

Functional Classification and Quantitative Analysis of Smart Connected Home Devices

Joseph Bugeja, Paul Davidsson, and Andreas Jacobsson

Internet of Things and People Research Center, Department of Computer Science and Media Technology,
Malmö University, Malmö Sweden,

{joseph.bugeja, paul.davidsson, andreas.jacobsson}@mau.se

Abstract—The home environment is rapidly becoming more complex with the introduction of numerous and heterogeneous Internet of Things devices. This development into smart connected homes brings with it challenges when it comes to gaining a deeper understanding of the home environment as a socio-technical system. A better understanding of the home is essential to build robust, resilient, and secure smart home systems. In this regard, we developed a novel method for classifying smart home devices in a logical and coherent manner according to their functionality. Unlike other approaches, we build the categorization empirically by mining the technical specifications of 1,193 commercial devices. Moreover, we identify twelve capabilities that can be used to characterize home devices. Alongside the classification, we also quantitatively analyze the entire spectrum of commercial smart home devices in accordance to their functionality and capabilities. Overall, the categorization and analysis provide a foundation for identifying opportunities of generalizations and common solutions for the smart home.

Index Terms—Classification; connected home; devices; IoT; smart home; survey; taxonomy; web mining.

I. Introduction

In recent years, the development of Internet of Things (IoT) technologies have contributed to transform the traditional home to a smart Internet-connected home. According to Gartner [1] there are currently 8.4 billion IoT devices as of 2017, which represents a 31% increase from the past year. Inspired by the rapid growth of the consumer IoT market, many commercial stakeholders, have produced a multitude of devices allowing for housework automation, enhanced entertainment, extended personal independence, and more.

As the number and heterogeneity of smart home devices keeps increasing at a fast pace, it becomes increasingly challenging to gain a deeper understanding of the smart home as a socio-technical system. Socio-technical systems combine technical infrastructure, e.g. the house, with human actors, e.g. the householders, who operate (in) this infrastructure and interact with it. Such a space, brings with it complexities especially as it tends to involve multiple fields each bringing its own

terminology, concepts, and assumptions. Consequently, research work in the smart home field is being segmented by multiple disciplines (e.g., networking, ubiquitous, and mobile computing) complicating the efforts towards a common understanding. Gaining a better understanding of the home is essential to build robust, resilient, and secure smart home systems. Thereby, this raises the importance for a logical organization of the devices into a classification. Classification approaches are critical to identify shared characteristics and help provide homogeneity and order within the existing diversity. Additionally, when the classification is used together with a quantitative analysis of the smart home domain this allows for attaining deeper insights into the smart home. For instance, it would help in identifying areas that are growing in adoption rate, new trends, and topics where research attention is needed.

There is little work on smart home taxonomies and most of the attempts are specific to industry solutions or market definitions that are dynamic. Specifically, with the existing classifications we observe four drawbacks. First, their scope tends to be limited to a particular category of devices, e.g., wireless sensor networks. Second, the most cited approaches are based on fixed parameters, e.g., memory constraints, an attribute that is likely to change with more resourceful devices. Third, there is a shortage of taxonomies that focus on actual smart connected home environments. Instead, some classifications rely on abstract models that oversimplify the complexity of real-world deployments. Finally, we observe that no scholarly work has been published that surveys the entire spectrum of commercial smart home devices.

We propose a classification model based on the functionality of smart home devices. Devices' functionality is attributed to the object's designer at the time the device is planned. Unlike other approaches, we build the categorization empirically by mining the technical specifications of 1,193 commercial smart home devices as featured in SmartHomeDB¹. Once classified, these devices are characterized on the basis of their capabilities. Alongside the classification, we analyze the entire spectrum of commercial smart home devices in accordance to their functionality and capabilities. Overall, the categorization and analysis provide a foundation for identifying opportunities of generalizations and common solutions for the smart home.

¹ <https://www.smarthomedb.com> [accessed April 23, 2018].

The remainder of this paper is structured as follows. In Section II we describe the technical composition of a smart connected home device. Next, we summarize related work on classifying IoT-based devices. The adopted research design approach is discussed in Section IV. Following that, the classification of devices and analysis results are presented in Section V. A reflection on the achieved results, possible uses, and a comparison of the proposed classification with the existing work is laid out in Section VI. Finally, in Section VII conclusions are drawn and directions for future work are specified.

II. Smart Connected Home Devices

A typical smart connected home setup consists of a multitude and heterogeneous Internet-connected devices.

Essentially, a smart connected home device is composed of a number of hardware elements. These include: input/output interfaces (e.g., the serial communication protocol Universal Asynchronous Receiver/Transmitter), memory/storage interfaces (e.g., flash storage), audio/video interfaces (e.g., HDMI), and networking protocols (e.g., Ethernet) allowing for connectivity to the home network and to the Internet [2]. Depending on its nature, a device may also be powered by a battery pack (e.g., 4xAA batteries), and it may comprise integrated sensors/actuators that receive information from the environment and use it to control it. Typically, home devices tend to be stationary (e.g., cooker), but can also include mobile and robotic appliances (e.g., vacuum cleaners). Schematically, Figure 1, depicts the capabilities of smart home devices.

Devices also support a number of software interfaces. In particular, those with high hardware specifications, e.g., smart home gateways, may offer: Application Program Interfaces (APIs) allowing for integrating the functionalities of a device, IFTTT (if this then that) allowing users to set automatic command generation in response to environmental changes, and native services accessible typically through a smartphone application or a Web browser. Additionally, devices may offer cloud/server support allowing for storage of data and communication with external service providers, and protocols that allow for remote operation and control.

In this work, we use some of the aforementioned capabilities to survey the smart connected home environment.

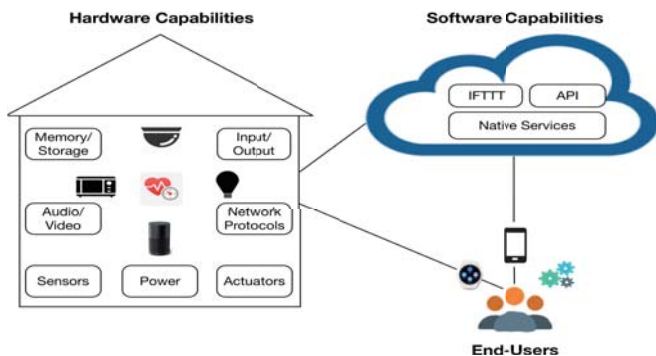


Fig 1. General description of a smart home device capabilities. An example of this could be a smart home consisting of an Internet-connected camera. This camera has integrated motion sensors, onboard storage, and networking protocols allowing it to be managed, by the householders, over the Internet, typically through a smartphone application.

III. Related Work

Several efforts have been made to develop a classification model for IoT devices as identified by Yoon et al. [3]. Below, we expand on the mentioned approaches and discuss their relevance to the smart home.

A. Classification based on functionality

The International Telecommunication Union (ITU) [4] classifies devices into four categories: data-carrying device, data-capturing device, sensing and actuating device, and general device. Data-capturing devices interact with physical objects through data-carrying devices (e.g., active RFID tags) or data carriers (e.g., bar codes) attached to physical objects. Sensing and actuating devices, e.g., smart meters, interact with physical objects in the environment. General devices, e.g., smartphones, have embedded communication and data processing capabilities. The ITU classification is generic and can be applied for classifying different IoT devices. Nonetheless, it fails to represent certain home devices, e.g. smart TVs, that embed sensors and function as well as gateway devices.

B. Classification based on data

Imagination Technologies Limited [5] classify IoT devices according to the type of content/data being managed. The classification includes three data types: machine to machine (M2M) data (e.g., sensor hubs), audio (e.g., wireless audio), and audio/video (e.g., connected cameras). The proposed classification can be used for profiling devices by different parameters such as power, CPU, and operating system requirements. However, its usefulness is limited for our classification requirements. Notably, as it ignores other data types, e.g., those related to movement information, that are often used for enhancing security/safety of householders. Furthermore, the adopted research methodology is not evident in the paper.

C. Classification based on resources

Bormann et al. [6] in RFC 7228 classify devices based on their degree of resource constraints, notably RAM and ROM requirements. The classification identifies three distinct classes of devices: Class 0 (C0), Class 1 (C1), and Class 2 (C2). C0 devices are the most constrained, while C2 devices are the least constrained supporting most of the existing protocols stacks. This classification introduces a systematic and precise approach to differentiate between different devices. However, it ignores resourceful devices, that arguably occupy a relatively large proportion of the smart home.

ENISA [7] extends RFC 7228 [6] to cover also high-capacity, resourceful, devices (i.e. devices beyond C2). The classification is intended to be used as a tool for understanding the security of the smart home environment. Alas, our aim is to have a generic classification that groups devices based on their functionality as the distinguishing factor.

D. Classification based on interactions

Moawad et al. [8] classify IoT objects into four levels (Level 0 – Level 3) depending on their abilities to interact with other objects. Level 0 and 1 objects, only receive or send information,

respectively. Level 2 can perform both operations with one object, while Level 3 extends the interaction to any other object. This taxonomy is useful as a basis for formalizing information flows in different use-cases. Nonetheless, it ignores parameters, such as protocol types and availability of sensors and actuators, that are useful for comparing different devices and understanding their effective capabilities.

E. Main observations

The existing classifications vary in terms of their coverage, scope, and method used to build the device hierarchy. We observe that the majority of the reviewed approaches are conceptual models that ignore the complexity and heterogeneity of the consumer IoT market. This limits their application to model real smart home deployments. Moreover, we observe that the reviewed work tends to be focused towards machine interpretable categorizations limiting their effectiveness when it comes to using them as an analysis tool oriented towards the human as the end-user. Finally, we note that the functionality and capabilities of devices necessary to describe and compare the technical dimensions of a device are not elaborated in previous work. Recognizing this, we develop a human-oriented classification, and conduct a survey of the commercial smart home domain.

IV. Research Methodology

The adopted research design method is detailed in the following sections.

A. Data collection

There are two main IoT product collection platforms, namely, *iotlist.co* and *smarthomedb.com* (SmartHomeDB). The platform, *iotlist.co*, is an online directory listing IoT devices on the market. This is not restricted to household items and does not detail the technical capabilities of devices. SmartHomeDB is an online platform focusing on smart connected home devices. Furthermore, it enlists the technical description of products, and has an open community-supported database. For these reasons, we use SmartHomeDB as our data-source.

At the time of our evaluation (as of May 2017) it consisted of a total of 1,193 IoT products, grouped into 87 different categories (each category representing a device type, e.g., kettle). Specifically, we crawl the entire database by following hyperlinks found within the 'Products' section of the database. Each webpage represented an actual device description. Here, *Bash* shell scripts were utilized for their efficiency. In particular, *curl* tool was used to download the content of each HTML page.

B. Data processing

In the data collection stage, we rely on web mining to extract the smart home devices' capabilities. Web mining is a process of analyzing Internet resources to extract useful information. Here, we extracted segments from each webpage that were related to each of the capabilities. In doing so, this involved processing the HTML text using standard data cleaning functions. Specifically, this involved tasks such as removing punctuations, converting uppercase to lowercase, and conflating whitespaces. Only the paragraphs that had at least one keyword related to a

particular capability were kept. In the actual implementation, *Python* programming language, and an open source framework – *scrapy* – was mainly used to extract data.

Once text preprocessing was completed, the output data, was normalized into the $[0, 1]$ range, with 1 indicating that the device supports a capability, and 0 otherwise. Each device was represented by a binary vector of size n , with n being a constant equal to the total number of surveyed capabilities ($n = 12$). The capabilities analyzed are identified in Section V.

The methodology of selected keywords was largely manual. Here, we looked at a number of device descriptions and identified frequent keywords pertinent to each capability. The majority of the capabilities were consistently identified with one keyword. However, for the rest, in particular to identify sensors and actuators, additional keywords, including a list of stems and synonyms of each were added. For example, we generated other forms of the word "speaker" (e.g., "speakers", "amplifier", and "sound") for the actuators capability. The output of the data processing stage consisted of a comma-separated value (CSV) file.

C. Data analysis

In preparation for the data analysis stage, the CSV file was loaded into IBM SPSS software. SPSS is widely used for statistical analysis. In our case, we used it for all the survey data analysis. The analysis followed a two-step approach.

First, a clustering variable was manually added to the dataset. This clustering variable represented the primary functionality the particular device implements. In our context, functionality signified the designed purpose of the device, e.g., that of providing 'Security and safety'. The main categories of functionality are based on existing scientific literature on smart homes, e.g., [9]. However, additional categories and further decomposition into sub-categories was manually done after investigating each device type. Example, 'Human-Machine Interface' was added as a functional category to cater for newer device types such as 'Voice Command Devices' (also known as 'Digital Voice Assistants').

The second stage involved the actual running of descriptive statistics on the dataset. Here, results were grouped by their clustering variable.

V. Results

In this section, we present the results achieved following the application of the research methodology in the form of a taxonomy and an accompanying quantitative analysis.

A. Functional classification

In this section, we classify devices based on their functionality, and identify their overall distribution (%) within the master dataset as follows (Figure 2):

Energy and resource management (32.6%): Smart energy and resource management hardware that track and adjust the indoor environment, e.g., in terms of consumption patterns of electricity, and accompanying actuators/motors. Plugs are the most frequent devices occupying 19% of this category.

Entertainment systems (16.2%): Smart entertainment and multimedia applications that provide on-demand access to

audio/video feeds from traditional broadcast channels, Internet, and enabled storage devices. Audio speaker devices represent the most common devices taking 51.3% of this category.

Health and wellness (2.7%): Smart healthcare, ambient assisted living, and wearable technologies that are mainly used for monitoring and improving the health and well-being of the householders. Scales are the prevalent devices representing 62.5% of this category.

Human-machine interface (5.0%): Devices in this category include intelligent (digital) personal assistants, handheld remote appliance controls, key fobs, and generic user-interaction points such as smartphones, tablets, and specialized terminals. Remote control devices represent the most popular devices in this category taking 85%.

Household appliances and kitchen aids (5.3%): Appliances that assist in home automation and household management

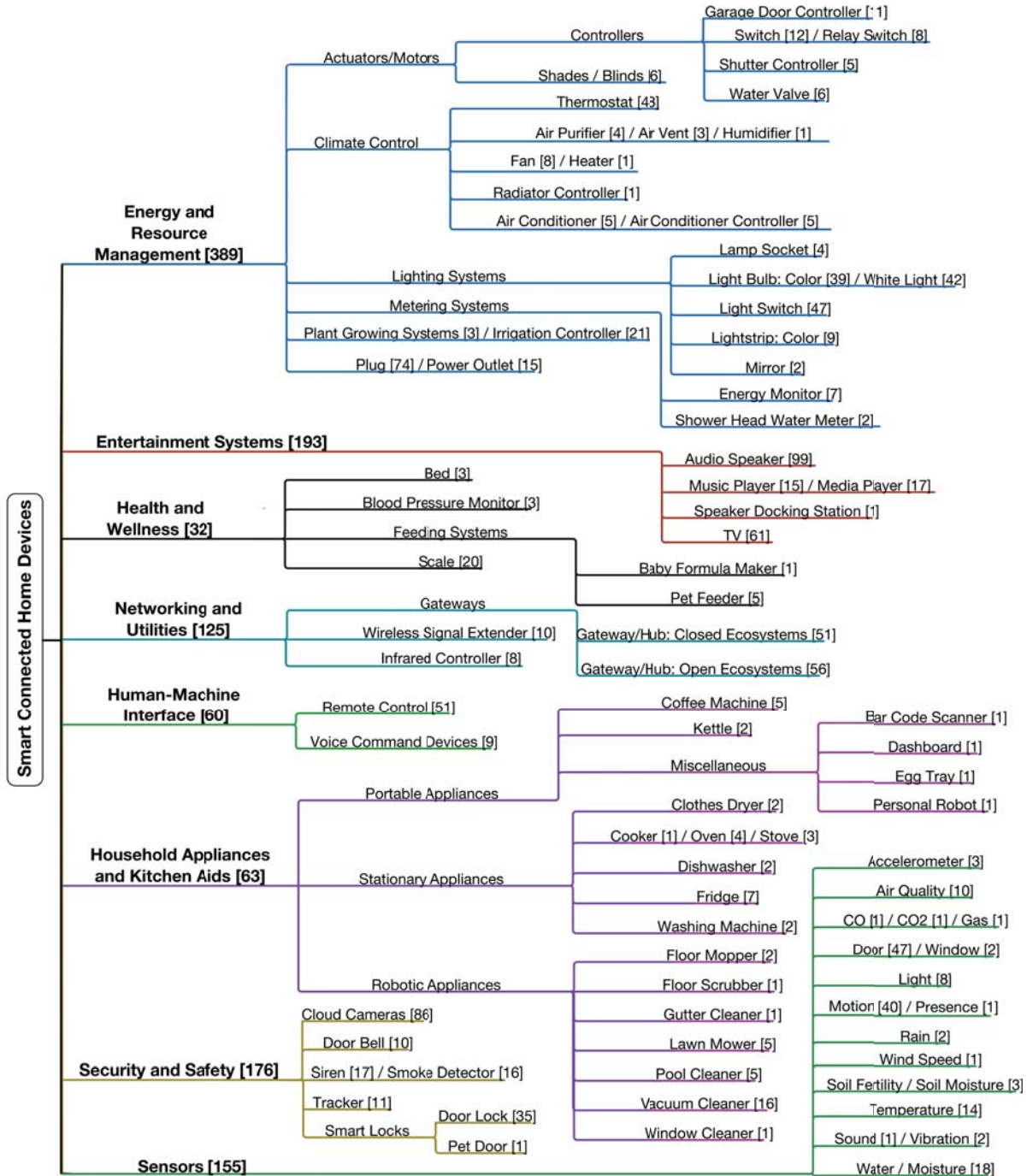


Fig 2. Hierarchical taxonomy of smart connected home devices. Each category represents actual devices that are available in the smart home market. The number enclosed in square brackets (“[]”) represents the total number of devices present in the master dataset for that item. To increase clarity and readability, the taxonomy adopts the same terms for item names as those used by SmartHomeDB. While the identified main categories are related to application areas, ‘Sensors’ represent standalone hardware units or components that can be embedded in different device types.

tasks, such as monitoring available food supplies, and assisting in housework chores such as cooking. Vacuum cleaners signify the most frequent device occupying 25.4% of this category.

Networking and utilities (10.5%): Equipment in this category facilitate communication and integration between different IoT devices and users. Open ecosystem gateway/hub devices are the most popular in this category taking 44.8%.

Security and safety (14.8%): Smart surveillance and security applications address matters related to the safety of a house and its occupants. Cloud camera devices are the most present devices occupying 48.9% of this category.

Sensors (13%): Devices that measure features of interest, typically environment parameters, e.g., temperature. Door sensors are the most frequent devices occupying 30.3% of this category.

B. Capabilities of smart connected home devices

In this section, we describe generic properties (capabilities) for characterizing smart connected home devices, group them according to their corresponding functional area, and calculate their overall support (%) within the entire smart home domain:

Device related capabilities:

Gateway functionality (67.9%): Identifies whether the device can function as a standalone, i.e. without requiring additional components, in particular third-party gateways or bridges.

Battery source (38.6%): Identifies whether the device relies on a battery source as its main energy source.

Integrated actuators (32.7%): Identifies whether the device features integrated actuators, e.g., a door electric motor.

Integrated sensors (34.5%): Determines whether the device includes integrated sensors, e.g., motion sensors.

Communication related capabilities:

Wireless protocols (91.5%): Determines whether the device supports radio protocols, e.g., Wi-Fi, to transfer information.

Wired protocols (23.1%): Specifies whether the device support wired protocols, e.g., Ethernet, to transfer information.

Cloud/Server (66.2%): Determines whether a smart home product is sending data to an external server, commonly managed by the manufacturer or service provider.

Software related capabilities:

API (2.9%): Programming instructions and standards that software developers can design software which integrates the functionalities of a smart home device.

IFTTT (12.8%): Online service that enables users to connect different web applications and smart home products together through conditional statements (“recipes”).

Accessibility related capabilities:

Smartphone accessibility (69.6%): Identifies whether there are software applications (apps), e.g., Android apps, that can be downloaded/accessed for controlling a smart home device.

Web browser accessibility (11.6%): Determines whether there is an open webpage for managing a device through a standard web browser.

Remote access (65.5%): Identifies whether an entity is able to control the smart home device even when not at home.

C. Quantitative analysis of devices’ capabilities

TABLE I. Matrix showing the percentage distribution of supported capabilities for each smart home category: E&RM, energy and resource management; E, entertainment systems; H&W, health and wellness; H&K, household appliances and kitchen aids; N&U, networking and utilities; S&S, security and safety; HMI, human-machine interface; S, sensors.

Capability	E&RM	E	H&W	H&K	N&U	S&S	HMI	S
Device								
Gateway functionality	51.7	99.0	100	100	98.4	71.6	51.7	27.7
Battery source	10.8	40.9	78.1	49.2	13.6	49.4	76.7	86.5
Integrated sensors	20.8	0.5	84.4	3.2	8.8	70.5	16.7	100
Integrated actuators	48.1	51.8	0	15.9	15.2	38.6	8.3	0.6
Communication								
Wireless protocols	91.0	97.4	100	63.5	96.0	93.8	91.7	88.4
Wired Protocols	3.6	82.4	0	0	55.2	17.6	3.3	0.6
Cloud/Server	77.1	12.4	68.8	49.2	75.2	90.9	30.0	91.0
Software								
API	3.6	0	6.3	3.2	5.6	2.3	1.7	3.2
IFTTT	15.4	0	12.5	7.9	21.6	15.9	15.0	12.9
Accessibility								
Web browser accessibility	12.6	0	18.8	0	22.4	9.7	18.3	17.4
Smartphone accessibility	63.2	95.3	84.4	57.1	76.8	69.9	50.0	56.8
Remote access	78.7	3.6	65.6	49.2	79.2	90.9	26.7	91.0

All smart home categories support wireless channels, but wired protocols are not supported in the ‘Health and wellness’ and ‘Household appliances and kitchen aids’ domain. Cloud/server support is prevalent in the ‘Security and safety’ and ‘Sensors’ category.

All devices in the ‘Household appliances and kitchen aids’ and ‘Health and wellness’ category function as standalone. Actuator support is mostly provided in the ‘Entertainment systems’ category. Excluding the ‘Sensors’ category, sensors are most present in the ‘Health and wellness’ class. 732 devices in the dataset are primarily powered by an A/C power cable/adaptor or mains supply. This corresponds mainly to devices ‘Energy and resource management’ devices.

IFTTT support is mostly supported in the ‘Networking and utilities’ category at 21.6% and being mainly utilized by Gateway/Hub devices. API support is overall the least popular capability. This is mostly used at 6.3% in the ‘Health and wellness’ category specifically in scales.

Accessibility through a smartphone application is more widespread than web browser accessibility. Benefiting the most from smartphone accessibility are ‘Entertainment systems’. Web browsers are most popular in the ‘Networking and utilities’ category. Remote access is supported by all categories, however it is least supported in ‘Entertainment systems’.

Table I summaries the surveyed devices’ capabilities across each category identified in Section V.

VI. Discussion

To substantiate our proposed categorization to the smart home environment, we apply the existing classification approaches in Section III to the processed master dataset.

Using the functionality-based approach data-carrying/data-capturing, sensors/actuators, and generic devices occupy about

13%, 17%, and 70%, respectively. Thus, we are able to classify all devices, but most of them fall into the generic device category. Using the data-based approach we observe that audio, audio/video, and M2M data take respectively, 13%, 22%, and 22%. Thus, more than 40% of the devices do not belong to any of the categories, and thereby cannot be represented directly with this approach. Moreover, using the resource-based approaches, we calculate that constrained devices occupy circa 30% of the database. With regards to the interaction-based approaches, we conjecture that with IFTTT and other technologies such as Stringify, devices will be capable of supporting advanced interaction possibly at the highest levels.

From this mapping, we observe that the current methods fail to distinguish between different high-capacity devices, e.g., cloud cameras, and in identifying helper components such as wireless signal extenders. Our classification addresses these gaps by focusing on the functionality of devices, and by characterizing devices further on the basis of their capabilities. Furthermore, we ensure applicability to the smart home domain by building the categorization directly from the mined data.

In terms of smart home device capabilities, we focused on properties that were present or that could be computed from SmartHomeDB. In relation to this, we observe that to build a comprehensive capability set that might be needed for conducting a rigorous security and privacy analysis, other capabilities may be required. Example, hardware capabilities related to storage and computation, and software capabilities such as operating system. Alas, these were not present in the utilized dataset.

The classification model has different uses. For the end-users and industry, the model can help enable quick understanding, differentiation, and identification of smart home devices characteristics and differences. Indeed, similar to the use-case proposed by Cheekiralla et al. [10] for wireless sensor network devices, the proposed categorization can be used to describe and compare home devices, on the basis of their functionality and capabilities as their distinguishing factors. Smart home developers can leverage the proposed model also as a guideline for improving certain properties of technologies. Moreover, end-users, e.g., householders, can use it as a catalogue of possible devices that could be used to implement their desired smart home functionality. In terms of scholarly work, the categorization provides a foundation for organizing devices in a logical and coherent manner. This can be used as a basis for identifying opportunities of generalizations and common solutions for the smart home. Common to all, the categorization serves as a tool to help bridge the communication between researchers and practitioners in the smart home development area. This is especially relevant since there are multiple disciplines originating from the IoT paradigm but as well from the commercial side, involved in the smart home.

The followed empirical approach leads us to have a more comprehensive account of the smart home. Nonetheless, we observe that there are multiple approaches for organizing devices. In our case, we adopted a design that is based on our systematic observation and interpretation of the smart home based on SmartHomeDB and guided by functionality as the driving factor. While the classification is focused on the smart home

domain we envisage that it can be possibly extended or adapted for other IoT applications, e.g., for IoT-based smart buildings, that share similar technical characteristics to the smart home.

VII. Conclusions and Future Work

The growth and heterogeneity of smart connected home devices alongside the projected market estimates raises the importance of an analysis of the current technological domain and a classification model that organizes devices logically.

In this paper, we presented a generic and novel classification of home devices based on the devices' functionality and their technical capabilities. The classification consists of eight distinct functional categories, and twelve unique capabilities. The categorization is constructed empirically by mining the technical specifications of 1,193 commercial devices.

Alongside the classification, we also analyzed the entire spectrum of commercial smart home devices in accordance to their functionality and capabilities. Together, the classification and analysis provide a foundation for identifying opportunities of generalizations and common solutions for the smart home.

As part of future work, we intend to use the classification to identify representative devices. These devices can then be investigated in detail, e.g., in terms of their collected data and associated information flows. A next potential avenue for future research is to show devices that are able to integrate and share information and instructions with each other. Such contribution is especially helpful for do-it-yourself (DIY) users and developers of smart home products and systems. Finally, we plan to extend this work and use it as a basis for conducting a quantitative risk assessment of the smart connected home.

Acknowledgment

This work has been carried out within the research profile "Internet of Things and People", funded by the Knowledge Foundation and Malmö University in collaboration with 10 industrial partners.

References

- [1] Gartner, "Gartner Says 8.4 Billion Connected 'Things' Will Be in Use in 2017, Up 31 Percent From 2016," 2016 [Online]. Available: <https://goo.gl/HKtLyP>.
- [2] P. P. Ray, "A survey on Internet of Things architectures," *Journal of King Saud University-Computer and Information Sciences*, 2016.
- [3] S. Yoon, J. Kim, Y. Jeon, "Security Considerations Based on Classification of IoT Device Capabilities," *Service Computation 2017, In The Ninth International Conferences on Advanced Service Computing*, 2017.
- [4] ITU-T, "ITU-T Rec. Y.2060 (06/2012) Overview of the Internet of things," pp. 1-22, 2013.
- [5] Imagination, "Internet of Things - Opportunities for device differentiation," pp. 1-11, 2015.
- [6] C. Bormann, M. Ersue, and A. Keranen, "RFC 7228: Terminology for Constrained-Node Networks," 2014.
- [7] ENISA (The European Network and Information Security Agency), "Security and Resilience of Smart Home Environments," 2015.
- [8] A. Moawad, et al., "Introducing conviviality as a new paradigm for interactions among IT objects," In *Workshop on AI Problems and Approaches for Intelligent Environments*, vol. 907, pp. 3-8, 2012.
- [9] T. D. P. Mendes et al., "Smart home communication technologies and applications: Wireless protocol assessment for home area network resources," *Energies*, vol. 8, no. 7, pp. 7279-7311, 2015.
- [10] S. Cheekiralla and D. W. Engels, "A Functional Taxonomy of Wireless Sensor Network Devices," In *The Second International Conference on Broadband Networks*, pp. 949-956, 2005.