A Common Architecture-based Smart Home Hardware Forensics for Scalable Investigations

Sungbum Kim  
Ajou University

Gwang Sik Lee  
Supreme Prosecutors Office of the Republic of Korea

Jian Song  
Supreme Prosecutors Office of the Republic of Korea

Insoo Lee  
Supreme Prosecutors Office of the Republic of Korea

Taeshik Shon  
tsshon@ajou.ac.kr

Ajou University

Research Article

Keywords: Digital Forensic, Forensic framework, Internet of Things, Smart home, Smart home platform

Posted Date: June 21st, 2024

DOI: https://doi.org/10.21203/rs.3.rs-4532459/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License

Additional Declarations: No competing interests reported.
Abstract

The smart home platform communicates with internet of things (IoT) devices, smartphones, and cloud servers to provide convenient services, storing user information and device operation and user behavior-related data. This data is crucial for criminal investigations, highlighting the importance of smart home forensics. Currently, advanced software and hardware technologies continue to be developed in the smart home market, and based on this, companies are releasing new services and devices. Therefore, scalable platform-oriented forensic research is needed for efficient digital investigation. This study identifies the components and structures of smart homes to derive a common architecture representing various environments. It proposes a three-stage smart home forensics framework: analyzing application functions to infer data, extracting and analyzing data from devices, and identifying data useful for criminal investigations. Its applicability is demonstrated with Samsung SmartThings and Xiaomi Mi Home platform testbeds.

1. Introduction

Advances in software and hardware technologies are making factories, vehicles, homes, and infrastructure smarter[1]. Companies and countries are focusing on research and development in areas such as networks, devices, and cybersecurity to expand smart industries[2-6]. For smart homes, home appliance manufacturers, mobile carriers, and construction companies are currently concentrating on smart home businesses. Numerous companies are launching smart home IoT devices centered on their proprietary and third-party platforms. Consequently, due to the expansion of services offered by smart homes, IoT devices are generating and storing significantly more data. This data is stored inside the device and server through communication between the IoT device and the cloud server of the smart home platform. The smart home platform stores user personal information, such as phone numbers, email, and home addresses, as well as information that can identify users' behavior such as logs and timestamps using devices. Therefore, the data acquired within the smart home environment is crucial from a forensic perspective because it can be used to identify suspects and prove criminal behavior during investigations.

Investigation agencies and research institutes conducted data acquisition and analysis studies centered on smart home IoT devices such as home cameras, artificial intelligence (AI) speakers, and smart plugs and smartphones to expand the scope of digital investigations[7-10]. However, it is difficult to apply existing research because IoT devices with new technologies are continuously being developed and applied. In addition, smart homes are currently being used around platforms, but existing studies have conducted forensic analysis research focusing on individual devices, which hamper their use in real-world investigations.

With the ongoing development of IoT devices, the smart home environment is also evolving in diverse ways. IoT devices that consist of various software and hardware generate and store different data depending on their functions. In addition, IoT devices are complexly connected to each other, so data is
stored distributedly. IoT forensics has several limitations, making it difficult to develop a standardized forensics framework\[11-14\]. The IoT environment in the smart home is established by users around smart home platforms such as Amazon Alexa and Google Home, and these smart home platforms have some common components. For example, most smart home platforms have products that provide security services such as smart cameras and door locks, and appliances that provide domestic services such as smart refrigerators and smart washing machines. In other words, Smart home IoT devices offer functionality within the scope of services related to the home. In addition, smart homes generally consist of network communication of essential components such as cloud servers, smartphone apps, and IoT devices. In this study, to compensate for the limitations of smart home IoT forensics, we derive a layer-based architecture that expresses smart home structures using smart home’s common components. The layer-based smart home architecture can be applied to the actual IoT environment during investigation to efficiently identify the complexly connected smart home environment. It also proposes a smart home forensic framework using the derived architecture. This framework analyzes the functions of IoT devices to infer what content of data is generated, acquires actual data, and classifies it according to data characteristics. By inferring the data, investigators can identify the target IoT device to find the artifacts needed for the event. Determining the target IoT device enables rapid and efficient investigation. In addition, by classifying data characteristics, the necessary evidence can be quickly identified according to the type of crime. We validate the proposed framework through Samsung SmartThings and Xiaomi Mi home platforms. The contributions of this study are as follows.

1) It classifies meaningful elements from a digital forensics perspective by identifying components that are common in smart home ecosystems. 2) Based on the identified components, a smart home common structure that can be continuously applied is derived as a layering architecture to lay the foundation for smart home forensics. This layered architecture makes it easy to identify complex IoT environments. 3) It proposes a forensic framework that can be applied collectively in a general smart home environment. The proposed framework derives forensic artifacts by inferring data generated in a smart home environment and extracting data from actual devices. The derived artifacts are classified into device use data, user data, and smart home environment identification data according to data characteristics. This classification can provide convenience for subsequent investigations. 4) Data inference and extraction are conducted on Samsung SmartThings and Xiaomi Mi Home platforms to derive a plan to utilize the derived artifacts. This scenario demonstrates the applicability of the proposed forensic framework in a general smart home environment.

2. Background

The concept of a smart home emerged as users began utilizing a wide range of IoT devices to access various services within their homes. In the context of smart home forensics, it is important to identify the components constituting the smart home environment. Smart homes generally follow the same structure as the one depicted in Fig. 1. IoT devices are interconnected seamlessly to offer diverse services, enabling users to control and manage them both within and outside their homes. These devices communicate with devices and cloud servers through various network protocols for smart home
services. Devices can be connected wirelessly, such as via Wi-Fi, as well as through wired connections like Ethernet. In addition, devices such as sensors are connected using low-power wireless communication such as Bluetooth, ZigBee, and Z-wave. These connected IoT devices generate a wide range of data, which is stored in the devices’ internal storage and on cloud servers. Data stored on cloud servers can be used to remotely control and manage apps installed on management devices such as smartphones and tablets. This section describes the components constituting the smart home environment.

Smart home IoT devices are the most basic form of smart home construction. These devices were released by embedding hardware (HW) and software (SW) such as central processing unit (CPU) and operating system (OS) in devices used at home such as conventional washing machines, Televisions, and speakers, and adding additional functions. It also offers new services through devices that are not typically found in homes before, such as crime prevention sensors and smart meters. Smart home IoT devices have device-specific functions, communicate with cloud servers, and can be controlled and managed through smartphone apps. Certain devices are also used for controlling other smart home devices.

Smart home services are classified by function by companies and users, and accordingly, users select and install IoT devices necessary to use the desired smart home services at home. For instance, devices such as home cameras and crime prevention sensors can be purchased and used for house crime prevention. As a result, IoT devices are categorized based on their services and used for specific purposes. This categorization allows for the inference of the type of data generated by these devices.

Smart home platforms have emerged to allow users to efficiently control smart home IoT devices as new IoT devices and services are being introduced. Currently, large companies are developing various devices and technologies to dominate the smart home ecosystem around their smart home platforms, and small and medium-sized companies are working with large companies to increase the compatibility to sell their IoT devices. While most platforms offer similar services, there are variations and distinctions among certain platforms. Therefore, it may be important to identify commonalities and differences in smart home platforms from a forensic perspective.

Cloud servers integrate and manage and control while communicating with smart home applications and enable scalable smart home environment configuration through software development kits (SDK) and application programming interface (API). Accordingly, various data are stored on the server. In particular, miniaturized and integrated IoT devices, it is dependent on cloud servers because there are numerous cases where there is no internal storage of the device. Therefore, when conducting smart home forensics, methods such as packet analysis can be considered to infer data on the cloud server.

The smart home has various components as described above, and data is exchanged between the components through various communication protocols. Protocols such as Ethernet and Wi-Fi can be used for communication protocols, and low-power communication such as Bluetooth and Z-wave also exist. In addition, applications installed in management devices such as smartphones and tablets are
used to control components, and hubs and gateways used for sensor linkage are also included. Protocols and devices used in this communication and management are significant in forensic analysis as they can capture the interaction structure among components.

3. Related Works

The smart home environment stores data generated through IoT devices and provides services to users using them. The data generated by IoT devices is useful from a forensic perspective because it occurs in residential areas. Accordingly, various forensic analysis studies have been conducted on IoT devices and smartphones. Gómez, Juan Manuel Castelo et al. conducted a forensic study on smart home kits consisting of smart home devices such as hubs and sensors [15]. They attempted to acquire data from Xiaomi Mi smart sensor sets and linked smartphones and acquired some smart home artifacts from smartphone applications. Boucheraud et al. conducted IoT environmental forensic research based on crime scenarios [16]. They used smart bulbs, cameras, sensors, scales, and wearable devices other than home automation devices such as AI speakers, raspberry pies, and hubs to configure the IoT environment in their homes. In addition, forensic analysis was performed through logical extraction using API and Android debug bridge (ADB) and physical extraction using joint test action group (JTAG) and Chip-off. Li et al. proposed a forensic process focused on physical extraction techniques for IoT devices [17]. They used 3D printers and pogo pins to connect to interfaces including JTAG and universal serial bus (USB) to perform tests on five data acquisition methods. Kim et al. performed data acquisition and analysis on smartphone applications linked to IoT devices. They connected Google Nest, SmartThings, and Kasa Cam to smartphone apps, and identified voice command information, phone information, and smart home device information through data analysis [18]. Iqbal et al. conducted a forensic analysis of five types of smart plugs. They conducted data analysis on smartphones connected to the smart plug, as well as performed data analysis by analyzing communication packets using Wireshark. Hutchinson et al. conducted data acquisition and analysis studies on August Smart Doorbell Pro and August Smart Lock Pro devices [19]. They conducted experiments by linking routed and escaped smartphones with devices and collected network traffic generated from IoT devices and smartphones using Wireshark. In addition, user data, user location data, and Doorbell camera photos were extracted through smartphone analysis that acquired administrator rights. Shin et al. conducted encrypted traffic collection and analysis between clouds and devices for AI speakers and developed a tool to acquire artifacts stored in the cloud through traffic decoding methods through five certificate injections [20]. Youn et al conducted a forensic study on Echo Show 2nd, an AI speaker with a display [21]. They extracted and analyzed data from Echo Show and proposed a smart display digital forensics framework.

Most of the smart home forensic studies conducted thus far have focused on IoT devices and smartphones such as AI speakers and smartwatches [22–28]. These studies have analyzed the data generated by IoT devices and smartphone applications in detail, which is meaningful in that it is a forensic methodology that can respond when a new device is released. However, research on device-oriented forensic analysis methods does not fundamentally solve the difficulties of smart home IoT forensics that there is no standardized forensic framework. In addition, smart home users purchase and
use devices around the smart home platform, however, as existing smart home research has focused only on IoT devices, there is a gap between research and actual investigation environments. Therefore, this study proposes a common architecture of smart homes and a framework used in a general smart home environment. We also validate the proposed framework by establishing a platform-level testbed.

4. Smart Home Common Architecture

For forensic frameworks that can be used in the latest smart home environment, the identification of common components and structures for smart homes is required. Smart homes operate around IoT. IoT devices include device elements such as cameras, memory, and OS from a forensic perspective. These devices have different types of data generated depending on the product family that provides similar services. These services are integrated and controlled by the platform, and each platform has different support services and devices depending on the company it operates. These platforms are linked between platforms using cloud servers, support APIs, and SDKs for compatibility and support management services through smart home apps. This smart home structure is connected through various wired/wireless communication and is managed through devices such as smartphones and tablets. As such, the main components can be divided into five components: cloud servers, platforms, services, IoT devices, and device components. Accordingly, in this study, a common smart home architecture was derived based on a layered architecture. The designed smart home architecture had five layers as shown in Fig. 2, and communication and management functions can be configured vertically because they can be applied to all layers. For example, AI speakers consist of parts such as microphones and speakers, such as the device components layer, and HW/SW such as CPU, memory, and OS. AI speakers play a crucial role in controlling and managing smart homes as a unified system. They offer home automation services at the service layer, which may vary in functionality depending on the specific smart home platform. If the smart home platform is a platform of a home appliance manufacturer such as the platform layer, then more home appliances can be linked than other platforms, and if it is an IT company, it can use its IT service linkage function. The smart home platform communicates with the cloud server through the application and controls it remotely, and new devices can be linked through APIs and SDKs (cloud layer). Smart homes can be controlled through devices such as smartphones and tablets (management), and communication such as Ethernet, 4G, and Wi-fi is used at each stage (communication).

The cloud layer located at the top, is an element for smart home operation, and these elements operate based on the cloud server of the smart home. The layer is a smart home application, SDK. It contains elements such as API and smart home-related data. The platform layer involves the provision of smart home platforms and includes platforms of companies such as home appliance manufacturers and mobile carriers. These platforms provide smart home services using services and devices from their or other companies through cloud servers. The service layer is a variety of services provided by the platform, including services such as home automation, crime prevention and security, and smart home appliances. The IoT device layer includes an IoT device for providing a smart home service. Various IoT devices are used to provide each service. For example, devices such as smart speakers and hubs are
used to provide home automation services and security services can be provided through devices such as smart door locks/doorbells and smart home cameras. The device components layer is a component of a smart home IoT device and includes all hardware such as sensors/actuators, cameras, CPUs, and software such as OS. Data stored in the device may be inferred using the components constituting the IoT device. In addition to the five layers, elements for interlayer communication and smart home management are also included. Low-power communications such as Bluetooth, Zigbee, and Z-wave, as well as Ethernet, 4G/5G, and Wi-Fi, are included for device-to-layer communication. It also includes mobile devices such as smartphones and tablet PCs, as well as hubs and gateways for smart home management and control. Smart home platforms consisting of these various layers provide services specific to their platforms through organic interactions between layers and components. This common architecture includes common elements of smart homes and can be added based on the characteristics of smart home platforms. Therefore, it can be applied to various smart home platforms and can be scaled for newly released smart homes.

5. Forensic Analysis Framework for Smart Home Ecosystem

The common architecture includes most components of smart homes environment. And we proposes a framework for smart home forensic using smart home environment information derived from common architecture like Fig. 3. The forensic framework consists of three stages, as shown in Fig. 4: 1) functional analysis-based data inference, 2) real-world device-based data identification 3) data identification available for criminal investigation. App service analysis is performed to infer data in the smart home environment. Smart homes can generally be controlled and managed through platform apps installed on management devices such as smartphones. Therefore, it is possible to infer data generated through common functions of IoT devices and platforms through app analysis. By inferring data, it is possible to narrow the scope of the analysis by selecting the target device, which can help to conduct an efficient forensic investigation. Second, the data to be analyzed is extracted and analyzed. Smart home data can exist in cloud servers or management devices as well as in-device storage and identifies where inferred data is stored. Finally, the identified data are classified according to characteristics. Data types can be classified into device use data, user data, and smart home environment identification data, and these data types vary depending on the event.

5.1. Stage 1: Guessing Data Through APP Service Analysis

In general, the smart home environment is managed and controlled through a smartphone application. IoT devices are linked to cloud servers using smartphones to communicate smart home-related data to provide services. Although there are sensor types that can only be used for low-power wireless communication, all IoT devices transmit and receive data around cloud servers and Wi-Fi because they communicate with servers through hubs in conjunction with smart hubs. Therefore, because most smart home services are provided according to the functions provided by smartphone applications, services, and device functions can be identified based on smartphone application function analysis and the resulting data can be inferred. There are various devices at the actual investigation site, but it is difficult
to extract and analyze data for all devices owing to various restrictions. Therefore, inferring the stored data is the basis for selecting a target device to perform data extraction. For example, for a residential intrusion incident, an analysis is conducted on devices such as home cameras, which are inferred to store images of outsiders rather than devices such as smart TVs and smart speakers.

Identification data in the smart home platform may be composed of two types: expected data for each device and common expected data for the platform. The expected data for each device is data generated to provide unique functions for each service and device, and the expected data varies depending on the device's function. Platform-common expected data is data generated throughout the component through platform-specific functions, and the expected data has different characteristics depending on platform-common functions and management device systems. For data in a management device system, expected data can be identified through various smartphone analysis studies. Therefore, it is possible to derive common expected data for each device and platform through smart home application function analysis.

5.2. Stage 2: Data Extraction and Analysis

The data inferred through functional analysis is verified through the data obtained by applying the actual device target data acquisition method, and the corresponding data storage location is identified. The location where data is stored can be classified into three categories: device, cloud, and management device.

When data is stored in a device, it can be extracted and identified by applying a data acquisition method for actual devices. Here, device software and hardware analysis should precede, and data extraction is possible if there is storage inside the device. In-device data extraction methods include logical extraction, such as extraction using a connection between IoT devices and PCs, and physical extraction through Universal Asynchronous Receiver-Transmitter (UART), JTAG, and Chip-off. When stored in the cloud, it can be identified through device-to-cloud communication packet analysis. When analyzing packets, the device must support Wi-Fi and can capture packets through the PC's mobile hotspot function. For management devices, smartphones are generally used, and data acquisition methods in smartphones are widely known through various forensic studies. After acquiring data in a smartphone, it can be extracted through application data analysis in a management device.

5.3. Stage 3: Classification Based on Data Characteristics

Smart home data can be composed of three types based on data characteristics through functional analysis and data identified through actual devices. Device usage data' is data that can identify the alibi of the person who used the device through the device usage log. The 'smart home environment identification data' can identify the configuration of the smart home environment, and more data can be obtained through other acquisition methods using the data. The 'smart home environment identification data' includes communication information, device specification information, device identification information, and a list of linked devices.
6. Case Studies for Forensics Analysis of Smart Home Platforms

To validate the proposed forensic framework, a smart home platform test bed was established, and forensic analysis was performed. The Samsung SmartThings and Xiaomi Mi Home test beds used in the experiment were configured as shown in Table 1. For management devices, they were linked using routed Samsung Galaxy S9+ (Android 10). These test beds acquire artifacts through data inference and extraction and derive manners to utilize artifacts through scenarios that were found in the home of a suspect in a fugitive murder case.

<table>
<thead>
<tr>
<th>Smart home platform</th>
<th>Device</th>
<th>Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Samsung SmartThings</td>
<td>AI speaker</td>
<td>sm-v310</td>
</tr>
<tr>
<td></td>
<td>Smart hub</td>
<td>IOT-V3P03</td>
</tr>
<tr>
<td></td>
<td>Door sensor</td>
<td>IOT-MPP03</td>
</tr>
<tr>
<td></td>
<td>Motion sensor</td>
<td>IOT-MTP03</td>
</tr>
<tr>
<td></td>
<td>Smart button</td>
<td>IOT-BTP03</td>
</tr>
<tr>
<td></td>
<td>Doorlock</td>
<td>ECK-300</td>
</tr>
<tr>
<td></td>
<td>Robot vacuum</td>
<td>VR30T80313W</td>
</tr>
<tr>
<td></td>
<td>Smart plug</td>
<td>PM-B550E-W</td>
</tr>
<tr>
<td></td>
<td>Air monitor</td>
<td>ACM-B1M0S</td>
</tr>
<tr>
<td></td>
<td>Smart watch</td>
<td>PM-B550E-W</td>
</tr>
<tr>
<td></td>
<td>Smart tag</td>
<td>EI-T5300BBEGKR</td>
</tr>
<tr>
<td></td>
<td>Smart Desk lamp</td>
<td>MT428D21K</td>
</tr>
<tr>
<td>Xiaomi Mi Home</td>
<td>IP camera</td>
<td>MJSXJ09CM</td>
</tr>
<tr>
<td></td>
<td>Robot vacuum</td>
<td>Xiaomi Dreame Robot vacuum F9</td>
</tr>
<tr>
<td></td>
<td>Smart TV</td>
<td>L55M5-5ASP</td>
</tr>
<tr>
<td></td>
<td>Air purifier</td>
<td>AC-M16-SC</td>
</tr>
<tr>
<td></td>
<td>Smart watch</td>
<td>SYB01</td>
</tr>
</tbody>
</table>

6.1. Samsung SmartThings

6.1.1 Guessing Data Through APP Service Analysis
Through smart home app function analysis, the function of smart home devices that provide services is identified to infer expected data that may occur based on the function. For smart things, the entire smart home environment is controlled through the “Smart-Things” application. Accordingly, the function of the linked device is identified through the Samsung SmartThings application. Figure 5 is the main feature tab of the Galaxy Home Mini, an AI speaker. Through this approach, we can ascertain the available functions of Samsung AI speakers. Samsung AI speakers offer features such as alarms, timers, reminders, and music playback. It can also be inferred from the setup function that networks information such as Wi-Fi mac addresses will be stored in addition to AI speaker location, preference information, and user accounts. Through this approach, the stored data can be inferred as shown in Table 2. This data may be stored in the device or other cloud servers and smartphones. This can be identified through actual data analysis.

Table 2
Expected data according to smart things IoT device type

<table>
<thead>
<tr>
<th>Device</th>
<th>Expected data</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI speaker</td>
<td>Setup info, Voice history, Alarm/timer/reminder info, Linked third-party info, Network info</td>
</tr>
<tr>
<td>Smart hub</td>
<td>Device on/off info, A List of devices linked to the smart hub</td>
</tr>
<tr>
<td>Door sensor</td>
<td>Real-time temperature info, Door open history info</td>
</tr>
<tr>
<td>Motion sensor</td>
<td>Motion detection record info, Temperature info</td>
</tr>
<tr>
<td>Smart button</td>
<td>Usage history, Settings info</td>
</tr>
<tr>
<td>Doorlock</td>
<td>Lock status info, door lock operation record</td>
</tr>
<tr>
<td>Robot vacuum</td>
<td>Scheduled cleaning settings info, cleaning history</td>
</tr>
<tr>
<td>Smart plug</td>
<td>Energy consumption, device usage records</td>
</tr>
<tr>
<td>Air monitor</td>
<td>Co2, temperature and humidity info by time zone</td>
</tr>
<tr>
<td>Smart watch</td>
<td>User health info</td>
</tr>
<tr>
<td>Smart tag</td>
<td>location info, recording instrument location</td>
</tr>
<tr>
<td>Smart Desk lamp</td>
<td>usage history, settings info</td>
</tr>
</tbody>
</table>

6.1.2. Data Extraction and Analysis

The data expected earlier can be divided into three types of storage: internal storage of the device, cloud server, and management device (smartphone). Accordingly, data acquisition was performed by targeting each data store.

6.1.2.1 Data in Device Storage
To perform data acquisition inside IoT devices, it is necessary to first verify the presence or absence of storage inside the device through device specification analysis. Through device specification analysis, devices with internal storage are AI speakers, smart hubs, robot vacuum cleaners, air monitors, and smartwatches. For AI speakers, 4GB Nand Flash memory (KLM4G1FETE-B041, 153ball) was identified, and data were acquired through chip-off as shown in Fig. 6. The separated chip could extract dump images after connecting to a PC through a flash memory reader. The dump image was mounted on Linux to identify the partition and extract the image. Because of the mount, a dump image and partition of 3.7 GiB capacity could be identified as shown in Fig. 7. In addition, as shown in Fig. 8, it is estimated that the 16th partition, known as the user partition, is based on the examination of the partition using FTK Imager.

Multiple artifacts could be obtained from the db file stored in the AI speaker dump image. As shown in Fig. 9 the ‘device_table’ table of ‘sc.db’ currently stores a list of smart home-linked devices. In addition, the Unique_table of cloud_pdm.db stores the Login id used to link the account as shown in Fig. 10. The login ID is usually an email address. In addition, it was confirmed that device setting information including the user's name, date of birth, and timer setting information was stored.

For robot vacuum cleaners, 4GB Nand Flash memory of the same model as AI speakers were identified, and dump images were similarly extracted using chip-off. In the robot vacuum cleaner image, cleaning records including cleaning start timestamps were stored in the History table of history_new3.db as shown in Fig. 11. In addition, information on the reservation cleaning time set by the user was stored.

For smart hubs, the same model, 4GB Nand Flash memory, could be identified and the image dumped, but internal data could not be verified because of disk encryption through LUKS, as shown in Fig. 12. In addition, for air monitors, Nand Flash memory was identified as a result of printed circuit board (PCB) analysis, but internal extraction was infeasible owing to the absence of a reader capable of reading the memory, and for smartwatches, Nand Flash memory could not be identified.

### 6.1.2.2 Cloud server

Most smart home IoT devices communicate with cloud servers through Wi-Fi. Accordingly, communication packets between the wall pad and the cloud may be captured through the PC mobile hotspot function. Because communication packets send and receive information related to device use and users, data stored in the cloud server can be inferred, which can be used in future warrant issuance screening. However, for packets from AI speakers, the contents were unknown because most packets were encrypted with TLS communication, as shown in Fig. 13, and other devices were also encrypted with TLS.

### 6.1.2.3 Smartphone

For a smartphone, it can be inferred that various information was stored because all IoT devices were managed through an application. The smartphone used in the experiment was rooted in advance, and
the dump image was acquired through ADB. Several db files were stored in the application data on the smartphone. In the devices table of CloudDb.db, a list of interlocking devices was stored as shown in Fig. 14. Furthermore, in the Messages table of NotificationDb.db, system notification message logs and timestamps such as hub connection messages and cleaning start/end messages in Fig. 15 were stored. In addition, user personal information linked to smart things, setting information for each device, and network information were stored.

6.1.3. Classification Based on Data Characteristics

Acquired data have different utilization methods depending on their characteristics. The characteristics of data can be largely divided into three categories: device usage data, user data, and smart home environment data. For device use data, it can be used when it is necessary to identify the device user’s behavior. User data can be used when it is necessary to verify the identity of the device user. Furthermore, for smart home environment identification data, it can be used when it is necessary to understand the smart home configuration environment. Accordingly, in this section, it is shown that data utilization methods according to crime scenarios can be derived according to data characteristics.

In this scenario, it is assumed that an in-house murder case has occurred in a two-person household, and the Samsung SmartThings smart home environment built in advance is investigated to track the alibi of the housemate, who is a prime suspect. The suspect stated that he was outside the home at the time the crime occurred. Previously, smart home components that can be configured in a home were identified through smart home architecture. Because the smart home environment can be used by connecting various IoT devices such as AI speakers, TVs, and sensors around the gateway, identification of components of the smart home environment must be preceded for investigation. Accordingly, in this scenario, the investigator identified and secured the AI speaker, robot vacuum cleaner, smart hub, smart button, and tablet used as a management device linked to SmartThings in the actual field.

The identified components of the smart home environment have different data depending on the type of platform, service, and IoT device. Therefore, it is possible to predict the stored data by performing a smart home functional analysis. Data inference can narrow the range of devices to be acquired during actual data acquisition and can handle the latest IoT devices. The investigators performed functional analysis of IoT devices linked with Samsung SmartThings to derive expected data, and accordingly, data acquisition was performed focusing on AI speakers and robot vacuum cleaners. Investigators found Nand flash memory of AI speakers and robot cleaners, extracted, and analyzed data through chip-off, and obtained some of the expected data by analyzing confiscated smartphone application data.

As mentioned above, the acquired data can be divided into device use data, user data, and smart home environment identification data, and the utilization method is different depending on the data characteristics. Here the primary objective is to track down the suspect’s alibi, device usage data can be used as the center. Accordingly, the investigator indirectly inferred that the suspect used the robot cleaner at home at the time of the crime through the cleaning timestamp and reservation setting time found in the robot cleaner. In addition, through the system notification message log of the tablet that
controlled the smart home, it was confirmed that the smart hub and the robot cleaner generated notifications at the corresponding time. In addition, the user account, name, and date of birth, which can be classified as user data, were collected from AI speakers to determine whether there were accomplices. In addition, by collecting a list of interlocking devices, which is smart home environment identification data, from AI speakers, we confirmed the possibility of securing additional evidence by securing smart lighting and security sensors.

Table 3
Devices and data used in the investigation (scenario)

<table>
<thead>
<tr>
<th>IoT Device</th>
<th>device usage data</th>
<th>Smart home environment identification data</th>
<th>user data</th>
</tr>
</thead>
<tbody>
<tr>
<td>AI Speaker</td>
<td>-</td>
<td>Linked device list</td>
<td>User account, name and date of birth</td>
</tr>
<tr>
<td>Robot vacuum</td>
<td>Cleaning timestamp, schedule setting time, system notification message</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Smart hub</td>
<td>system notification message</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Smart Button</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
</tbody>
</table>

6.2. Xiaomi Mi Home

6.2.1. Guessing Data Through APP Service Analysis

Xiaomi Mi Home also infers expected data that can occur depending on the function through the functional analysis of the smart home app. For Xiaomi, the entire smart home environment is controlled through the ‘Mi Home’ app. Figure 16 is the main function tab of the IP camera. In addition to the main function of video recording, the IP camera provides features such as person detection and replay of previous monitoring records. In this way, the stored data can be inferred as shown in Table 4. This data may be stored in the device and may be stored in other cloud servers and smartphones. This can be identified through actual data analysis.
Table 4
Expected data according to Xiaomi Mi home IoT device type

<table>
<thead>
<tr>
<th>Device</th>
<th>Expected data</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP camera</td>
<td>Recorded video, person detection record, device settings</td>
</tr>
<tr>
<td>Robot vacuum</td>
<td>cleaning record, area cleaned</td>
</tr>
<tr>
<td>Smart TV</td>
<td>TV usage history, third-party app information</td>
</tr>
<tr>
<td>Air purifier</td>
<td>PM2.5, temperature, humidity</td>
</tr>
<tr>
<td>Smart watch</td>
<td>user information</td>
</tr>
</tbody>
</table>

6.2.2. Data Extraction and Analysis

6.2.2.1 Data Stored on the Device

Through device specification analysis, devices with internal storage are IP cameras and smart TVs. For an IP camera, as images were stored in a secure digital (SD) card, data can be extracted through an SD card reader. Because of connecting to the PC through the reader, several jpeg and mp4 files were confirmed as shown in Fig. 17. Through this analysis, it was discovered that the IP camera automatically records a video every minute and stores thumbnail (.jpeg) and video (.mp4) files.

Furthermore, because of analyzing the ‘/MIJIA_RECORD_MOTION/2022101814/motion_record_msg’ file in the SD card directory with Hex Editor, it was confirmed that the time stamp when a person was detected was saved in little-endian as shown in Fig. 18.

For smart TVs, UART pin and Nand Flash memory could be identified, however, data acquisition was infeasible owing to technical limitations.

6.2.2.2 Cloud Server

Similar to other IoT devices, Xiaomi Mi Home's IoT devices communicate with the cloud server via Wi-Fi. Accordingly, communication packets between each device and the cloud server were acquired through packet capture. However, similar to Samsung SmartThings, there were some packets encrypted with TLS, and there was no meaningful data in other packets.

6.2.2.3 Smartphone

The smartphone used in the experiment targeting the Xiaomi platform was also pre-rooted, and the dump image was acquired through ADB. Some db files were stored in the app data of the smartphone. In cn_6607694827_typelist_v2.db, user ID, device notification message, and timestamp are stored as shown in Fig. 19.

6.2.3. Classification Based on Data Characteristics
For the Xiaomi Mi Home, some artifacts were previously obtained from IP cameras and management devices. For the IP camera, videos and photos were stored on the SD card and people detection timestamps were established. Because this data was stored through device use, it can be classified as device use data. Device usage data is important because it can determine the alibi of the person involved in the case. For example, for a house invasion, the automatically recorded video shows the intruder’s face, clothing, and body type, and the time stamp of the person detection shows the time the intruder entered the house.

7. Discussion

In this study, we proposed an architecture that generally expresses the smart home environment from a forensic perspective based on the components of the smart home. The architecture is organized through five layers: cloud, platform, service, IoT device, and device component, and each of these layers is a common element of the smart home environment. The smart home common architecture includes all the components of the smart home and can be the basis for smart home forensic research because it is designed to fit the smart home architecture. In addition, our proposed architecture can be added-on even when new devices, services, and platforms are released, so it can be expanded.

This study proposed a forensic framework through the previously derived smart home common architecture, and the framework consists of three parts. In the function analysis-based smart home data inference step, it is possible to identify the functions supported by the IoT device and the functions of each platform through the functional analysis of the smart home application. Samsung and Xiaomi operate through the SmartThings and Mi home apps, respectively, and both apps include the ability to remotely control linked devices. Accordingly, expected data was derived through functions supported by interlocking devices and platforms. Through this functional analysis, data in the smart home can be inferred, which can help set and plan target devices in the forensic readiness stage.

The inferred data can be stored in the cloud and management devices as well as the internal storage of IoT devices. Accordingly, different data acquisition methods are applied to extract the actual data. When stored in the device’s internal storage, data can be extracted by applying various acquisition methods to devices with internal storage. In this study, data inside the actual device was extracted from Nand Flash and SD cards in Samsung and Xiaomi devices, and data such as user information and linked device list could be identified. Although not performed in this study, it seems that more artifacts can be obtained if deleted files can be recovered through file system recovery of acquired images [29,30]. Other devices such as the Samsung Smart Hub could not verify data owing to disk encryption, and small devices such as smart plugs and smart tags could not obtain data owing to the lack of storage. In addition, devices such as Xiaomi Smart TV could not acquire data due to technical limitations. Currently, numerous data acquisition method studies have been conducted. However, as mentioned above, as new technologies are applied to IoT devices, a new methodology is required through the investigation of SW/HW of IoT devices, which requires additional research on the latest IoT devices in smart homes.
If it is stored in the cloud, it can be verified by analyzing the packet between the IoT device and the cloud server. In this study, packets could not be checked with TLS encryption, and unencrypted packets may not seem important because they are not meaningful data. However, there are cases where artifacts have been identified from packets, which can help with warrant review. In particular, if a certificate can be inserted into the device, tools such as Burp suite can be used to decrypt TLS packets, and data stored in the cloud server can be extracted through a replay attack.

When stored in a smartphone app, data can be extracted by extracting images from smartphones linked with Samsung and Xiaomi applications. The research team confirmed that a large amount of app data is stored in smartphones, but there is a limitation that is difficult to acquire for smartphones at home. However, for smartphones, it is possible to obtain information from IoT devices that lack storage, such as sensors and door locks, therefore, if the smartphone can be acquired in the field, it can be strong evidence.

Finally, the data derived in this manner are used differently depending on the type of crime. Accordingly, the research team defined the data characteristics by dividing them into device use data, user data, and smart home environment data. Device usage data can be used to specify a suspect's alibi, and user data can help identify victims and suspects. In addition, smart home environment data can be used by investigators to accurately examine the smart home environment, enabling efficient investigations.

8. Conclusion

Owing to the development of SW/HW technology, various IoT devices are being released. In particular, advances in HW, such as sensing technology, allow devices to collect various data and provide better quality services. Therefore, home appliances used at home are not only released as IoT devices but also as new devices are developed, to allow users to use convenient functions at home. Accordingly, smart home vendors are expanding the smart home environment through smart home platforms. The smart home that emerged in this manner provides customized services through communication with a cloud server based on IoT devices. At this time, as various data are generated and stored in IoT devices, various studies have been conducted on smart homes from a forensic perspective. However, as the latest devices are released, the scope of smart homes is expanding, and scalable smart home forensic research is required accordingly.

In this study, components and structures of the smart home were identified to derive a forensic method suitable for the evolving smart home environment, and based on this, a smart home architecture that can be commonly applied in the smart home environment was derived. The common architecture of smart home platforms is composed of five layers such as cloud, platform, service, IoT device, and device components, and each of these layers is a common element of the smart home environment. Based on the common architecture of these smart home platforms, a forensic framework for smart home environments was proposed. The forensic framework has three steps, each consisting of predictive data inference, actual data acquisition, and data identification that can be used for criminal investigation. In
this study, Samsung and Xiaomi smart home test beds were established to verify the forensic framework. The expected data in the smart home was identified through the functional analysis of the smart home app, and some of the expected data was obtained and verified for the built Samsung and Xiaomi SmartThings smart home test beds. Afterward, the identified data was classified according to data characteristics such as device usage data, smart home environment identification data, and user data. The smart home composition classification and common architecture of smart home platforms established through this study contain common components in the smart home environment, therefore, they can be collectively applied and utilized in the smart home environment during investigations. In the future, the smart home digital forensic field is of great significance as a study based on. In addition, when an investigator conducts an investigation on smart home IoT devices in the field, it can be used as a reference for use as key evidence by tracking data collection and analysis and user behavior.

Declarations

Author Contribution

S. K: Methodology, Investigation, Formal analysis, Writing - original draft. G. L: Writing - review & editing, Validation. J. S: Writing - review & editing, Validation. I. L: Writing - review & editing, Validation. T. S: Conceptualization, Writing - review & editing, Resources, Supervision, Validation.

Acknowledgments

This work was supported by Supreme Prosecutors' Office of the Republic of Korea.

References

12. Alam MN, Kabir MS (2023), May Forensics in the Internet of Things: Application Specific Investigation Model, Challenges and Future Directions. In 2023 4th International Conference for Emerging Technology (INCET) (pp. 1–6). IEEE

Figures

Figure 1

Structure of Smart Home Ecosystem
Figure 2

Smart Home Common Architecture

Figure 3

Smart Home Forensics Diagram
Figure 4

Forensics Analysis Framework Based on Smart Home Platform Common Architecture
Figure 5

Galaxy Home Mini Function of Samsung SmartThings Smartphones
Figure 6

Nand Flash memory in Galaxy Home Mini PCB

Figure 7

Mounted Galaxy Home Mini Dump Image
Figure 8

Galaxy Home Mini Partition List (FTK Imager)

<table>
<thead>
<tr>
<th>id</th>
<th>name</th>
<th>groupid</th>
<th>type</th>
<th>vid</th>
<th>mmnn</th>
<th>version</th>
<th>nick</th>
<th>color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>spk</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>smsn (1) [0MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>tup (2) [27MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>csa (3) [8MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>ramdisk2 (4) [20MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>ramdisk1 (5) [8MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>param (6) [8MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>cm (7) [4MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>boot (8) [16MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>recovery (9) [16MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>csc (10) [270MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>system-data (11) [158MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>nad_fw (12) [20MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>nad REFER (13) [1MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>boot_debug (14) [5MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>kernel_debug (15) [10MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>user (16) [1844MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>roofts (17) [1306MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>steady (18) [0MB]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 9

'Device_table' table in 'home/owner/apps_rw/com.samsung.lux-st-service/data/sc.db'
Figure 10

'Unique_table' table in 'home/owner/apps_rw/com.samsung.tizen.samsung-cloud/data/.cloud_pdm.db'

Figure 11

'History' table in '/home/owner/apps_rw/org.tizen.ocfd/data/history/history_new3.db'

Figure 12

Hub V3 Partition Disk Encryption for Samsung SmartThings

Figure 13

Packets generated when using AI speakers
Figure 14

‘devices’ table in ‘data/com.samsung.android.oneconnect/databases/CloudDb.db’

Figure 15

‘Messages’ table in ‘Data/com.samsung.android.oneconnect/databases/NotificationDb.db’
Figure 16

IP camera function in Xiaomi Mi Home app

Figure 17

Multimedia files (JPEG, MP4) stored on the IP camera memory card
Figure 18

IP camera people detection timestamp

Figure 19

DB file (cn_6607694827_typelist_v2.db) stored in IP camera memory card