Challenges in the Data Collection for Diagnostics of Smart Buildings

Sanja Lazarova-Molnar and Nader Mohamed

Abstract The rise of smart buildings, i.e. buildings equipped with latest technology and built according to cutting-edge architectural advances, implies increased buildings' complexity. For this reason, both new and retrofitted buildings are often susceptible to new and unforeseen faults, whose timely detection and servicing can significantly affect buildings performance. Many Fault Detection and Diagnosis (FDD) methods are data-driven, where the quality of collected data can significantly affect the accuracy of results. However, data collection for FDD of buildings is a challenging task as needed data is not typically readily available. In this paper we focus on the data collection for FDD of smart buildings. This forms the motivation of this paper, i.e. to identify the challenges that relate to data collection processes for FDD of buildings, as well as propose workarounds of how to tackle the more important ones. Furthermore, we also look into how new technologies can be useful for this goal.

Keywords Smart buildings · Diagnostics · Data collection · Challenges

1 Introduction

Smart buildings are slowly becoming reality. The definition of a smart building can be very broad and varied, and it can be stretched to include various aspects of buildings. We define smart buildings [1] as buildings that have been automated

S. Lazarova-Molnar(✉)
Center for Energy Informatics, University of Southern Denmark, Campusvej 55, 5230 Odense, Denmark
e-mail: slmo@mmmi.sdu.dk

N. Mohamed
Middleware Technologies Labs, P.O. Box 33186, Isa Town, Bahrain
e-mail: nader@middleware-tech.net

K.J. Kim and N. Joukov (eds.), *Information Science and Applications (ICISA) 2016*,
Lecture Notes in Electrical Engineering 376,
DOI: 10.1007/978-981-10-0557-2_90

and retrofitted to reduce their energy footprint and CO_2 emissions without compromising the comfort of occupants [2]. In general, buildings can become smarter in two ways: 1) by using intelligent ICT solutions, or 2) by retrofitting the building using new and better materials or constructional solutions. Usually, to optimize and better meet the preset performance targets, a combination of both approaches is deployed. In this context, in one of our previous works we emphasize the need of providing assessable solutions for smart buildings, as currently it is very difficult to predict the effect of any of these potential building improvements [3]. Due to the significant amount of technology involved in making buildings intelligent, as well as the use of new and better materials, which have, however, not yet stood the test of time; the likelihood of components and subsystems failing, sometimes with hard to predict consequences, has been increasing as well. This represents both a challenge and an opportunity to save on the energy-related costs. Correct and timely attention and intervention to faults in buildings presents a significant opportunity to save on energy related costs. The estimate is between 15% and 30% of the energy consumption cost [4]. Due to the high complexity of buildings, most of the Fault detection and diagnosis (FDD) methods are data-driven, or a hybrid between data-driven and model-driven methods. This implies that the data collection processes have a significant position for the accuracy of the methods. However, the data collection processes that target data relevant to diagnostics, as our experience has shown, are far from trivial, and it is a real challenge to gather it.

In this paper we focus on the main data collection issues that relate to Fault Detection and Diagnostics (FDD) of smart buildings. The goal is to identify the main streams of data, and associated challenges with each of them. Furthermore, we also propose workarounds to some of them, as well as a vision on how new technologies can benefit these processes. We begin with a brief overview of existing FDD approaches for buildings.

2 Fault Detection and Diagnostics for Smart Buildings

As previously stated, Fault Detection and Diagnostics (FDD) is a burning issue for smart buildings. It has been shown that by deploying automated FDD, the operation cost of buildings can be significantly decreased [4]. These facts have triggered a significant amount of research in the field of FDD for buildings. In the following we review the most significant findings in the problem area.

A thorough overview of the FDD approaches for smart buildings is presented in [4] and, consequently, in [5]. The overview, although somewhat dated, is still the most recent thorough overview of FDD methods for building systems, and as such provides a broad classification of them. According to it, the main categorization of FDD methods is in Quantitative Model-Based, Qualitative Model-Based and Process History Based methods. The main strengths of the quantitative methods is that they are based on sound physical models, however, their complexity is usually very large, and are often deemed as intractable. The high complexity implies that very often, for FDD purposes, it is resorted to simplified models. The process history based methods are typically based on machine learning algorithms and to a large extent their quality depends on the quality of collected data.

In a more recent, however less exhaustive overview of methods implemented in tools addressing the problem of diagnostics of buildings [6], authors also point out the main challenges associated with the process of detecting and isolating of errors, summarized as follows:

- Cost and ease of developing building models,
- Scalability and portability of diagnostic solutions,
- Inability to capture relevant data, and
- Lack of knowledge about the data generated by building automation systems.

To summarize, based on our findings and literature survey, diagnostics methods for buildings, in general, can be categorized as: Data-driven methods, Model-driven methods and Hybrid method that combine the former two categories. Model-driven diagnosis methods for buildings are methods that have been developed solely on basis of physical models, where relations are strictly quantitatively described. Data-driven diagnosis methods derive the relationships and predictive models based on historical and ongoing data collection. Hybrid methods contain elements from both model-driven and data-driven methods. In the following we highlight some of the more relevant methods in all three categories.

Data driven diagnosis methods typically deploy machine learning algorithms, which based on collected data from energy consumption and faults, can detect and diagnose faults in real-time. One data-driven approach is presented by Namburu et al. [7], where the authors present an approach based on support vector machines, principal component analysis, and partial least squares to isolate faults. In one of the latest works in this problem domain by Fan et al. [8], a framework for knowledge discovery in Building Automation System data for the purpose of diagnosis is presented. A number of techniques are being deployed, among which genetic algorithms and quantitative association rules mining.

Model-based diagnosis mostly relies on sound physical models that accurately describe and quantify the relations of control and output parameters of subsystems. Typically, model-based approaches impose a number of simplifying assumptions on the system, one of which is the deterministic behavior. O'Neill et al. present a full model-based real-time energy performance monitoring and diagnosis system, which has also been implemented and deployed [9]. The system has identified opportunities for saving ca. 30% of energy cost. However, it has also encountered a number of shortcomings, among others also that the effort needed to calibrate the model was significant.

One hybrid method, that combines model-driven and data-driven approaches, is presented in the work by Du, Jin and Yang, in [10]. There, the authors describe a fault diagnosis approach of sensors for temperature, flow rate and pressure in variable air volume (VAV) systems based on wavelet neural network, which is a combination of wavelet analysis and neural network. The authors claim that the combination of both approaches has significantly improved the accuracy, as compared to each of the methods separately. With similar results, Li and Wen [11] have applied a combination of Principal Component Analysis (PCA) and Wavelet transform (as pre-processing step to PCA) for fault detection in air handling units (AHU), which has yielded better results than applying the conventional PCA-method.

One significant aspect of fault diagnostics of buildings is the participation and feedback of occupants, also as part of the data collection processes. There is a significant amount of research that shows that occupants' feedback can provide a meaningful clue to the diagnostics puzzle. In one of our previous works, we detail the effect that occupants can have on building's performance and the opportunity that lies in them [12]. Goins and Moezzi demonstrate how an occupants' complaints handling process can contribute towards diagnosing performance problems [13]. There, a complaint is defined as "statement that a condition is unsatisfactory or could be improved". The results of the study which confirm the links between the occupants' comfort and building's performance are certainly encouraging to utilize this kind of data, besides the standard automatically metered, to enhance the accuracy of the diagnostics processes. In the following section we identify and present the challenges in the data collection processes for FDD with respect to smart buildings.

3 Data Collection for FDD of Smart Buildings

Data collection is apparently a significant part in every research project on diagnostics of buildings. However, we were not able to find sufficient number of resources on this specific topic. Among the few works that have been published, we have identified a slightly relevant Master thesis [14], where data collection for diagnostics in general has been studied. The thesis was aimed at minimizing the cost by facilitating automatic analysis of diagnostics data. The importance of the data collection for FDD of buildings stems from the complex nature of buildings, where relevant data is linked to both people and entities of technical nature. This brings a lot of challenges, as people are often difficult to cooperate with, so on many occasions non-intrusive data collection is the only option. Furthermore, unlike collection of data for forecasting behavior or performance of buildings, FDD is a problem that goes deeper, as its goal is to discover a cause for a certain failure or malfunction, which sometimes can be unseen before. This fact implies that both data and extensive meta-data is needed to capture and identify the "why?". In [15], Schumann et al. identify three main streams of data in buildings:

1) Metered data, obtained from meters and sensors,
2) User feedback, obtained either through real-time feedback or surveys, and
3) Expert knowledge, formalized in a manner that can be utilized by the selected method, most often in a form of "if-then" rules

Common sources of automatically collected data in a building, as specified by the first stream, are: HVAC, Refrigeration, Power Meter, Utility, Lighting, Financial, Transaction, Occupancy, and Weather. More on the various types of sensing systems and their categorization can be found here [16]. However, in order to support the machine learning methods, there is a need of event logging of failures and other rare occurrences, which can only be properly collected by building's support staff. Our experience has been that this would be very difficult to introduce, as building management staff would preferably spend their time on actually correcting faults, as opposed to reporting about them. Nevertheless, this defines a vital stream of data for FDD that we term as:

4) Event log, i.e. log of failures and other relevant events (e.g. reconfigurations of systems or subsystems).

To support the FDD processes, various alternative sources could also provide useful data, e.g. anonymized data on Internet usage of occupants, or the number of wireless units connected to an access point could both provide useful insights. For this reason, we identify a fifth stream of data as:

5) Implicit Data, i.e. data obtained from various implicitly relevant sources.

In [17], Middelkoop in a chapter on Data Collection for Buildings' Diagnostics has identifed a set of extensive guidelines for high resolution data collection. In the following we summarize a significant subset of those:

- Detailed and structured logs must be made, implying that every historical event needs to be archived along with its meta-data,
- Points must be properly named and tagged,
- Data should be push-collected based on change of value (CoV), this is to decrease the burden of stress on both network and systems,
- Points must be properly configured (Names, Units, CoV), meaning that all devices need to be configured with proper unites and scales, as well as how and when a CoV occurs,
- Points must be verified, thus ensuring that point names and descriptions match,
- Sensors must be calibrated,
- Changes in hardware/configuration/calibrated must be logged and have immediate meta-data updates, this is especially important for diagnosing misconfigurations,
- Meta-data updates must be additive and not replace old data, as all data is relevant for diagnosis.

In the same book [17], in another chapter the following additional issues have been identified:

- Conflicting and inconsistent tagging of data,
- Fragmented data, due to using too short intervals that lead to memory overruns and premature termination, or due to improper and ineffective data management of file maintenance, and
- Presence of noise and its elimination.

All of the afore-mentioned guidelines and problems identify some of the challenges that we have been facing as well. Apparently, some of these issues can be mended by using data analytics tools, but even these tools would certainly be more efficient if parts of the problems are tackled during the data collection processes. Additionally, we were able to also identify the following challenges:

- Necessity of a formal assessment of quality of available data for FDD of buildings, and
- Formal way of blending in expert knowledge for FDD purposes.

Table 1 Relevant data streams for FDD of buildings and related challenges

Data Strea m	Metered Data	User Feed-back	Expert Knowledge	Event Log	Implicit Data
Challenges	- availability and collection of meta-data - tagging of data - calibrating of sensors and meters	- intrusive - incon-sistent - subjective	- intrusive - could be subjective - difficult to formalize	- intrusive - difficult to determine how to de-scribe events	- easily avail-able - accessibility and privacy issues

To summarize, in Table 1, we present the five relevant data streams, along with the challenges associated with each of them. All of the data streams can be utilized to validate each other, as in theory they need to be consistent with each other. This validation alone can also signal anomalies, and often it is an important step in the FDD processes. In the following section we focus on the non-intrusive event logging, as this has appeared to be a significant challenge for the FDD oriented data collection.

4 Non-intrusive Event Logging as a Challenge

One of the main challenges in FDD-related data collection, as we have pointed out in the previous Section 3, is the lack of historical data on failures and other relevant events, as well as the lack of will for providing it. Our workaround for this issue is to simulate faults when the building is not in full use (the time period would depend on the type of building, i.e. for schools it could be summer vacation, or weekends for office buildings, etc.). This could certainly be helpful for a certain types of faults, i.e. misplaced or non-operating sensors. Furthermore, this approach would further trigger another question: "*What is an optimal testing scenario that would yield the most useful data?*", i.e. how to get most useful data with the lowest number of simulated faults in a shortest time. Apparently, this would depend on the nature of the faults and one would need to incorporate domain expert knowledge when developing the faults simulation scenario.

Another approach that we intend to attempt, although time-consuming, is to browse through purchase data for a given building. Typically, there we could also obtain partial data, even though slightly inaccurate, for purchasing of spare parts, as well as payments for repairs and maintenance. This should provide a better insight into causes of anomalies when no more adequate data is available. We are also going to investigate the options of having these processes automated.

The problem of inadequate event logging we aim to also tackle by crowdsourcing with a user-friendly mobile app that would require minimal effort to report, and would rely on an indoor positioning system. We envision it as a smart application that would utilize historical data to guess what a user would be reporting based on a number of parameters, such as location, time of day, type of event in the room, etc. in order to minimize occupant's effort to report. Apparently it would also need to deal with natural language processing for free-text or free-speech.

The combination of all of these approaches should provide useful data to the FDD methods. The bright side is that along with the new technologies for enhancement of buildings, there are also new technologies being developed that could also support the data collection processes. In the following section we provide an overview of this, as well as a vision of how these new technologies could be utilized for this purpose. We especially focus on how this data could be collected in a non-intrusive manner.

5 New Technologies in Support of the Data Collection Processes

One of the traditional approaches used to enhance reliability of engineering systems is the use of redundant resources to provide reliable and fault tolerant operations. This approach can be used to have more accurate collected information in smart buildings for more reliable diagnostics. For example, multiple redundant sensors can be installed to collect more data that can be used to compare readings and validate the accuracy of the current measurements and situations in smart buildings. In addition, they can be easily used to detect faults. However, this approach can significantly increase the cost of monitoring and control systems in smart buildings.

New technologies such as the Internet of Things (IoT) [18] can effectively contribute to solve some of the issues in the diagnostics processes of smart buildings. The IoT aims to interconnect our everyday life devices such as smartphones, smart watches, thermostats and sensors [19]. It provides them with information processing capabilities to enable computers to sense, integrate, present, and react to all aspects of the physical world. The IoT can enable plug-n-play capability for FDD systems. This facilitates fast deployment of different devices for FDD systems in a cost-effective manner. In addition, this enabling mechanism can be used to deploy temporary systems with extra devices for fault detection and diagnostics in existing smart buildings to conduct periodic checkups or to fix existing noticeable problems [20].

Personal mobile devices such as smartphones can provide extra capabilities in collecting data for diagnostics of smart buildings. Generally, smartphones can provide sensing data as users move in different areas. This approach of collecting sensing data is called participatory sensing [21]. Participatory sensing can be used to collect data about the environment, weather, and mobility as well as any other sensory information that collectively forms some knowledge of the current situation or configuration of a certain environment. It is also called people-centric sensing [22] as people play an important role with their personal devices and movement in collecting sensory information. Smartphones usually have multiple sensors such as a thermometer and light sensors as well as processing and communication capabilities that can be utilized to collect current temperature, lighting, and occupancy levels. This collected data can be used to enhance the diagnostic process in smart buildings.

Another new mobile device that can be used to collect data is the smart watch. Smart watches are wrist worn computers that run mobile operating systems and apps that can achieve multiple functionalities. They can have cameras and an array of sensors such as thermometers, accelerometers, altimeters, barometers, compasses, and GPS. In addition, they can communicate with other devices using Bluetooth and Wi-Fi. Although the concept of smart watches has been around for long time, they took years for the technology to advance enough for cost-effective and suitable implementation [23]. Due to their computation, communication, and sensing capabilities they have been proposed for use in a number of applications such as mobile health [24, 25], monitoring human behavior [26], and systems monitoring [27]. One of the advantages of smart watches over smartphones is that they are usually continuously attached to the human body. This provides more accurate sensing capabilities within the areas where the building's occupants are usually available.

With the utilization of smartphones and smart watches in smart buildings, more accurate virtual sensors [28] can be developed for diagnostic processes. Virtual sensors are logical sensors that provide economical alternatives to costly physical sensors. A virtual sensing system uses information available from other devices such as fixed physical sensors, smartphones, and smart watches to calculate an estimate of the quantity of interest. In this regard, there are two approaches of virtual sensing: analytical virtual sensing and empirical virtual sensing [29]. The analytical virtual sensing approach is based on the calculation of the measurement estimate using approximations of the physical laws including those that involve the distances of the used physical devices. The empirical virtual sensing approach is based on the calculations of the measurement estimate using the available current and previous measurements. Virtual sensors can provide low-cost sensing capabilities while expanding the data collection process for more accurate smart buildings diagnostics.

As more data is collected from smart buildings for analysis, storing and processing this data will require huge resources as well as advanced software that implements innovative algorithms for accurate fault diagnostics. This can be very costly for smart buildings owners. Cloud computing can provide a scalable and cost-effective platform for such needs [30]. Cloud computing is an emerging commercial IT infrastructure model that offers to eliminate the need for clients to maintain in-house high-cost hardware, software, and network infrastructures [31]. It also reduces, or even eliminates, the high-cost of recruiting technical professionals to support these infrastructures and operate the in-house IT solutions. Smart buildings can use different services that can be provided by cloud service providers such as data storage services, processing services, and fault diagnostics services. One of the advantages of this model is that as the cloud service provider can collect more data from multiple smart buildings; it can enhance the fault detection and diagnostic processes. In addition, this will enable cloud service providers to implement automated knowledge systems for smart building diagnostics [8] and other advanced mechanism such as fault detection analysis [32] for the benefit of the clients, the smart buildings.

6 Conclusions

In this paper we have explored the issue of data collection for diagnostics of smart buildings. This is a matter that has a significant place in every research done on the FDD topic, yet it is still flawed and poses a lot of questions. For these reasons, we aimed to summarize the challenges that accompany the data collection processes for FDD of smart buildings. We also hope that this would provide a basis for a future platform for joint efforts in overcoming these issues and, hopefully, sharing data and knowledge to support the research efforts in this domain.

Acknowledgement This work is supported by the Innovation Fund Denmark for the project COORDICY.

References

1. Snoonian, D.: Smart buildings. Spectrum, IEEE **40**, 18–23 (2003)
2. Agarwal, Y., Weng, T.: From buildings to smart buildings—sensing and actuation to improve energy efficiency. IEEE Design & Test of Computers, 36–44 (2012)
3. Jørgensen, B.N., Kjærgaard, M.B., Lazarova-Molnar, S., Shaker, H.R., Veje, C.T.: Challenge: advancing energy informatics to enable assessable improvements of energy performance in buildings. In: Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems, pp. 77–82. ACM (2015)
4. Katipamula, S., Brambley, M.R.: Review article: methods for fault detection, diagnostics, and prognostics for building systems—a review, Part I. HVAC&R Research **11**, 3–25 (2005)
5. Katipamula, S., Brambley, M.R.: Review article: Methods for fault detection, diagnostics, and prognostics for building systems—a review, part ii. HVAC&R Research **11**, 169–187 (2005)

6. Belfast, J.: Fault Diagnostics Tools for Commercial Buildings—Applications, Algorithms and Barriers. Energy Engineering **111**, 57–78 (2014)
7. Namburu, S.M., Azam, M.S., Luo, J., Choi, K., Pattipati, K.R.: Data-driven modeling, fault diagnosis and optimal sensor selection for HVAC chillers. IEEE Transactions on Automation Science and Engineering **4**, 469–473 (2007)
8. Fan, C., Xiao, F., Yan, C.: A framework for knowledge discovery in massive building automation data and its application in building diagnostics. Automation in Construction **50**, 81–90 (2015)
9. O'Neill, Z., Pang, X., Shashanka, M., Haves, P., Bailey, T.: Model-based real-time whole building energy performance monitoring and diagnostics. Journal of Building Performance Simulation **7**, 83–99 (2014)
10. Du, Z., Jin, X., Yang, Y.: Fault diagnosis for temperature, flow rate and pressure sensors in VAV systems using wavelet neural network. Applied Energy **86**, 1624–1631 (2009)
11. Li, S., Wen, J.: A model-based fault detection and diagnostic methodology based on PCA method and wavelet transform. Energy and Buildings **68**, 63–71 (2014)
12. Lazarova-Molnar, S., Kjærgaard, M.B., Shaker, H.R., Jørgensen, B.N.: Commercial Buildings Energy Performance within Context: Occupants in Spotlight
13. Goins, J., Moezzi, M.: Linking occupant complaints to building performance. Building Research & Information **41**, 361–372 (2013)
14. Mukhtar, Z., Tahir, M.I., Land, M.R., Olausson, M.M.: Diagnostics Data Collection and Analysis
15. Schumann, A., Hayes, J., Pompey, P., Verscheure, O.: Adaptable fault identification for smart buildings. In: Artificial Intelligence and Smarter Living (Year)
16. Kjærgaard, M.B., Lazarova-Molnar, S., Jradi, M.: Towards a categorization framework for occupancy sensing systems. In: Proceedings of the 2015 ACM Sixth International Conference on Future Energy Systems, pp. 215–216. ACM (Year)
17. Capehart, B.L., Brambley, M.R.: Automated Diagnostics and Analytics for Buildings. Fairmont Press, Inc. (2014)
18. Zanella, A., Bui, N., Castellani, A., Vangelista, L., Zorzi, M.: Internet of things for smart cities. Internet of Things Journal, IEEE **1**, 22–32 (2014)
19. Chaqfeh, M., Mohamed, N.: Challenges in middleware solutions for the internet of things. In: 2012 International Conference on Collaboration Technologies and Systems (CTS), pp. 21–26. IEEE (2012)
20. https://www.realcomm.com/advisory/advisory.asp?AdvisoryID=586
21. Burke, J.A., Estrin, D., Hansen, M., Parker, A., Ramanathan, N., Reddy, S., Srivastava, M.: Participatory sensing, UCLA: Center for Embedded Network Sensing (2006)
22. Tuncay, G.S., Benincasa, G., Helmy, A.: Autonomous and distributed recruitment and data collection framework for opportunistic sensing. ACM SIGMOBILE Mobile Computing and Communications Review **16**, 50–53 (2013)
23. Rawassizadeh, R., Price, B.A., Petre, M.: Wearables: has the age of smartwatches finally arrived? Communications of the ACM **58**, 45–47 (2014)
24. Boletsis, C., McCallum, S., Landmark, B.F.: The use of smartwatches for health monitoring in home-based dementia care. In: Human Aspects of IT for the Aged Population. Design for Everyday Life, pp. 15–26. Springer (2015)
25. Kalantarian, H., Alshurafa, N., Nemati, E., Le, T., Sarrafzadeh, M.: A smartwatch-based medication adherence system. In: 2015 IEEE 12th International Conference on Wearable and Implantable Body Sensor Networks (BSN), pp. 1–6. IEEE (2015)

26. Kalantarian, H., Sarrafzadeh, M.: Audio-based detection and evaluation of eating behavior using the smartwatch platform. Computers in Biology and Medicine **65**, 1–9 (2015)
27. Vega, A., Ljubic, S.: Smartwatch as a remote server monitoring device: implementation and interaction design. In: 2015 38th International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO), pp. 566–571. IEEE (2015)
28. Kabadayi, S., Pridgen, A., Julien, C.: Virtual sensors: abstracting data from physical sensors. In: Proceedings of the 2006 International Symposium on World of Wireless, Mobile and Multimedia Networks, pp. 587–592. IEEE Computer Society (2006)
29. Liu, L., Kuo, S.M., Zhou, M.: Virtual sensing techniques and their applications. In: International Conference on Networking, Sensing and Control, ICNSC 2009, pp. 31–36. IEEE (2009)
30. Hanna, E.M., Mohamed, N., Al-Jaroodi, J.: The cloud: requirements for a better service. In: 2012 12th IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), pp. 787–792. IEEE (2012)
31. Fox, A., Griffith, R., Joseph, A., Katz, R., Konwinski, A., Lee, G., Patterson, D., Rabkin, A., Stoica, I.: Above the clouds: A Berkeley view of cloud computing. Dept. Electrical Eng. and Comput. Sciences, University of California, Berkeley, Rep. UCB/EECS 28, 13 (2009)
32. Capozzoli, A., Lauro, F., Khan, I.: Fault detection analysis using data mining techniques for a cluster of smart office buildings. Expert Systems with Applications **42**, 4324–4338 (2015)