Adversarial Machine Learning based Partial-model Attack in IoT

Zhengping Luo
University of South Florida.
Email: zhengpingluo@usf.edu.

Shangqing Zhao
University of South Florida.
Email: shangqingzhao@usf.edu.

Zhuo Lu
University of South Florida.
Email: zhuolu@usf.edu.

Yalin E. Sagduyu
Intelligent Automation Inc.
Email: ysgduyu@i-a-i.com.

Jie Xu
University of Miami.
Email: jiexu@miami.edu.

ABSTRACT

As Internet of Things (IoT) has emerged as the next logical stage of the Internet, it has become imperative to understand the vulnerabilities of the IoT systems when supporting diverse applications. Because machine learning has been applied in many IoT systems, the security implications of machine learning need to be studied following an adversarial machine learning approach. In this paper, we propose an adversarial machine learning based partial-model attack in the data fusion/aggregation process of IoT by only controlling a small part of the sensing devices. Our numerical results demonstrate the feasibility of this attack to disrupt the decision making in data fusion with limited control of IoT devices, e.g., the attack success rate reaches 83% when the adversary tampers with only 8 out of 20 IoT devices. These results show that the machine learning engine of IoT system is highly vulnerable to attacks even when the adversary manipulates a small portion of IoT devices, and the outcome of these attacks severely disrupts IoT system operations.

CCS CONCEPTS

• **Security and privacy → Mobile and wireless security**; • **Networks → Network reliability**.

KEYWORDS

Internet of Things, wireless security, machine learning, adversarial machine learning, data fusion

ACM Reference Format:
Zhengping Luo, Shangqing Zhao, Zhuo Lu, Yalin E. Sagduyu, and Jie Xu. 2020. Adversarial Machine Learning based Partial-model Attack in IoT. In *2nd ACM Workshop on Wireless Security and Machine Learning (WiseML '20), July 13, 2020, Linz (Virtual Event), Austria.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3395352.3402619

1 INTRODUCTION

Internet of Things (IoT) is cast as a system of networked devices embedded with sensors to gather and interchange data, and execute complex tasks [1–3]. As technology is advancing from web2 (social networking web) to web3 (ubiquitous computing web), IoT, as an extension of Internet into the physical world, becomes the core technology to connect sensors and actuators into an integrated network [4–6]. Applications of IoT include but are not limited to smart home, smart warehouse, vehicular networks, environmental monitoring, and perimeter security [3].

Wireless Sensor Network (WSN) is often considered as the building block for the IoT systems. However, the sensing devices in IoT are prone to failures [2]. While information exchange among heterogeneous sensing devices/actuators and hubs/data centers is mandatory, many applications of IoT have strict timing, security, reliability requirements. Therefore, how to ensure the real information sensed by sensors/actuators to be securely received by the hub/data center in a wireless environment is critical for both the security and reliability of the IoT systems.

As the scale of IoT systems grows rapidly with more devices added, *machine learning* has started playing a key role in the processing and learning from large-scale data generated by IoT devices [7]. While machine learning helps with efficient operation of IoT systems, the other side of the coin is concerning adversaries may also employ machine learning as powerful means to launch attacks against IoT infrastructures. The study of machine learning under adversaries is referred to as *adversarial machine learning* [8].

In this paper, we focus on the security of *data fusion/aggregation* process in IoT. Original data or information is collected through IoT devices such as actuators, RFID, switches, and sensors. Then multiple IoT devices report their data to a hub or data center to aggregate and report the aggregated results to the cloud or data analysis center. In this paper, we consider a scenario where multiple devices report their data to a fusion center, and the fusion center makes a binary decision based on the received information. In this scenario, we show that the adversary can employ machine learning techniques to launch an attack by controlling only a small part of the devices, which we call the *partial-model attack*.

The main contributions of this paper are listed as: (i) We introduce a machine learning based partial-model attack in IoT data fusion process, where the adversary aims to disrupt decision making of IoT data fusion process by taking advantage of the IoT device properties; (ii) We present numerical experiments to validate the proposed attack framework and demonstrate that the successful attack ratio is high even when a small portion of sensors are controlled by the adversary. For instance, in a scenario where 8 out of 20 devices are controlled by an adversary, the hit ratio reaches up to 83%; (iii) We discuss potential ways of defending against machine learning based partial-model attack in IoT systems.

WiseML '20, July 13, 2020, Linz (Virtual Event), Austria

ACM ISBN 978-1-4503-8007-2/20/07…$15.00
https://doi.org/10.1145/3395352.3402619

2 BACKGROUND AND RELATED WORK

2.1 IoT architecture

IoT finds rich applications including but not limited to Industrial Internet of Things (IIoT), smart home, smart city, healthcare, and transportation. Basic components that enable the IoT benefits include: (i) *hardware* (heterogeneous sensors and actuators); (ii) *middleware* (data aggregation/fusion center and storage devices); (iii) *presentation* (visualization and other analysis tools that enable access to different platforms) [2, 9, 10].

Radio Frequency Identification (RFID) is a major source of data for many IoT systems [11, 12]. It is a technology that employs electromagnetic fields for data transfer and automatic object detection. With the RFID tag, items can be detected by reading their labels. Once sensor data is collected, the next step is to transfer data for storage and processing. Cloud computing is the storage and computing center of IoT, where data analytics are based on. Users can access the cloud computing and generate visual presentation of the collected data [13]. Moreover, cloud platforms provide device lifecycle management for IoT and can provide digital twin version of real systems [14].

We illustrate the general architecture of IoT systems in Fig 1 as three layers: *IoT things* (physical devices), *IoT network*, and *IoT cloud and application*. The bottom layer consists of the physical devices, the second layer focuses on the infrastructure such as network and data aggregation, the last (highest) layer is the user-oriented layer.

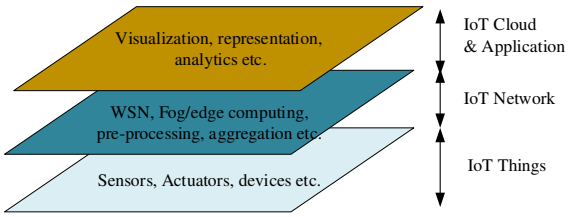

Figure 1: The generalized layers of typical IoT systems.

2.2 Adversarial machine learning and IoT

Machine learning, especially *deep learning*, has attracted tremendous attention since its successful application in image recognition [15]. Recently, deep learning has started finding applications also in wireless systems, including waveform design, signal analysis and security [16]. Devices within an IoT system often generate data continuously and simultaneously at high rates. To deal with this large-scale data, machine learning offers automated means to process and analyze data, and make decisions [2]. Unlike traditional statistical models, machine learning provides a way to learn parameters from the data, and it can make decisions based on both historical data and real time streaming data. Besides, given the heterogeneity of IoT systems, machine learning can be performed in either central or distributed fashion.

Despite its strengths, machine learning itself has many vulnerabilities that might be exploited by malicious users [17, 18]. The security problem of IoT systems are critical for both the users and owners of IoT infrastructure. Machine learning in the presence of adversaries is studied under the emerging area of *adversarial machine learning* [8, 17–19, 22]. The shared nature of wireless medium makes machine learning especially vulnerable to various attacks built upon adversarial machine learning. In wireless domain, adversarial machine learning has been applied to launch different types of attacks [21], including inference (exploratory) attack [20, 23], evasion attack [24–27], poisoning (causative) attack [27–29], Trojan attack [30], and spoofing attack [31]. These attacks are stealthier (more difficult to detect) and operate with lower footprint compared with conventional wireless attacks such as a jamming [32]. Adversarial learning can also be used to augment training data with synthetic data samples [33].

Security of IoT has drawn increasing attention. Many of the security studies of IoT are centered around two fronts [34]: sensing end-devices and connecting protocols. Strategies such as improving security through firewall and mobility policies have been presented in [35]. Intrusion detection mechanisms formulated as anomaly detection have been discussed for IoT systems in [36]. The privacy issues of IoT have been considered in [37].

Major security vulnerabilities and challenges of IoT can be summarized as follows [14]:

- *Sub-system heterogeneity*: Devices, sensors, actuators and sub-controlling components within IoT systems are heterogeneous. Thus, it is challenging to integrate them into one system, and security measures required for different subsystems might also differ from each other. It is necessary to find a common strategy to control all the heterogeneous sub-systems.
- *End-device reliability*: End-devices of IoT systems are distributed in real world and they can be influenced by various environmental factors that may cause them to fail to function, report wrong sensing results, or even lose control to malicious users. How to ensure the reliability of the end-devices is critical to the security of IoT.
- *Data security*: Data security in IoT systems involves multiple concerns such as safe transmission (how to guard the sensed information such that it can be safely transferred to the cloud/processing center) and safe operation of data center.

Our paper aims to employ adversarial machine learning techniques to launch attacks against the *IoT data fusion process*. The adversary first learns a machine learning model and then based on the learned model it crafts adversarial inputs. Detailed information on this attack is given in the next section.

3 ADVERSARIAL MACHINE LEARNING-BASED PARTIAL-MODEL ATTACK

3.1 IoT data fusion

Data gathering from multiple devices is a critical step in IoT systems. Compared with making decisions from a single data source, collecting data from multiple sensing devices may help filter out noises and other deviations [38]. The requirements of IoT data fusion are summarized as follows [39]: (i) *Context-aware*: It is necessary to support adaptive and flexible services. The context information like location, weather and other environmental factors may change.

Thus, data fusion at IoT systems needs to calibrate and adapt to these changes. (ii) *Privacy preserving*: It is necessary to protect privacy of sensitive IoT information such as personal habits or industrial secrets. (iii) *Reliable*: It is necessary to detect, remove or replace unreliable devices, as the sensed information may be noisy or contaminated. (iv) *Real time*: It is necessary to make timely decisions to support real-time operations of data fusion. (v) *Verifiable*: It is necessary to keep the fusion result verifiable to the user or public.

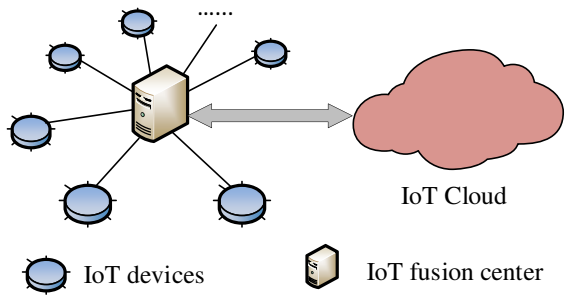

Figure 2: The simplified sensing and decision model of IoT.

In this paper, we consider a general scenario of IoT data fusion as shown in Fig.2. Multiple IoT devices report their collected information to a data fusion center, and the data fusion center makes a decision (classification or regression), and then transfers the decision output to the IoT cloud for further analysis.

3.2 Partial-model attack

The sensing devices play a key role in IoT systems. Therefore, secure and efficient communication between sensing devices and the data fusion center is needed to support IoT operations. However, sensing devices are prone to failures and manipulation by adversaries. We propose an adversarial machine learning based attack model as follows. We assume that the adversary controls a small number of IoT devices, and the adversary knows the decision output of the IoT fusion center. However, the adversary has no knowledge about the decision process in the IoT fusion center. Data is exchanged between the IoT devices fusion center and other edge computing center over wireless channels. Thus, it is possible for the adversary to hijack the over-the-air transferred information.

We assume that there are n IoT sensing devices that report their information to the IoT data fusion center to aggregate and output a decision. In the meantime, m of these devices are controlled by a malicious adversary. By controlling a small portion of IoT devices based on machine learning techniques, the adversary aims to take advantage of the failures of other normal/un-manipulated sensing nodes, further flip or change the decision output significantly. One application is spectrum sensing data falsification by some rogue nodes in cooperative spectrum sensing [29, 40]. As IoT devices may fail or report confused information, the adversary can detect this kind of uncertainty and further take advantage of this uncertainty to expand the impact of the attack.

For a successful attack, the adversary first needs to learn about the potential true decision/classification output in the IoT fusion center from controlled IoT sensing devices and historical decision/classification output. Therefore, the adversary needs to build first a machine learning model to infer the potential decision state based on the information collected from controlled IoT sensing devices. For the attack model, the inputs come from the sensing results of controlled devices. The output are the adversarial vectors crafted from the inputs.

The next question is when to launch the attack. In our proposed partial-model attack, the adversary does not launch the attack at each round of inputs. The attack should be launched when controlled devices sensed possible confused signals, i.e., when the learned decision model by the adversary is less certain about the decision. Consider a convolutional neural network (shown in Fig.4), in which feature vectors are served as inputs of the framework. The next layers are the convolutional layers, which consist of convolutional and pooling operations. The main objective of the convolutional layers is to extract more complicated features (e.g., silhouettes in image recognition). The fully connected layers follow the convolutional layers and aim to find the optimal combination of the previous features. The last layer of the framework is named as SoftMax layer. The number of neurons in the SoftMax layer is equal to the output classes.

The value in SoftMax layer is considered as "confidence value" and reflects how "confident" the trained model is towards the output [41]. Therefore, we employ the largest value of the neuron output in the SoftMax layer, which is also the value of the final decision output for the model, as an indication about the certainty of des. The output value of the SoftMax layer is the result of a squashing function, which limits the output within the range between 0 and 1. Mathematically the standard SoftMax function is defined as: $\sigma(z_i) = \frac{e^{z_i}}{\sum_{j=1}^{c} e^{z_j}}$, in which $\sigma(z_i)$ is the output, or confidence value of the final decision towards the i_{th} class. c is the total number of output classes. Each output value in the SoftMax layer gives the "confidence" of the decision output towards each class [41]. When the confidence value is beyond a certain threshold, malicious inputs are generated and sent to the IoT data fusion center, otherwise normal data are sent to avoid being detected.

After the learning step, the adversary infers the potential true decision output. Then through manipulating the information of controlled IoT devices sent to the IoT data fusion center, the adversary has the possibility to compromise the IoT data fusion center. There are different ways to craft malicious inputs (a.k.a. adversarial inputs) for the controlled IoT sensing devices as shown in[17, 18]. The basic idea of crafting adversarial inputs is to move the input towards the decision boundary of the learned classification model such that the modification is minimized.

The overall attack framework is shown in Fig.3. Normal IoT devices report the collected data directly to the IoT fusion center, while manipulated IoT devices need to report the collected information to the adversary. The adversary first learns an attack model and then in later rounds decides whether to launch attack, or not. When the adversary launches the attack, it reports the manipulated inputs to the IoT fusion center, otherwise it reports the original data. The proposed partial-model attack is not a mathematically guaranteed attack. The key for this attack to succeed depends on

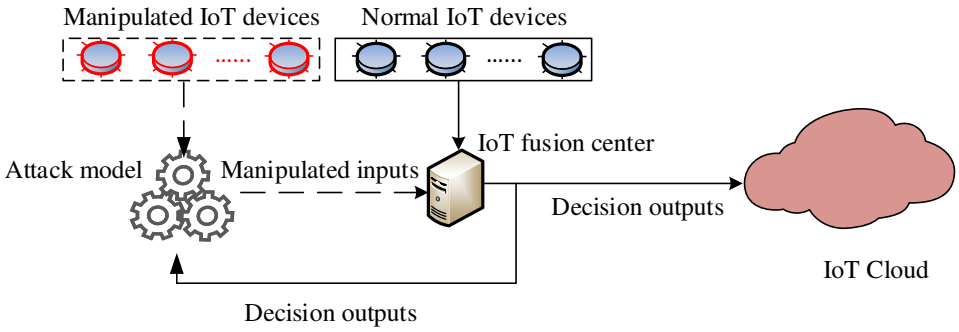

Figure 3: The machine learning based partial-model attack in IoT.

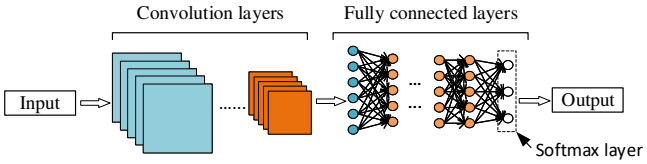

Figure 4: A typical convolutional neural network.

the overall uncertainty among IoT devices. In the next section, we present simulation results to evaluate the proposed attack.

4 EXPERIMENTATION AND ANALYSIS

We conduct detailed simulations of the IoT data fusion process and analyze the performance of the partial-model attack in this section.

4.1 Experimental configurations

The IoT fusion center in our attack model collects information from a set of sensing devices, aggregates them to make a decision and delivers the decision to the IoT cloud for further data analysis. In our simulation, the decision model in the IoT data fusion center is set as a binary classification model as many sensing tasks are binary, such as switches and signal sensing devices. We assume that there are 20 IoT sensing devices such as RFID. 10000 data samples are collected from two Gaussian distributions, each of the distribution corresponds to one class. The first 2000 data samples serve as training data for the adversary. The remaining data samples are used for evaluation. To make the simulation consistent with real environment uncertainties, the mean and deviation of the Gaussian distribution for each device are set as a random number within a given range (e.g., to represent the potential differences in oscillators when spectrum data is sensed). The adversary employs a 5-layer neural network as the learning model. The implementation of the learning model is based on TensorFlow.

In the IoT data fusion center, we employ Support Vector Machine (SVM) as the fusion rule, which is one of the most popular statistical classification models. It is worth noting that other fusion rules such as multi-layer perceptron neural network, decision tree, etc. can also be used as the fusion rule. Due to the limitation of space, we consider SVM in this paper.

4.2 Attack performance analysis

The performance of the proposed machine learning based partial-model attack is measured by hit ratio (namely, attack success ratio), which is defined as:

$$\text{Hit ratio} = \frac{\text{The number of successful attacks}}{\text{Total number of attack instances}}.$$

The successful attack here corresponds to those input samples that successfully flip the decision of IoT fusion center that would have been made when no attack is launched.

We first evaluate the hit ratio by varying the number of controlled devices m. The results under different confidence threshold value of 0.60,0.75 and 0.9, respectively, are shown in Fig.5. The hit ratio increases with m. In particular, when m is 8, which is less than half of the total number of devices, n, the hit ratio is 72% while the confidence threshold value is 0.75. As m further increases, the hit ratio approaches to 1. When the threshold is too large, the number of attacks will decrease dramatically, thus the overall hit ratio will also decrease.

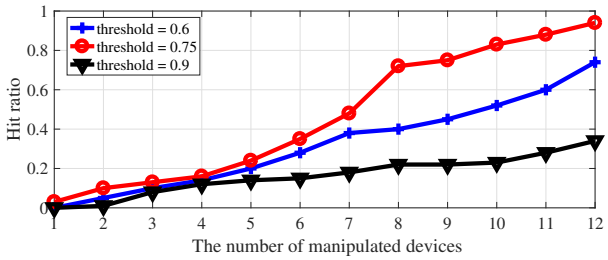

Figure 5: The relationship of the number of controlled sensing devices with the hit ratio.

Next, we evaluate the relationship between the hit ratio and the confidence threshold. The results are shown in Table 6. Four different scenarios when m is 6, 8, 10, or 12 are considered. We observe that when the confidence threshold is set to near 0.5 (the confidence threshold is always larger than 0.5 due to our binary model), the attack success ratio is comparatively lower than the case when the threshold is set around 0.7. The reason is that when the learned model is less certain about the decision output, it has higher

probability that the learned model makes mistakes in inferring the potential true decision in the IoT data fusion center.

When the threshold is set as 0.7, the hit ratio approaches to 83% when $m = 8$. The hit ratio decreases dramatically when the confidence value increases beyond 0.8. The reason is that when the threshold is set too high, the number of attacks increases and thus it becomes difficult for the adversary to take advantage of the uncertainty of other normal IoT sensing devices.

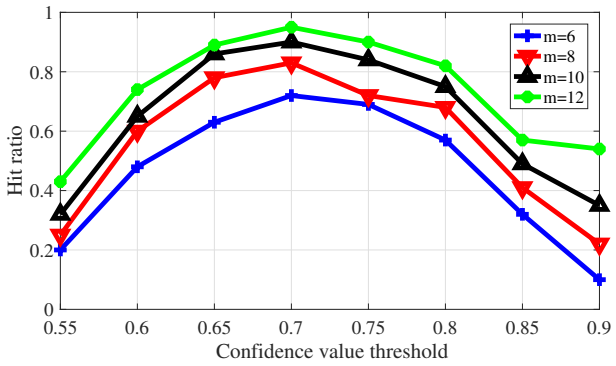

Figure 6: Hit ratios under different confidence thresholds and different number of manipulated devices.

Our simulation results demonstrate that the partial-model attack is likely to succeed when the IoT devices or information exchange involve uncertainties. The adversary takes advantage of the "uncertainty" of other normal devices and increases its hit ratio. Therefore, robust security mechanisms are needed in future IoT system design.

5 DISCUSSION

As machine learning provides IoT systems with powerful means of learning from data and solving complex tasks, it also raises security concerns due to its vulnerability to adversarial manipulation. Our proposed adversarial machine learning based partial-model attack model focuses on the IoT data fusion process and equips the adversary with the capability to launch successful attacks even when the adversary controls a small part of the IoT devices by exploiting the performance uncertainty of the IoT devices or the communication channel. How to counter the proposed attack is our future work. Below, we provide several potential mechanisms for defense: *Deploying robust anomaly detection mechanism in the IoT fusion center.* This is a direct method to defend the IoT systems against the partial-model attack. However, in this attack, all the manipulated devices cooperate with each other to launch the attack. Thus, how to design an anomaly detection method to detect a set of devices is a challenge. *Improving privacy protection in every level of the IoT infrastructure.* In the partial-model attack, the key to learn a partial model to mimic the fusion center is the availability of the output of the fusion center. Thus, the decision information can be kept as private and secure by deploying a privacy protection mechanism.

Using machine learning to attack the IoT systems is detrimental to the IoT security. On the other hand, machine learning can be also employed as a defense method [42]. Therefore, it is important to understand the interaction of machine learning techniques used for attack and defense, and game theory can be used as mathematical means to study the conflict of interest driven by machine learning.

6 CONCLUSION

How to protect the privacy and security of IoT systems from the data collecting stage to the final visualization and application stage is essential to the successful adoption of IoT. In this paper, we introduced an adversarial machine learning based partial-model attack strategy, which mainly sits in the data collecting and aggregating stage of IoT systems. We use the machine learning based model to infer the potential decisions or aggregate results and then launch attacks by manipulating the data of the controlled IoT devices. Simulations show that the attack is highly successful even with a small part of manipulated IoT devices.

Acknowledgement: The work at USF was supported in part by NSF CNS-1717969.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The internet of things: A survey," *Computer networks*, vol. 54, no. 15, pp. 2787–2805, 2010.

[2] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of things (iot): A vision, architectural elements, and future directions," *Future generation computer systems*, vol. 29, no. 7, pp. 1645–1660, 2013.

[3] W. Khan, M. Rehman, H. Zangoti, M. Afzal, N. Armi, and K. Salah, "Industrial internet of things: Recent advances, enabling technologies and open challenges," *Computers & Electrical Engineering*, vol. 81, p. 106522, 2020.

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

[5] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things journal*, vol. 1, no. 1, pp. 22–32, 2014.

[6] Z. B. Celik, E. Fernandes, E. Pauley, G. Tan, and P. McDaniel, "Program analysis of commodity iot applications for security and privacy: Challenges and opportunities," *ACM Computing Surveys (CSUR)*, vol. 52, no. 4, pp. 1–30, 2019.

[7] M. S. Mahdavinejad, M. Rezvan, M. Barekatain, P. Adibi, P. Barnaghi, and A. P. Sheth, "Machine learning for internet of things data analysis: A survey," *Digital Communications and Networks*, vol. 4, no. 3, pp. 161–175, 2018.

[8] Y. Vorobeychik and M. Kantarcioglu, *Adversarial machine learning*. Morgan & Claypool Publishers, 2018.

[9] J. K. D. Barriga, C. D. G. Romero, and J. I. R. Molano, "Proposal of a standard architecture of iot for smart cities," in *International Workshop on Learning Technology for Education Challenges*. Springer, 2016, pp. 77–89.

[10] H. Liu, H. Ning, Y. Zhang, Q. Xiong, and L. T. Yang, "Role-dependent privacy preservation for secure v2g networks in the smart grid," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 2, pp. 208–220, 2013.

[11] A. P. Plageras, K. E. Psannis, B. Gupta, C. Stergiou, B.-G. Kim, and Y. Ishibashi, "Solutions for inter-connectivity and security

in a smart hospital building," in *2017 IEEE 15th International Conference on Industrial Informatics (INDIN)*. IEEE, 2017, pp. 174–179.

[12] C. L. Stergiou, A. P. Plageras, K. E. Psannis, and B. B. Gupta, "Secure machine learning scenario from big data in cloud computing via internet of things network," in *Handbook of Computer Networks and Cyber Security*. Springer, 2020, pp. 525–554.

[13] C. Stergiou and K. E. Psannis, "Recent advances delivered by mobile cloud computing and internet of things for big data applications: a survey," *International Journal of Network Management*, vol. 27, no. 3, p. e1930, 2017.

[14] F. Firouzi, B. Farahani, F. Ye, and M. Barzegari, "Machine learning for iot," in *Intelligent Internet of Things*. Springer, 2020, pp. 243–313.

[15] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[16] T. Erpek, T. J. O'Shea, Y. E. Sagduyu, Y. Shi, and T. C. Clancy, "Deep learning for wireless communications. in development and analysis of deep learning architectures." Springer, 2020, pp. 223–266.

[17] N. Papernot, P. McDaniel, I. Goodfellow, S. Jha, Z. B. Celik, and A. Swami, "Practical black-box attacks against machine learning," in *Proceedings of the 2017 ACM on Asia conference on computer and communications security*, 2017, pp. 506–519.

[18] A. Kurakin, I. Goodfellow, and S. Bengio, "Adversarial examples in the physical world," *arXiv preprint arXiv:1607.02533*, 2016.

[19] Y. Shi, Y. E. Sagduyu, and A. Grushin, "How to steal a machine learning classifier with deep learning," in *IEEE Symposium on Technologies for Homeland Security (HST)*, 2017.

[20] Y. Shi, Y. E. Sagduyu, K. Davaslioglu, and R. Levy, "Vulnerability detection and analysis in adversarial deep learning. in guide to vulnerability analysis for computer networks and systems." Springer, 2018, pp. 211–234.

[21] Y. E. Sagduyu, Y. Shi, T. Erpek, W. Headley, B. Flowers, G. Stantchev, and Z. Lu, "When wireless security meets machine learning: Motivation, challenges, and research directions," 2020, available on arXiv:2001.08883.

[22] Y. Shi, Y. E. Sagduyu, T. Erpek, K. Davaslioglu, Z. Lu, and J. Li, "Adversarial deep learning for cognitive radio security: Jamming attack and defense strategies," in *IEEE ICC 2018 Workshop on Promises and Challenges of Machine Learning in Communication Networks*, 2018.

[23] T. Erpek, Y. E. Sagduyu, and Y. Shi, "Deep learning for launching and mitigating wireless jamming attacks," *IEEE Transactions on Cognitive Communications and Networking*, vol. 5, no. 1, pp. 2–14, March 2019.

[24] B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, and S. Ulukus, "Over-the-air adversarial attacks on deep learning based modulation classifier over wireless channels," in *Conference on Information Sciences and Systems (CISS)*, 2020.

[25] ——, "Channel-aware adversarial attacks against deep learning-based wireless signal classifiers," 2020, available on arXiv:2005.05321.

[26] Y. E. Sagduyu, Y. Shi, and T. Erpek, "IoT network security from the perspective of adversarial deep learning," in *IEEE SECON Workshop on Machine Learning for Communication and Networking in IoT*, 2019.

[27] Y. E. Sagduyu, T. Erpek, and Y. Shi, "Adversarial deep learning for over-the-air spectrum poisoning attacks," *IEEE Transactions on Mobile Computing*, no. 1, pp. 2–14, 2019.

[28] Y. Shi, T. Erpek, Y. E. Sagduyu, and J. Li, "Spectrum data poisoning with adversarial deep learning," in *IEEE Military Communications Conference (MILCOM)*, 2018.

[29] Z. Luo, S. Zhao, Z. Lu, J. Xu, and Y. E. Sagduyu, "When attackers meet AI: Learning-empowered attacks in cooperative spectrum sensing," 2019, available on arXiv:1905.01430.

[30] K. Davaslioglu and Y. E. Sagduyu, "Trojan attacks on wireless signal classification with adversarial machine learning," in *IEEE DySPAN Workshop on Data-Driven Dynamic Spectrum Sharing*, 2019.

[31] Y. Shi, K. Davaslioglu, and Y. E. Sagduyu, "Generative adversarial network for wireless signal spoofing," in *ACM WiSec Workshop on Wireless Security and Machine Learning*, 2019.

[32] Y. E. Sagduyu, R. Berry, and A. Ephremides, "Jamming games in wireless networks with incomplete information," *IEEE Communications Magazine*, vol. 49, no. 8, pp. 112–118, 2008.

[33] K. Davaslioglu and Y. E. Sagduyu, "Generative adversarial learning for spectrum sensing," in *IEEE International Conference on Communications (ICC)*, 2018.

[34] Z. B. Celik, L. Babun, A. K. Sikder, H. Aksu, G. Tan, P. McDaniel, and A. S. Uluagac, "Sensitive information tracking in commodity iot," in *27th USENIX Security Symposium (USENIX Security 18)*, 2018, pp. 1687–1704.

[35] S. Kubler, K. Främling, and A. Buda, "A standardized approach to deal with firewall and mobility policies in the iot," *Pervasive and Mobile Computing*, vol. 20, pp. 100–114, 2015.

[36] B. B. Zarpelão, R. S. Miani, C. T. Kawakani, and S. C. de Alvarenga, "A survey of intrusion detection in internet of things," *Journal of Network and Computer Applications*, vol. 84, pp. 25–37, 2017.

[37] D. Goad, A. Collins, and U. Gal, "Privacy and the internet of things–an experiment in discrete choice," *Information & Management*, p. 103292, 2020.

[38] S. Stillman and I. Essa, "Towards reliable multimodal sensing in aware environments," in *Proceedings of the 2001 workshop on Perceptive user interfaces*, 2001, pp. 1–6.

[39] W. Ding, X. Jing, Z. Yan, and L. T. Yang, "A survey on data fusion in internet of things: Towards secure and privacy-preserving fusion," *Information Fusion*, vol. 51, pp. 129–144, 2019.

[40] Y. E. Sagduyu, "Securing cognitive radio networks with dynamic trust against spectrum sensing data falsification," in *IEEE Military Communications Conference (MILCOM)*, 2014.

[41] M. Fredrikson, S. Jha, and T. Ristenpart, "Model inversion attacks that exploit confidence information and basic countermeasures," in *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 2015, pp. 1322–1333.

[42] R. Doshi, N. Apthorpe, and N. Feamster, "Machine learning ddos detection for consumer internet of things devices," in *2018 IEEE Security and Privacy Workshops (SPW)*. IEEE, 2018, pp. 29–35.