Seminar

By Yaregal T

June 2023

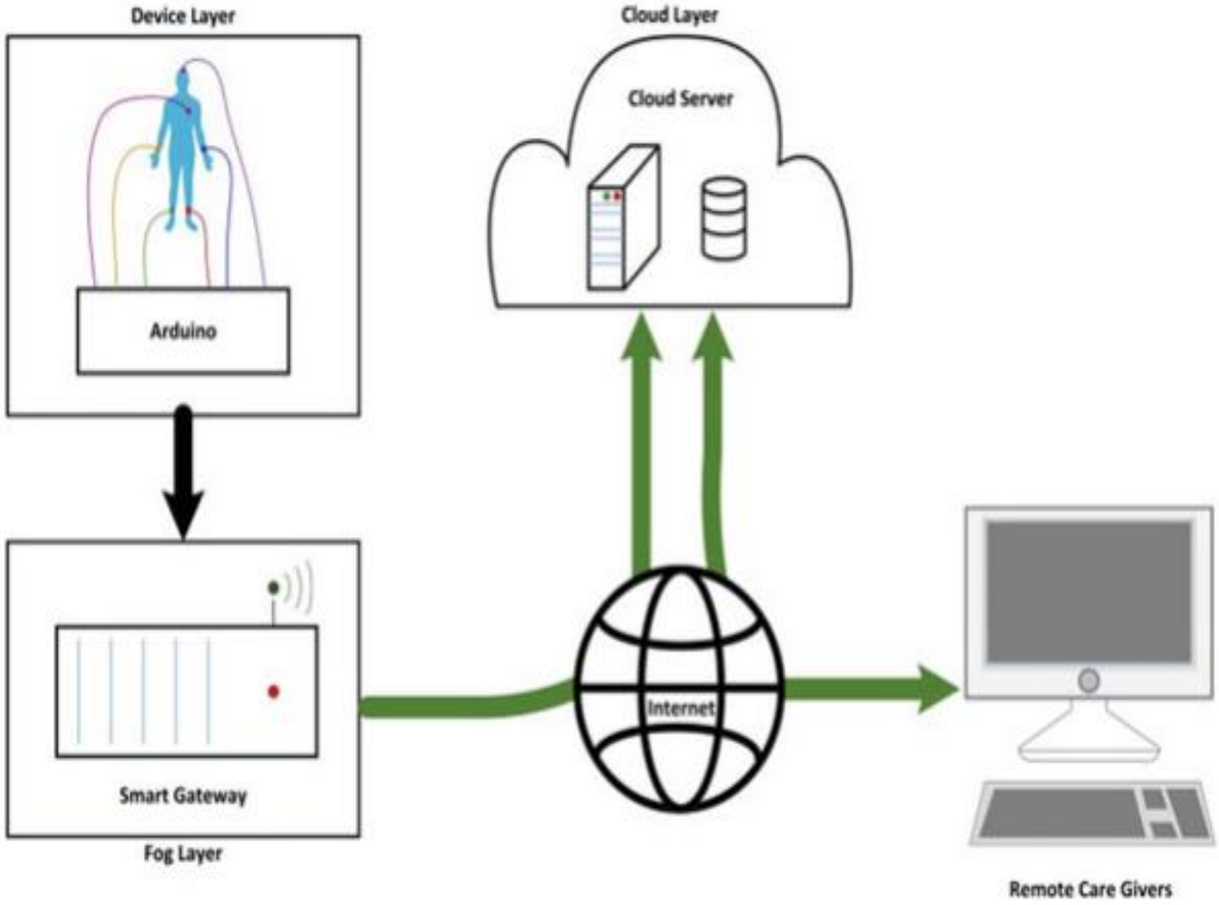
Title

- Survey on IoT Security Analysis Using Lightweight Machine Learning

Introduction

- the Internet of Things (IoT) is "a system of interrelated computing devices, mechanical and digital machines, objects, animals, or people that are provided with unique identifiers and the ability to transfer data over a network without requiring human-to-human or human-to-computer interaction".
- The IoT encompasses a large range of devices ('things'), among which everyday household electronics, such as dishwashers, fridges, smart cameras, smartwatches, smart glasses, smart TVs, and smart light bulbs. Wearable devices can monitor heart rate, steps, and spent calories to name just a few 'smart' features introduced by IoT devices

What is IOT


Device Layer
Arduino
Cloud Layer
Cloud Server
Smart Gateway
Fog Layer
Internet
Remote Care Givers

What is IOT Cont..

- imagine your smart TV is hacked by someone and its record audio in your Salone or bedroom, imagine your IoT connected car is hacked, or imagine an intruder successfully hacked your IoT supported door, from these simple cases you can imagine the consequence that successful penetration of security holes in an IoT supported devices will cause to the society in a variety of aspects.
- to this the security problems are causing a lot of problems not only on the IoT system in which the flaw or the bug is found but also big systems in which this buggy IoT system is communicating.
- As an example, in October 2016, the distributed denial of service attack on Dyn, a company controlling and managing several DNS services, brought down most of America’s Internet, and was caused by an IoT botnet (Mirai).
- Security problems are massively increasing because the amount of linked smart devices constantly grows with their use of different standards, the heterogeneity of the devices, their different implementation way [6]- [8], making the security testing operations daunted.

Problem Statement

- In this rise of the Internet of Things (IoT), where billions of devices are expected to be integrated into horizontal applications [4][5]. Security in IoT software is not second thing to do since the IoT systems are improving our day-to-day life in a way that our day-to-day life depended on it.
- Give the IoT devices limitation such as computation limitation, storage limitation, power limitation most security measures are not effective in IoT
- IoT security is a major concern these days because many IoT devices are not designed with security in mind[6].
- Due to their heterogeneity, distributed, constrained environment, run on different platforms its difficult to capture security flows in IoT [3][11].

Objectives

- **General Objective**
  - Perform a survey on IoT Security Analysis Using LightWight Machine Learning
- **Specific Objective**
  - I summarize and provide a taxonomy of recent work using machine learning to enhance the security property of IoT systems and how machine learning can help to detect security attacks.
  - I identify the weaknesses that still exist in current research and the discrepancies between these weaknesses and the requirements of the IoT setting

Methodology

- In the first stage I will analyze the existing state of the art IoT testing methodology and examine if they give enough emphasis to the security testing of the IoT system to achieve this I will perform a systematic mapping study (SMS).
- Based on the study I will designing a framework that gives higher emphasis to the security of the IoT system which works a long side the testing methodologies so that it will be test everything once.
- To evaluate the performance of the framework by developing an IoT system
- Since our research is motivated to solve security problems in the IOT during the stages of testing by unveiling as many vulnerabilities as possible and the IoT testing methodologies are analyzed using the systematic mapping study (SMS) and a designing of new IoT testing framework which will have the additional capability to discover as many security vulnerabilities as possible during the testing stage in addition to unveiling other requirements of the system for this reason the methodology used for this reason I will use Design Science research approach.

Literature Review

- The IoT is a complete ecosystem that contains a heterogeneous devices and connections, a huge number of users, and a large amount of data. To identify the potential vulnerabilities that exist within an IoT system, it is necessary to look at the whole IoT ecosystem and the behaviors exhibited in the ecosystem
- I grouped the IoT security issues in to three categories:
  - To identify the distinctiveness of each IoT device in the network.
  - To investigate the network behaviors in IoT.

To identify the distinctiveness of each IoT device in the network

- Each and every device in the IoT ecosystem will often have a fixed features such as physical characteristics or services that it provides, power they consume and etc.
- based on this futures it is possible to profile a device to uniquely identify the device from other devices found in the same IoT ecosystem.
- On paper [21], an IoT digital camera have been used to take photographs and record audio/video and could even link with social networking data sources if permitted access. The CCD sensor in the digital camera has a unique sensor pattern noise (SPN) which could be used to create a unique fingerprint of the device.

Cont..

- Such fingerprints for IoT devices could also be identified based on the device users, which can be further analyzed using techniques such as deep learning
- I classified the papers in to identify the distinctiveness of each IoT device in the network to three categories
  - (1), Device Identification Using DL,
  - (2) Service Fingerprint Extraction Using DL,
  - (3) Device Integrity Testing Based on DL

Device Identification Using DL

- Traditional methods of IoT device identification are by using serial numbers, IMEI codes, or other static identifiers
- Deep learning has the potential to identify subtle differences between classes when considering a large feature set to characterize data and therefore could be effective for device identification
- Deep learning methods can extract features from the signal or traffic produced by the device in order to recognize and identify the device.

Cont.

- Work in [11] proposes the method of using deep CNNs to automatically extract features to identify the capture device.
  - They calculate residual noise in the image by subtracting the denoised version of the image from the image provided.
  - The residual noise is then used as input to the CNN model to extract and identify distinct features from various device types.
- Work in [12] uses a CNN to extract model-related features and then uses a support vector machine (SVM) to predict the camera model.
- In both of these cases, the role of deep learning is primary for the feature extractor

Cont.

- Similar techniques have also been applied for audio device identification [13]
- Radio fingerprinting has also been studied where devices are identified by their wireless radio device properties[14]
- In [14], Yu et al. propose a solution using partially stacking-based convolutional DAE to classify devices through reconstructing a high-SNR signal. Based on RF fingerprinting techniques
- Bassey et al. proposed a framework to detect unverified smart devices with deep learning [15].
  - First, they use a convolutional neural network to automatically extract high-level features from RF traces; then, they perform dimensionality reduction and decorrelation on deep features. Finally, they use clustering techniques to classify IoT devices.

Service Fingerprint Extraction Using DL

- Due to the dynamic nature of IoT networks, it can be difficult to maintain static fingerprints for devices as they are connected or removed from the network. Therefore, establishing a dynamic behavior baseline is essential.
- Fingerprinting IoT devices can also be a challenge due to the heterogeneous nature of IoT devices, protocols, and command interfaces.
- Service fingerprints identify IoT devices based on the services that they provide, which then generates a profile that can be used to identify the type of device that it is likely to be

Cont.

- Typically, this would be achieved using system logs, web traffic and or their battery consumptions as inputs to extract behavioral fingerprints
- Meidan et al. propose an IoT device classification framework based on HTTP packet analysis [16]. They perform this as a two-pass classification to
  - firstly distinguish between IoT devices and non-IoT devices and
  - then perform a fine-grain classification model to

Cont.

  - differentiate between nine distinct IoT devices.

- In [17], the authors propose to approximately model IoT behavior by the collection of communication protocols used, and the set of request and response traffic sequences observed, from which device features are then extracted from the network traffic.
- Finally, features are aggregated using a statistical model as a base profile for device identification.
- In [18], the proposed scheme extracts up to 23 features from each packet, from which they form a fingerprint matrix and use a random forest to develop a classification model.

Cont.

- [19] proposes to use information from network packets to identify devices. They observed that packet interarrival time (IAT) is unique among devices. They extract and plot the IAT graph for packets where each graph contains 100 IATs. Then, they use the CNN to learn features from device graphs and distinguish different devices.
- Another study in [20] attempts to automatically identify the semantic type of a device by analyzing its network traffic. First, they define a collection of discriminating features from raw traffic flows, and those features are used to characterize the attributes of devices then, they use a LSTM-CNN model to infer the semantic type of a device.

Cont.

- Due to the large variety of devices and manufacturers in IoT setting, other researchers [21] argue that traditional intrusion detection methods cannot suitably detect compromised IoT devices given the scale of devices being monitored.
- They propose DI¨OT, a self-learning distributed anomaly-based intrusion detection system, to identity compromised devices. DI¨OT can effectively build devicetype-specific behavior profile with minimal human efforts.
- Federated learning is utilized in DI¨OT to efficiently aggregate behavior profiles across devices. Compared with traditional machine learning, in the works described using deep learning, features are often automatically extracted from raw device traffic

Device Integrity Testing Based on DL

- Hardware Trojans are a major security concern where hardware can be accessed by untrusted third-party.
- Traditional methods include one-class anomaly detection, two-class classification, clustering, and outlier-based, utilizing training data such as on-chip sensor data and on-chip traffic data.
- Research on the topic of deep learning-based hardware Trojan detection methods is limited but increasing, with many currently based on simple neural networks as an anomaly detector

Cont.

- In works such as [22], they use power consumption data as the model input. To reduce the noise in data acquisition, wavelet transforms are used. A neural network is used to distinguish between normal chip power consumption and deviation in chip performance where a Trojan may be present
- Wen et al. [23] use self-organizing maps (SOMs) to detect hardware Trojans. They employ Hotspot to catch the steady-state heat-map from running IC. Then, a 2-dimensional principal component analysis (PCA) is used to extract features from the heat-map. the SOM is used to automatically distinguish Trojan-infected chips.

Cont.

- Work by [25] proposes to extract features from netlists; for each netlist, they get 11 features. Then, the deep multilayer neural network is used to find out malicious netlist. However, the role they play is as an anomaly detector with predefined features
- It is suggested that further research of deep learning in this application is still required.

Network behaviors in IoT

- Here, I focus on the modelling of network behaviors as a result of IoT devices, including device access control, connection-related activities, firmware upgrades, and remote access and control of devices
- In particular, it would be beneficial to develop a model that can identify malicious behaviors across the network so as to block remote access.
- Deep learning has recently been used to attempt to identify such attacks. Meidan et al. [26] use deep autoencoders to build normal behavior profiles for each device. They extract statistical traffic features and train autoencoders with features from benign traffic

Cont.

- When applied to new traffic observations for a new IoT device, there exists a bigger reconstruction error on the trained autoencoder which can be used to indicate that the device could be compromised.
- Similar approaches used in Kitsune [27] use ensembles of autoencoders to identify anomalies in IoT such as Mirai. Both of the above approaches assume that normal traffic activity can be approximately reconstructed, while an anomaly would cause large reconstruction error.

Cont.

- Other methods use the CNN to automatically identify malicious traffic in IoT. In [28], they turn the payload in the traffic packet into a hexadecimal format and visualize it into a 2D image. Then, they employ a lightweight CNN framework called Mobile Net to extract features from traffic images and malware classification.
- To deal with the volume of traffic needed to analyses this in a DDoS setting, in [29], they propose a deep learning lightweight DDoS detection system called LUCID
- authors in [30] employ damped incremental statistics as basic features. They then use triangle area maps (TAMs)-based multivariate correlation analysis (MCA) to generate grayscale images as training data from normalized traffic features. They then use these as input to a CNN to learn a model for detecting anomalies.

Model Data Abuse in IoT Environment

- Data gathered by IoT networks can be of great value, and abuse of this data can result in serious consequence, e.g., the case made against Cambridge Analytical.
- It is therefore crucial that IoT devices manage data responsibly
- Data leakage can occur from the generation of data, the use of data, and the transmission/ storage of data over the IoT network.
- For example, data collection by smart meters will reflect home usage patterns for electricity, gas, or water, which if leaked could expose attackers to information about when the house is occupied or not.

Cont.

- Five context parameters related to IoT data privacy are proposed by [31]: place (“where”), type of collected information (“what”), agent (“who”), purpose (“reason”), and frequency (“persistence”).
- e. In [32], the authors propose a deep and private-feature learning framework called deep private-feature extractor (DPFE). Based on information theoretic constraints, they are training a deep model which allows the user to prevent sharing sensitive information with a service provider and at the same time enables the service provider to extract approved information using the trained model.

Cont.

- Similar work in [33] proposes a feature learning framework that leverages a double projection deep computation model (DPDCM). Different from the traditional deep learning framework, they
- use double projection layers to replace the hidden layers, which can learn interactive features from big data. Furthermore, they design a training algorithm to fit the DPDCM model. To improve the learning efficiency, cloud computation is used. They also propose privacy-preserving DPDCM based on BGV encryption to protect personal data

Cont.

- Works by [34] demonstrate that decentralized federated learning can improve data privacy and security, while reducing economic cost. Works in [35] integrate deep reinforcement learning algorithms and the federated learning framework into an IoT edge computing system.
- The main focus of their work is to improve the efficiency of the mobile edge computing system. They design a framework called "InEdge AI" to maximize the collaborative efficiency among devices and edge nodes. With this framework, learning parameters can be exchanged efficiently for better training and inference
- Their framework can reduce unnecessary system communication while at the same time carry out dynamic system level optimization and application-level enhancement

Cont.

- Wang et al. studied a broad range of machine learning models optimized with gradient descent algorithms [36]. Their research first analyses the convergence bound of distributed gradient descent algorithms. Then, they propose an algorithm to reach the best trade-off between local and global parameter learning while given limited resource budget.

Data Integrity in IoT with Deep Learning.

- In an IoT setting, upholding integrity is vital to ensure that there is consistency between the actual, physical observation, and the transmitted data or signals that represent this activity.
- False data injection (FDI) is an attack against a cyberphysical system by modification of the sensor data, which could include SCADA (supervisory control and data acquisition) systems used widely in sectors supporting critical national infrastructure
- Works in [37] use deep learning algorithms to learn the behavior feature model from historical sensor data and employ the learned model to infer the FDI behavior in real time.

Cont.

- Similar work was proposed by Wang et al. WangH2018 used a two-stage sparse scenariobased attack model to detect attack in smart grid given incomplete network information. To effectively detect established cyber-attacks, they develop a defense mechanism based on interval state model. In their model, they use a dual optimization method to model the lower and upper bounds of each state parameter
- which will maximize variation intervals of the system variable. At last, they employ the deep learning model to properly learn nonlinear and nonstationary behavior features from historical electric usage data.

Deep Learning Methods for IoT Security

- I will summarize the methods using deep learning techniques to enhance IoT security.
- There are three major techniques of deep learning used for IoT security
  - Feature Learning Process
  - Deep Learning for Device Feature Extract
  - Network Behavior Modelling with Deep Learning.

RESEARCH GAPS

- Efficiency:- With the development of deep learning methods, various new architectures surpass state-of-the-art performance. However, many of them have not necessarily been developed for the IoT setting. To fully adapt these algorithms to an IoT setting would certainly help to improve the performance of recent studies [41, 42].
- Adaptive:- A static trained model cannot easily adapt to changing conditions and so could result in an increase in false positives and false negatives. Another consideration is that many IoT devices may be deployed in a wide scale of areas. The properties of the environment where IoT are deployed may vary from each other. Retraining a deep learning model for each setting not only costs a lot of time but also requires further labelled training data.
- Heterogeneous Data:-IoT devices produce different type and scale of data, such as data from signal frequency and network traffic, they will have different formats. Even data of same type may differ in scale, such as packets number and bytes number. Although they all belong to network features, they use different scale. How to handle those heterogeneous data is an ongoing problem [43, 44].

Cont.

- Resource Efficient Deep Learning:- modification on deep learning model itself, compressing or pruning the original deep learning model and result cache, preventing duplicate computation by sharing result among devices
- Lifelong Learning:- lifelong learning capabilities are crucial for computational learning systems and autonomous agents interacting in the real world and processing continuous streams of information. the IoT setting, with dynamically changing environments and low-powered devices, lifelong learning is needed to create more intelligent and efficient agents. Catastrophic forgetting is when a machine forgets what it learned before as it learns new things, which makes its performance worse over time. This happens because the machine thinks the old information is still the same even though new information is being added.

Conclusion

- According to this survey, deep learning has demonstrated significant potential in the IoT environment. The investigation of IoT device security aspects using deep learning technology is the main emphasis of this survey. Deep learning-based device fingerprinting and profiling in particular Ire thoroughly covered. To enhance feature mapping for device identification, a method for semantically meaningful device modeling was put forth. Finally, I talked about the issues and research directions I want to investigate in our future work.

Thank you For your Attention if any question you are well come