Attack vs. Defence for AI-Powered Systems in Smart Homes

Mahmoud Alkawareet 001302274, MD Abdul Raihan Tanzim 001341954, MD Juobiar Hossain 001265120, Sabiha Ahmed Toba 001340510

I. Introduction

In recent years, AI-powered solutions, which automate and streamline a variety of domestic chores, are now essential components of smart homes. Security cameras and smart thermostats use machine learning algorithms to understand user preferences, maximise performance, and enhance overall usefulness. However, there are additional risks and vulnerabilities brought about by this greater reliance on AI. These systems are vulnerable to internal and external threats as they grow more networked, which could jeopardise security, privacy, and even the home's structural stability. AI-powered smart homes face, focusing on technologies like AI model poisoning, AI manipulation in IoT systems, and 5G networks are increasing the attack surface in these systems. Federated learning could help protect privacy by decentralizing training and keeping data at the source Chen et al. (2023).

The security of machine learning-based assets in smart homes is examined in this paper from both an attack and defence perspective. Understanding the probable dangers to AI systems, creating an attack graph model to show potential attack vectors, and suggesting preventative measures to lessen these risks will be the main objectives. By analysing and discussing the interplay between attack and defence, the report aims to provide a comprehensive approach to securing AI-powered systems in the smart home environment.

II. Use Case Scenario

Several AI-powered gadgets collaborate to produce a smooth and effective living space in a typical smart home. A smart thermostat, for instance, learns the homeowner's preferred temperature using machine learning and modifies the heating or cooling system appropriately. Like this, security cameras with AI capabilities are employed to keep an eye on the house and identify any strange activity, alerting homeowners or local authorities. Additionally, users may use voice commands to operate various smart devices thanks to voice assistants like Google Assistant and Amazon Alexa.

These gadgets increase convenience, but they also present several risks. Unauthorised access to these devices could allow a hacker to alter their behaviour. For example, they may use voice assistants to access classified information, turn off security cameras, or change the temperature settings. As smart homes become more connected, the risk of such attacks grows, making it essential to design robust defence strategies.

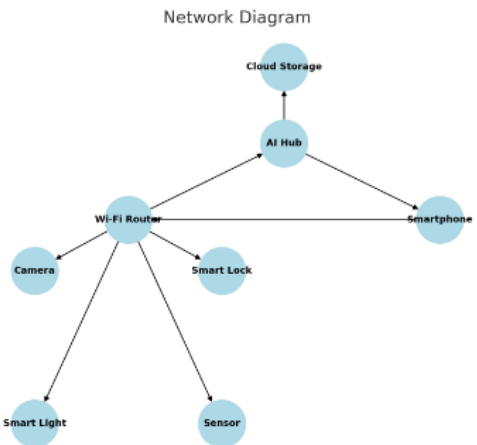

Figure 1: Network Diagram

The network diagram shows smart home network, showcasing the connections with various devices like cameras, smart locks, lights, and sensors, by communicating through a Wi-Fi router and AI hub. These connections, while offering convenience, also introduce potential vulnerabilities that require robust security measures.

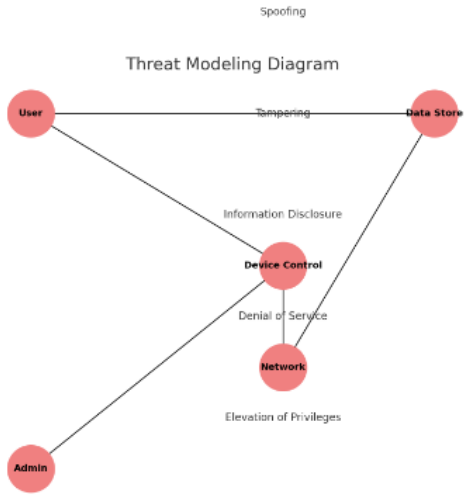

Figure 2: Threat Modelling Diagram

This threat modelling diagram highlights the user data, system and service availability by visualizing attack and their effects within a system.

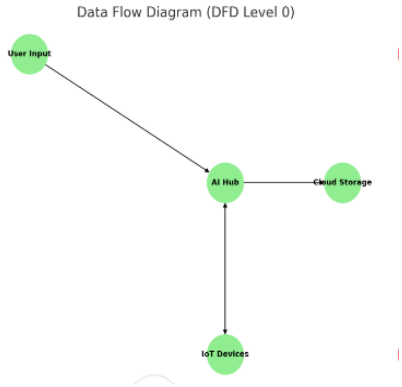

Figure 3.1: Data Flow Diagram Level 0

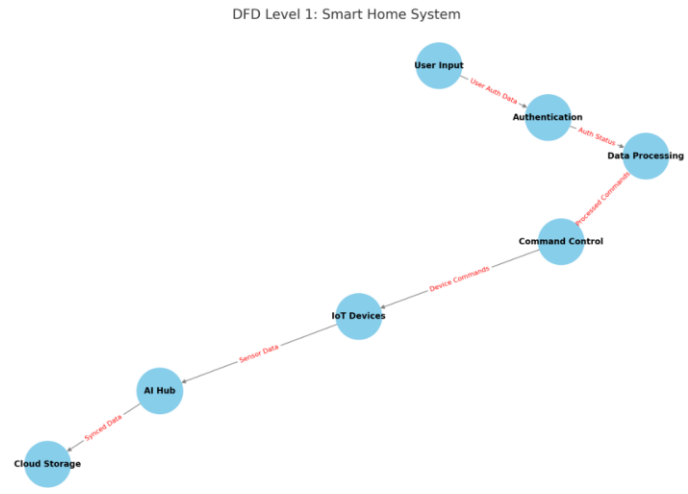

Figure 3.2: Data Flow Diagram Level 1

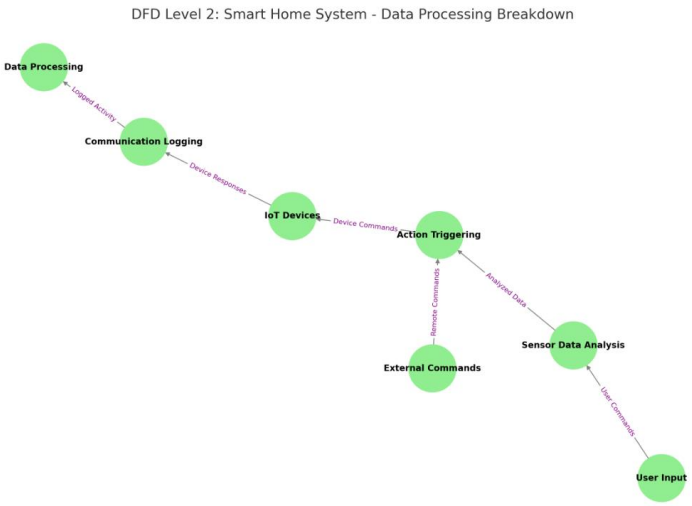

Figure 3.3: Data Flow Diagram Level 2

III. Analysis and Discussion

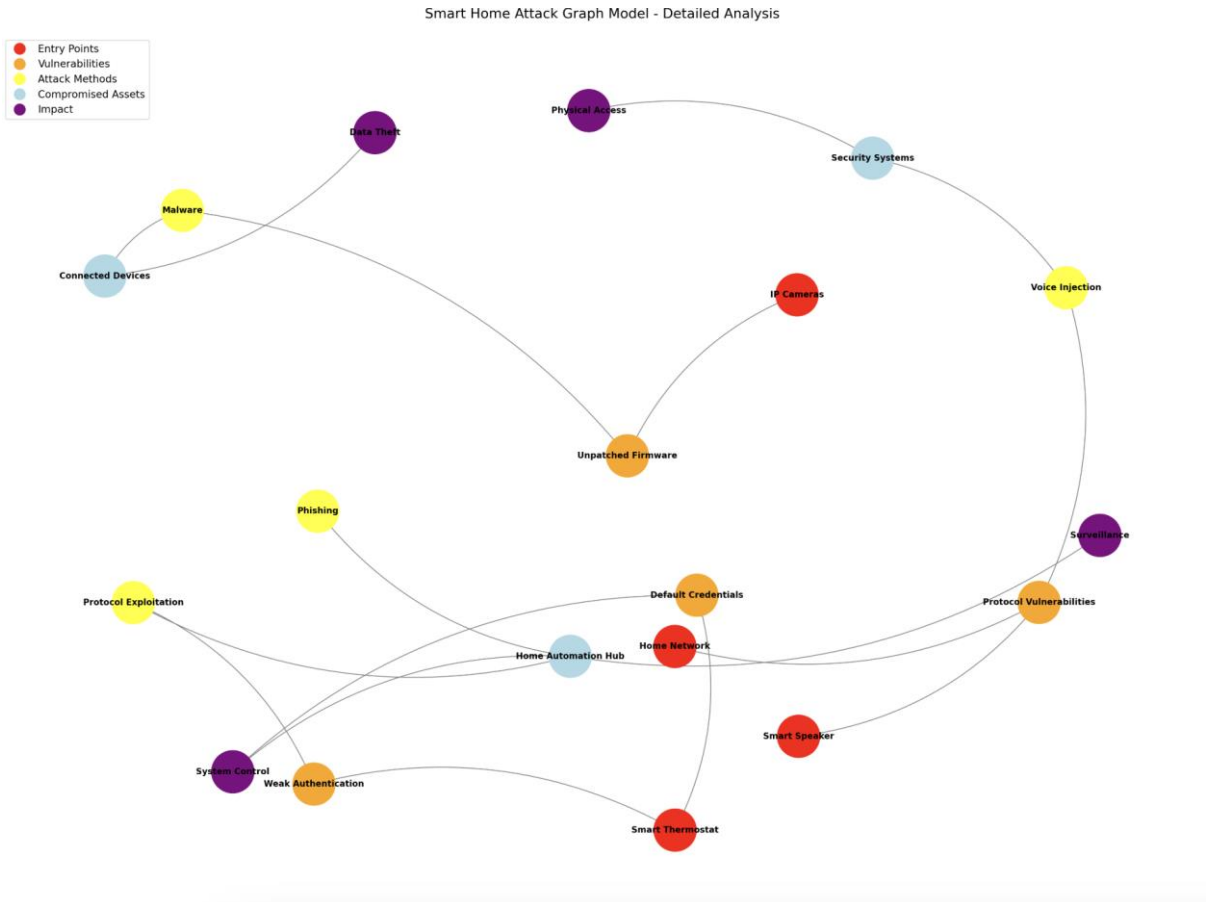

Figure 4: Attack Graph Model

We developed an attack graph model to better comprehend the possible risks to a smart home driven by AI. This model graphically depicts how an attacker could exploit system vulnerabilities and demonstrates the relationships between different attack techniques.

The attack graph for a smart home consists of several key components:

- **Entry Points**: These are the initial attack vectors, such as compromised devices (e.g., a vulnerable smart speaker) or weak network security protocols.
- **Exploitation**: Once inside the network, attackers may leverage various techniques, such as injecting malicious data into a model, manipulating device behaviour, or compromising communication channels between devices.
- **Persistence**: Attackers may attempt to maintain access by creating backdoors, disabling security features, or infecting devices with malware that provides continued access.
- **Impact**: The final stage involves the consequences of the attack, such as privacy violations, physical harm (e.g., tampering with security systems), or even financial loss.

This attack graph provides a useful tool for understanding how different attack methods are interconnected and how attackers might escalate their control over the system.

•Attack Strategies

Potential dangers to AI-powered systems in smart homes were found to include the following attack techniques:

- **Adversarial Attacks**: These assaults entail tampering with a machine learning model's input to make it provide inaccurate classifications or predictions. An intruder may be able to avoid detection if, for example, a smart security camera is shown manipulated data that prevents it from identifying a person or object (Szegedy et al., 2013).
- **Data Poisoning**: This kind of attack involves the hacker altering the data that is used to train machine learning algorithms. They can impair the model's performance and lead to incorrect judgements by introducing harmful input. Data poisoning may cause a smart thermostat in a smart home to make inaccurate temperature adjustments (Biggio et al., 2012).
- **Inference Attacks**: These attacks seek to retrieve private user data and other sensitive information from the AI model. To obtain personal data from a smart home system, for instance, attackers could reverse-engineer a trained model (Shokri et al., 2017). This could entail retrieving voice data from a voice assistant or security camera surveillance footage in a smart home.
- **Model Inversion**: Model inversion is a more advanced type of inference attack that enables attackers to recreate training data from a model that has been trained. In smart homes, where the training data may include private information like user preferences or behavioural patterns, this is especially problematic (Fredrikson et al., 2015).

•Defence Strategies

To mitigate the risks posed by these attack strategies, several defence mechanisms can be employed:

- **Adversarial Training**: Using this method, models are trained to identify and withstand hostile inputs. The model can be strengthened against malevolent attacks by adding adversarial cases during the training phase (Goodfellow et al., 2014). Adversarial training may be able to guarantee that smart cameras and sensors in a smart home continue to operate as intended even in the face of changed input data.
- **Data Validation and Anomaly Detection**: Finding odd patterns in data can assist in spotting and averting hostile assaults and data poisoning. A smart thermostat, for example, may keep an eye out for abrupt and inexplicable changes in user behaviour and sound an alarm if the data is unusual (He et al., 2017). Anomaly detection could also be used by security cameras to identify anomalous activities or distorted images.
- **Encryption and Secure Communication**: It is possible to stop hackers from capturing and changing data sent over a network by making sure that every communication between smart home devices is encrypted. Protecting the integrity of data in models used for vital systems, like home

security or health monitoring, is especially crucial (Zhang et al., 2019). Lightweight cryptographic algorithms that are suitable for resource-constrained devices in smart homes. Cryptographic protocols like ChaCha20 or LEA (Lightweight Encryption Algorithm) could be implemented to secure data in devices such as thermostats or cameras without sacrificing performance. Look into how Xu et al. (2017) reviews lightweight cryptography in IoT and smart home security contexts.

- **Access Control and Monitoring**: Implementing strong access control mechanisms, such as multi-factor authentication, can help limit the risk of unauthorized access. Monitoring systems can also track and log suspicious activity, enabling quicker identification and response to potential breaches (Gollmann, 2011).

•Discussion on Balancing Security and Efficiency

In smart home environments, balancing security and efficiency is crucial. Many smart devices are resource-constrained, meaning that implementing complex security measures can lead to performance degradation. For example, adding extensive encryption or advanced anomaly detection algorithms may introduce delays in processing data or reduce battery life for devices like smart thermostats or security cameras.

To address this issue, lightweight cryptographic protocols and efficient machine learning algorithms that do not require excessive computational power should be prioritized (Xu et al., 2017). Moreover, defence mechanisms should be adaptive, scaling security measures based on the device's capabilities and the potential risks it faces. For example, more sensitive devices, like security cameras, may require stronger protections, while less critical devices, like smart lights, can afford simpler security measures.

IV. Conclusion

Numerous automation and convenience advantages result from the incorporation of AI into smart home systems. But it also brings with it fresh privacy and security issues. The possible attack vectors that could target AI-based assets in smart homes have been examined in this paper, along with suggested defence tactics to lessen the risks. To guarantee that smart home devices can continue to operate at their best while being shielded from malevolent attacks, the relationship between security measures and system efficiency needs to be thoroughly examined. We can create safe and effective AI-powered smart home environments by putting strong defences in place and considering the limitations of smart gadgets.

V. Team Contribution

In this project, the team members divided responsibilities based on their individual strengths and interests. Below is a breakdown of each team member’s role and contributions to the overall project:

- **Mahmoud Alkawareet**: took the initiative to investigate and evaluate different attack tactics. This member was principally in charge of researching adversarial attack methods and model poisoning. Finding recent studies and articles on model modification techniques was part of his research, and he shared his findings with the team. He and Md Abdul Raihan Tanzim also offered feedback on the attack graph model and helped write the report's "Attack Strategies" section.
- **Md Abdul Raihan Tanzim**: investigated inference attacks and the possibility of data leakage in smart home AI systems in close collaboration with Team Member 1. He helped create a flowchart for the attack graph model and examined how attackers could obtain private information by looking at the outputs of the ML models. They also provided input on adversarial attacks and helped to improve the explanation of inference attacks.
- **MD Juobiar Hossain:** centred on defensive strategies, such as adversarial training and data validation. He oversaw investigating defence tactics that counteract adversarial examples and

poisoning attacks, with a focus on data validation techniques that are practical for real-time application in smart home systems. When writing the first draft of the "Defence Strategies" section, MD Juobiar Hossain and Sabiha Ahmed Toba made sure that every defensive tactic included the appropriate citations.

- **Sabiha Ahmed Toba:** worked together with MD Juobiar Hossain to investigate performance and security trade-offs and access control systems. She investigated access control strategies that stop unwanted model manipulation, such as monitoring and logging systems. The "Discussion on Balancing Security and Efficiency" part was also written by Sabiha Ahmed Toba, who included information on the trade-offs of putting defences in place in smart devices with limited resources.

References

- Biggio, B., Fumera, G., & Roli, F. (2012). Poisoning Attacks in Data Mining. *ACM Computing Surveys*, 44(2), 1-35.
- Fredrikson, M., Jha, S., & Rahmati, A. (2015). Model Inversion Attacks that Exploit Confidence Information and Basic Countermeasures. *Proceedings of the 22nd ACM SIGSAC Conference on Computer and Communications Security*, 1322–1333.
- Gollmann, D. (2011). *Computer Security*. Wiley.
- Goodfellow, I., Shlens, J., & Szegedy, C. (2014). Explaining and Harnessing Adversarial Examples. *Proceedings of the International Conference on Machine Learning*.
- He, Z., Wu, J., & Yang, M. (2017). Anomaly Detection for Smart Homes: A Survey. *IEEE Internet of Things Journal*, 4(3), 1005-1018.
- Shokri, R., Stronati, M., Song, L., & Shmatikov, V. (2017). Membership Inference Attacks Against Machine Learning Models. *Proceedings of the 2017 IEEE Symposium on Security and Privacy*.
- Szegedy, C., Zaremba, W., Sutskever, I., Bruna, J., & Fergus, R. (2013). Intriguing Properties of Neural Networks. *Proceedings of the International Conference on Learning Representations*.
- Xu, J., Liu, M., & Chen, X. (2017). Lightweight Cryptographic Solutions for Internet of Things Security: A Survey. *IEEE Access*, 5, 25657-25678.
- Zhang, Y., Yang, W., & Li, X. (2019). Secure and Efficient Data Sharing in Cloud-Based IoT Systems. *Future Generation Computer Systems*, 93, 389-397.