265–285.

[33] R. McCoppin and M. Rizki, Deep learning for image classification, in *SPIE Defense + Security*, International Society for Optics and Photonics, 2014, 90790T–90790T.

[34] E. C. McLaughlin, Alexa, what other devices are listening to me?, http://edition.cnn.com/2017/01/12/tech/voice-technology-internet-of-things-privacy/index.html.

[35] T. Oluwafemi, T. Kohno, S. Gupta and S. Patel, Experimental security analyses of non-networked compact fluorescent lamps: A case study of home automation security, in *Proceedings of LASER*, 2013, 13–24.

[36] L. Pang, I. Tchoudovski, A. Bolz, M. Braecklein, K. Egorouchkina and W. Kellermann, Real time heart ischemia detection in the smart home care system, in *Proceedings of 27th Annual International Conference of the Engineering in Medicine and Biology Society (IEEE-EMBS 2005)*, IEEE, 2006, 3703–3706.

[37] N. Papernot, P. McDaniel, A. Sinha and M. Wellman, Towards the science of security and privacy in machine learning, arXiv preprint, arXiv:1611.03814.

[38] O. M. Parkhi, A. Vedaldi and A. Zisserman, Deep face recognition, in *Proceedings of British machine vision conference (BMVC)*, 1 (2015), 6pp.

[39] A. Rahmati, E. Fernandes, K. Eykholt, X. Chen and A. Prakash, Heimdall: A privacy-respecting implicit preference collection framework, in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, ACM, 2017, 453–463.

[40] B. Rashidi, C. Fung, A. Nguyen and T. Vu, Android permission recommendation using transitive bayesian inference model, in *Proceedings of European Symposium on Research in Computer Security*, Springer, 2016, 477–497.

[41] S. Roy, J. DeLoach, Y. Li, N. Herndon, D. Caragea, X. Ou, V. P. Ranganath, H. Li and N. Guevara, Experimental study with real-world data for android app security analysis using machine learning, in *Proceedings of the 31st Annual Computer Security Applications Conference*, ACM, 2015, 81–90.

[42] U. Rührmair, F. Sehnke, J. Sölter, G. Dror, S. Devadas and J. Schmidhuber, Modeling attacks on physical unclonable functions, in *Proceedings of the 17th ACM conference on Computer and communications security*, ACM, 2010, 237–249.

[43] U. Rührmair, J. Sölter, F. Sehnke, X. Xu, A. Mahmoud, V. Stoyanova, G. Dror, J. Schmidhuber, W. Burleson and S. Devadas, Puf modeling attacks on simulated and silicon data, *IEEE Transactions on Information Forensics and Security*, 8 (2013), 1876–1891.

[44] M. S. Shahriar and M. S. Rahman, Urban sensing and smart home energy optimisations: A machine learning approach, in *Proceedings of the 2015 International Workshop on Internet of Things towards Applications*, ACM, 2015, 19–22.

[45] F. Siegemund, A context-aware communication platform for smart objects, *Lecture notes in computer science*, (2004), 69–86.

[46] K. Simonyan and A. Zisserman, Very deep convolutional networks for large-scale image recognition, arXiv:1409.1556.

[47] R. Sommer and V. Paxson, Outside the closed world: On using machine learning for network intrusion detection, in *Proceedings of 2010 IEEE Symposium on Security and Privacy (SP)*, IEEE, 2010, 305–316.

[48] I. Strategy and P. Unit, Itu internet reports 2005: The internet of things, *Geneva: International Telecommunication Union (ITU)*.

[49] A. van Berlo, A. Bob, E. Jan, F. Klaus, H. Maik and W. Charles, *Design Guidelines on Smart Homes*, A COST 219bis Guidebook (October 1999).
[50] D. Ventura, D. Casado-Mansilla, J. López-de Armentia, P. Garraizar, D. López-de Ipina and V. Catania, Ariima: A real iot implementation of a machine-learning architecture for reducing energy consumption, in Proceedings of International Conference on Ubiquitous Computing and Ambient Intelligence, Springer, 2014, 444–451.

[51] Y. Wang, Z. Cai, G. Yin, Y. Gao, X. Tong and G. Wu, An incentive mechanism with privacy protection in mobile crowdsourcing systems, Computer Networks, 102 (2016), 157–171.

[52] S. Wüller, U. Meyer, F. Förö and S. Wetzel, Privacy-preserving conditional random selection, in Proceedings of 2015 13th Annual Conference on Privacy, Security and Trust (PST), IEEE, 2015, 44–53.

[53] T. Xu, J. B. Wendt and M. Potkonjak, Security of iot systems: Design challenges and opportunities, in Proceedings of the 2014 IEEE/ACM International Conference on Computer-Aided Design, IEEE Press, 2014, 417–423.

[54] T. Yu, V. Sekar, S. Seshan, Y. Agarwal and C. Xu, Handling a trillion (unfixable) flaws on a billion devices: Rethinking network security for the internet-of-things, in Proceedings of the 14th ACM Workshop on Hot Topics in Networks, ACM, 2015, Article No. 5.

[55] G. Zhang, C. Yan, X. Ji, T. Zhang, T. Zhang and W. Xu, Dolphinattack: Inaudible voice commands, arXiv:1708.09537.

[56] L. Zhang, Z. Cai and X. Wang, Fakemask: A novel privacy preserving approach for smartphones, IEEE Transactions on Network and Service Management, 13 (2016), 335–348.

[57] Q. Zhang and B. Li, Discriminative k-svd for dictionary learning in face recognition, in Proceedings of 2010 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 2010, 2691–2698.

[58] Y. Zhang, A. Juels, M. K. Reiter and T. Ristenpart, Cross-vm side channels and their use to extract private keys, in Proceedings of the 2012 ACM conference on Computer and communications security, ACM, 2012, 305–316.

[59] X. Zheng, Z. Cai, J. Yu, C. Wang and Y. Li, Follow but no track: Privacy preserved profile publishing in cyber-physical social systems, IEEE Internet of Things Journal, 4 (2017), 1868–1878.

[60] H. Zhou, K. M. Hou and D.-C. Zuo, Real-time automatic ecg diagnosis method dedicated to pervasive cardiac care, Wireless Sensor Network, 1 (2009), 276–283.

Received October 2017; revised December 2017.

E-mail address: maojian@buaa.edu.cn
E-mail address: lqxiao529@buaa.edu.cn
E-mail address: bianjd@buaa.edu.cn