Understanding the Willingness to Share Building Data by a Social Study based on Privacy Calculus Theory

Yanhui Xu
The Hong Kong Polytechnic University
stephen.xu@connect.polyu.hk

Dan Wang
The Hong Kong Polytechnic University
dan.wang@polyu.edu.hk

ABSTRACT

In recent years, machine learning (ML) based building analytics have been developed for diverse building services. Yet, the widespread sharing of building data, which underpins the establishment of ML models, is not a common practice in the buildings industry today. Clearly, there are privacy concerns. There are studies on protecting building data, e.g., to k-anonymize building data; yet these studies are computational methods. The root causes of why building operators are or are not willing to share data are unclear.

In this paper, we study the problem of *willingness to share building data*. First, we justify our study by investigating the field to show that data sharing is indeed limited. Second, we examine the issue of the willingness to share building data from the perspective of a social study. We observe that the *intention to disclose* (i.e., decision making on data sharing) is not only based on *perceived risks*, but also on *perceived benefits*. We leverage the privacy calculus theory and present a systematic study. We develop hypotheses, design a questionnaire, conduct a survey involving 95 building operators and service providers around the world, and analyze the results, wherein we quantify how various factors influence the willingness to share building data. Third, we use trust, an important factors to the intention to disclose, to develop a trust model with differentiable trust levels. Such model provides building operators a mechanism to share data besides a 0-and-1 choice. We present a case study where we enhance an existing building data anonymization platform, PAD with the trust model. We show that the enhanced PAD has a substantially smaller computation workloads.

CCS CONCEPTS

• **Security and privacy → Social aspects of security and privacy**;

KEYWORDS

smart buildings, privacy calculus, data disclosure, trust

ACM Reference Format:
Yanhui Xu and Dan Wang. 2022. Understanding the Willingness to Share Building Data by a Social Study based on Privacy Calculus Theory. In *Proceedings of The 9th ACM International Conference on Systems for Energy-Efficient Buildings, Cities, and Transportation, Boston, MA, USA, November 9–10, 2022 (BuildSys '22),* 10 pages. https://doi.org/10.1145/3563357.3564059

BuildSys '22, November 9–10, 2022, Boston, MA, USA
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9890-9/22/11…$15.00
https://doi.org/10.1145/3563357.3564059

1 INTRODUCTION

Over the past few years, there has been a proliferation of building analytics, i.e., machine learning (ML) based approaches to enhance the efficiency and operational performance of buildings . Some common topics are: reducing the energy consumption of HVAC systems, maintaining thermal comfort for occupants, effectively utilising the space for building real estate, detecting indoor occupancy, and improving lighting control .

Despite these encouraging works, the widespread sharing of building data, which underpins the establishment of ML models, has remained largely elusive. Such data are usually confined within specific academic groups working on building-related research problems or held by external contractors of commercial buildings who are engaged in monitoring the day-to-day activities that take place in the building. Often, there are strict clauses attached to these engagements that prohibit the sharing of building data to be used for other purposes. Between building operators and service providers, there are lengthy negotiations on which specific data are needed by the service providers and data are shared on a need-to-know basis. Though it may not be always possible to make the building data fully public, we argue that it is particularly important for building service providers (as data consumers) to have certain autonomy in using the building data. For example, it is very common that service providers (including academicians) develop machine learning algorithms that need to use knowledge learned from the data of other buildings.

We investigated recent publications in two computer science publication venues and two built environment publication venues. We observed that less than 1/4 of the data-driven publications in computer science venues used shared data. The percentage is even smaller in the built environment venues, at less than 1/7. We further investigated a few existing shared datasets. We observed that they are restricted in the sense that they can only be used for a very limited number (usually only one) of building analytics applications.

Clearly, the unwillingness to share data is due to potential concerns on privacy and confidentiality. For example, it has been found that indoor occupancy data can be used to infer ownership of zones at the workplace through linkage attacks [41]. In another example, historical household electricity data can leak information about the personal habits and daily routines of residents [26].

There are existing studies in which attempts have been made to mitigate privacy concerns, such as by protecting privacy in data sharing based on k-anonymity and other methods. However, these studies assume that people are willing to share data in the first place. Computational methods can then be used to balance the utility of the data that were shared and the privacy risk associated with doing so. Unfortunately, there has been no study on the root causes affecting the willingness to share building data. Our work fills this gap in the literature.

In this paper, we study the problem of the *willingness to share building data*. The challenge is to understand the process by which people make the decision to share data. We thus resort to conducting a social research study on people's attitude towards privacy and their actual decision making behavior under circumstances involving privacy. We study a number of social theory frameworks. In this study, we draw on the privacy calculus theory since it best reflects the decision making process where the *intention to disclose* is based not only on *perceived risks*, but also on *perceived benefits*. This paper presents a systematic study. We developed hypotheses for a number of influential factors such as perceived severity, trust, building operation services, building occupant services, and smart ecosystems. We developed a questionnaire and asked an independent professional company to conduct a survey. We collected 95 responses from building operators and building service providers of diverse backgrounds on their levels of expertise and experience. We present a rigid data analysis of the results: first, we show that the overall design of our questionnaire is successful, i.e., the respondents could understand our questions in the way that we expected; and second, we present the results of our hypotheses, i.e., whether the influential factors have an impact on the intention to disclose and the magnitude of that impact. Our statistics reveal three interesting insights. First, that the perceived benefits can contribute as much as 130.90% in the decision making process as compared to the perceived risks. Second, the impact of building occupant services is twice of building operation services. Third, trust is important and can directly influence the intention to disclose.

The social research study provides scientific understanding of the factors towards decision making. We further investigate how our research results can be used. Our results show that trust is an important factor. Thus, we develop a trust model with differentiable trust levels. Intuitively, our model allows a differentiation on building data so that building operators can share different parts of data besides a 0-or-1 choice. We apply this trust model to enhance a recent privacy-preserving building data publishing framework, PAD [37]. PAD k-anonymizes building data records and it has high computational workloads when the amount of data is large. We show that with our trust model, less amount of data needs to be anonymized. Specifically, 45.3% of the total data need to be anonymized if there is moderate trust. With strong trust, 15.2% of data need to be anonymized. Our trust model-enhanced PAD can enjoy substantially reduced computational workloads.

The contribution of the paper can be summarized as:

- We show that data sharing in the building industry has been very limited to date. We investigate the problem of the willingness to share building data. To the best of our knowledge, we are the first to study this problem.
- We conduct our investigation from the perspective of social research. We show that the intention to disclose relates not only to perceived risks, but also to perceived benefits. We leverage a privacy calculus theory that exploits the decision-making process by weighing the benefits and risks.
- We use trust, one influential factor observed in our study, to design a trust model and we apply it to a PAD system. We show that the amount of data to be anonymized decreases to 15.2% if there is strong trust and the computational workloads of the PAD system substantially decreases.

Table 1: The number of publications (#), data-driven publications without data sharing (w/o) and with data sharing (w) in conferences & journals

Conference	#	w/o	w	Journal	#	w/o	w
e-Energy 2018	42	5	1	Appl Energy 318-320	75	3	0
e-Energy 2019	38	1	2	Appl Energy 315-317	102	0	1
e-Energy 2020	38	4	0	Appl Energy 312-314	178	4	0
e-Energy 2021	29	2	0	Appl Energy 309-311	180	1	0
Buildsys 2018	23	4	0	Energy Build 266-268	98	23	1
Buildsys 2019	38	10	3	Energy Build 263-265	50	6	1
Buildsys 2020	38	4	2	Energy Build 260-262	80	8	3
Buildsys 2021	28	4	0	Energy Build 257-259	105	15	2

2 BUILDING ANALYTICS AND BUILDING DATA SHARING: MOTIVATION FOR STUDY

Building Analytics: Currently, the maintenance, operation and control of buildings are based mainly on the principles of physics. Recently, big data and ML technologies have been developed for building services, such as predictive maintenance, operation efficiency, and others. Below are a few examples of building analytics:

- BLF [31]: Building load forecasting predicts the cooling load demand of a building at a certain time. BLF can improve HVAC operation and control.
- COP [45]: The coefficient of performance (COP) of a chiller captures the cooling power output of this chiller given a certain injection of electricity. COP prediction can improve HVAC operation and control.
- NILM [22]: Non-intrusive load monitoring uses the electricity load of a time period to infer what kinds of appliances are in use and when they are used.
- FDD [12]: Fault detection diagnosis (e.g., HVAC systems or lighting systems) takes the indoor environmental data and the mechanical readings of a piece of building equipment to detect the abnormal condition of this equipment.
- HAR [6]: Human activity recognition takes the readings of various sensors to recognize the specific activity at a zone.
- TMD [35]: Thermal model development takes indoor environmental data to develop a thermal model that simulates the thermal status in a building.

Building Data Sharing: Data sharing in the fields of computer vision (CV) and natural language processing (NLP) is common and they substantially accelerate the development of ML models and algorithms. While building analytics have been shown to be effective, data sharing is not common in the building industry sector. We present an analysis of building data sharing practices.

We first analyze the data sharing practices of the publications in two computer science publication venues (ACM e-Energy and ACM Buildsys) and two built environment publication venues (Applied Energy and Environment and Buildings). Table 1 shows the results. We see that in the computer science venues, of the 34 data-driven publications only 8 shared data, or less than 1/4. The proportion is even smaller in the built environment venues, where of the 60 data-driven publications only 8 shared data, or less than 1/7.

There are datasets which can be shared for non-commercial and educational purposes. We further analyze to what degree these existing datasets support common building analytics applications. We study the six aforementioned applications.

Table 2: Smart building datasets and corresponding supported applications

App Dataset	BLF	COP	HAR	FDD	TMD	NILM
BLOND	×	×	×	×	×	○
Build-FDD	×	×	×	○	×	×
CASAS	×	×	○	×	×	×
CU-BEMS	○	×	×	×	○	×
Genome	○	×	×	×	×	×
MFRED	×	×	×	×	×	○

- BLOND [22]: collected data from a German office building between October 2016 and May 2017. It includes records on the electricity consumption of 15 kinds of appliances.
- Build-FDD [12]: collected data from three laboratories in the USA. It contains fault data collected from three kinds of devices: air handling units (AHU), ventilation air volume systems (VAVs), and rooftop units (RTU).
- CASAS [6]: published 66 datasets collected from smart homes in Japan, Mexico, Paris, and Milan. CASAS contains records of different kinds of indoor activities: telephone use, hand washing, meal preparation, eating and medication use, cleaning, etc.
- CU-BEMS [35]: collected data from an office building in Bangkok, Thailand from July 2018 to December 2019. It contains three kinds of energy consumption data (i.e., air conditioning load, lighting load, and plug load) and three kinds of indoor environmental data (i.e., indoor temperature, relative humidity, and ambient light).
- Genome [31]: collected the electricity load data from 1,238 buildings in the USA and Europe between 2014 to 2016. These buildings served 13 different functions (e.g., educational buildings, government buildings, hospital buildings, etc.).
- MFRED [29]: collected the electricity load data from 390 apartments in the USA from January to December 2019. It contains data on appliances such as refrigerators, space heaters, light bulbs, and some entertainment devices.

We observe that these datasets are limited in the sense that it is quite difficult to use them in a wide range of building analytics applications. Table 2 shows how such data can be used to support the aforementioned six building analytics.

That there is little sharing of data and existing datasets are designed to support limited applications was what motivated this study: to understand the willingness to share building data.

3 SOCIAL RESEARCH AND RELATED WORK

3.1 Social Research on Privacy

Social research on privacy is the study of the decision-making behavior of people under circumstances involving privacy. Intrinsically, there is a "privacy paradox" [4], i.e., how people evaluate the trade-off in their decision-making process. Various theoretical frameworks on privacy were developed (e.g., privacy calculus, prospect theory, etc.) to study the driving forces (e.g., trust, perceived severity, better services) behind the making of decision. We first briefly present privacy theories and why we choose to adopt the privacy calculus framework. We then present some studies using the privacy calculus theory.

Privacy Calculus As early as 1968, it was found that privacy is related to the behavior of withdrawing to protect certain information from the outside world [43]. On the other hand, when people are interacting with the outside world or developing social relationships, some forms of disclosure behavior are also required. In 1977, Laufer and Wolfe proposed the "calculus of behavior" theory to explain that people weigh the *perceived benefits* and *perceived risks* of their disclosure behavior to decide to what extent they will disclose or withdraw personal information [24]. The trade-off in the disclosure behavior of people was formally named "privacy calculus" in 1999. There are studies providing extensive details of the perceived benefits (e.g., trust) and perceived risks (e.g., perceived severity) related people's privacy concerns, as well as studies on empirical supports in various contexts (e.g., e-Commerce, SNS, and location-based services).

Communication Privacy Management (CPM) theory In [43], privacy disclosures are regarded as one of the primary approaches to developing and maintaining social relationships. CPM was adopted to study how families manage the issue of privacy to maintain family relationships (e.g., topic avoidance). For example, married couples avoid talking about the experience of miscarriage to maintain their relationships [5]. Intrinsically, CPM theory puts forward a description of how people erect rule-based boundaries to disclosing their information with third parties when developing and maintaining relationships. CPM theory has been used in other settings, such as employer-employee relationships, customer-retailer relationships, etc. In our study, we emphasize the consideration of benefits and risks. Relationship management between building operators and service providers may be a future study.

Prospect Theory (Nobel Prize 2002) is widely used to analyze decision-making behavior when the decision makers are facing risky choices [17]. The key finding of Prospect Theory is an asymmetric value function, whereby people can be risk-averse but also risk-seeking depending on whether the choices lead to a gain or a loss. The asymmetry of this function explains the "irrationality" of people when making decisions. For example, in the choice of having a 100% chance to gain $450 or a 50% chance to gain $1,000, people chose the former even though the expected gain was higher in the later. Prospect theory has also been used to analyze the decision-making process in relation to privacy. In this paper, we explore the forces driving the determination of benefits and risks; thus, we adopted the privacy calculus theory over the prospect theory.

Contextual Integrity is about studying information flows under certain circumstances (e.g., in an interview, it might not be appropriate to ask questions related to religion since religion is very private matter) [34]. The contextual integrity theory states that an appropriate flow of information can be regulated with a five-parameter setting, i.e., regulations should be designed on the subject, sender and recipient of data, the type of information and the principles of transmission. Specifically, privacy norms have been developed in certain contexts (e.g., to govern the actions of people in an online space), and context integrity theory has been applied to detect and identify violations of privacy due to the context change. Again, we are more interested in the forces driving the decision to disclose data than how the information flows should be regulated. Thus, contextual integrity is not suitable for our study.

3.2 Studies using the Privacy Calculus Theory

We now present studies using the privacy calculus theory. The following steps are commonly flowed when conducting a social study: proposing hypotheses, designing a questionnaire, recruiting participants, collecting data (i.e., the responses of the participants), validating the data, analyzing the data, and presenting the findings.

Social Network Services (SNS), such as Facebook and Twitter, are highly dependent on user information and are well-known for bringing about privacy concerns. The use of such services is highly dependent on people's perceptions of the benefits and risks involved, and the privacy calculus theory has been applied to gauge such perceptions. For example, there is a study on information disclosure through mobile applications. In that study, it was first hypothesized that personalized services and self presentation are factors of perceived benefits, and perceived severity and perceived control of information are factors of perceived risks. To collect data, a questionnaire was designed and a group of Facebook users were recruited. In the data analysis, it was shown that personalized services are the dominant factor in the perceived benefits, and that perceived control is the dominant factor in the perceived risks [40]. Follow up studies have been conducted on a number of additional factors, e.g., the influence of culture and gender to better understand the factors of perceived benefits and perceived risks.

IoT Services may raise concerns about privacy. The privacy calculus theory has been used to study the intention to adopt IoT services and to explain the trade-offs when making decisions on whether to disclose information. For example, it has been found that the privacy concerns of customers will occasionally not have a significant influence on the intention to adopt an IoT service; instead the customers' trust in the service providers could overcome their concerns about the risks to their privacy [15].

Healthcare Applications also raise privacy concerns. There is a large body of research using the privacy calculus theory. Studies have shown that having trust that a product has been well-developed plays a direct positive role in the intention to adopt a healthcare application. Other findings include the discovery that the current health status of users would influence their willingness to make disclosures [20].

4 THEORETICAL FRAMEWORK

The assumption of privacy calculus is that the *intention to disclose* is based on the calculus of the *perceived benefits* and *perceived risks*.

We first clarify the scenario on data disclosures. A *building operator* is the data owner, who represents an aggregate of the stakeholders in a building, (e.g., building owners or tenants). In this paper, we use a building operator to represent the collective decision-making on data disclosures of the building. A *building service provider* is the data consumer, who can develop services for the benefit of the building. Here, a service also represents the aggregate benefits, e.g., an AI service to enhance the occupancy comfort for a commercial building; an energy conservation service for government buildings in a smart city campaign; a piece of research to develop new understandings on building learning models.

A building operator makes a decision to *share* building data to a building service provider. This sharing means that building service providers have certain autonomy in using the building data. This gives flexibility to building service providers beyond a strict Non-Disclosure Agreement (NDA) where the data can be used on pre-defined terms and a need-to-know basis. We argue that such sharing is necessary for the overall benefits of the smart building industry. For example, it is very common that service providers (including academicians) develop machine learning algorithms that need to use knowledge learned from the data of other buildings.

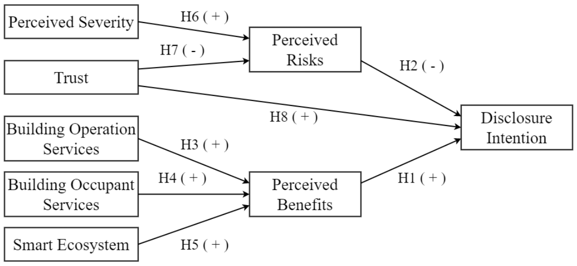

Figure 1: Research Model

Note that this sharing does not necessarily mean that the building operator discloses the data to fully public. Under the privacy calculus theory framework, data sharing will need to have some minimum agreements, e.g., usage for non-commercial and educational purposes since disclosing data fully public indicates that the perceived risk can be infinite and thus is beyond the expressiveness of the privacy calculus theory. Signing strict NDAs indicates that the perceived risk is zero; and as said, it does not benefit the overall industry. In this paper, we study the in-between scenarios and we strive to understand what the concerns/factors are to trigger the decision-making process to share data in these scenarios.

Clearly, the process by which the building operators make the decision to disclose data is a dependent variable to a number of factors. In the privacy calculus theory, these factors are called *constructs*. In this paper, we hypothesize eight constructs. Our overall privacy calculus research model is shown in Fig. 1; and we will explain the choices of these constructs shortly. We comment that there are other constructs; yet we want to limit the complexity of our research model and the length of our questionnaire. We leave investigating other constructs to a future study.

We also comment that the privacy concerns in this paper include confidentiality concerns, i.e., building operators do not want to disclose data sometimes because they are concerned about revealing their potential problems in operation and control. Scientifically, this is a confidentiality factor rather than a privacy factor. As an early study, we use privacy to include both privacy and confidentiality concerns since building operators would calculate both factors when making decisions to share or not to share data.

4.1 Antecedents[1] of the Disclosure Intention

In our proposed privacy calculus model, there are three basic constructs; namely, perceived benefits, perceived risks, and disclosure intention. Perceived benefits are generally related to customized services, financial rewards and the promotion of the public image, etc. For example, Bee'ah, a company in the Middle East, upgraded their office building by adopting the smart building services provided by Johnson Control and Microsoft. This upgrade improved the quality of the office, and also reduced the building energy consumption by 5% and water consumption by 20% [1]. Perceived risks are the degree to which people believe that there is a potential loss associated with the disclosure of their data (e.g., through misused by service

[1] Antecedent originated from behavioral psychology. It indicates the stimulus that cues an organism into performing decision-making behavior. In social studies, the factors that are significantly related to a behavior are the *antecedents* of this behavior.

providers [8] or by being sold to third parties without permission [44]). For example, in a study conducted in Orizaba, Mexico, it was specifically stated that the disclosure of the 10-month building data was subject to individual negotiation due to perceived risk involving privacy concerns [36].

Similar to previous studies, we put forward to the following hypotheses on the positive effect of perceived benefits and negative effect of perceived risks on the willingness to disclose building data:

H1. Perceived benefits are positively related to the intention to share building data with a smart building service provider.

H2. Perceived risks are negatively related to the intention to share building data with a smart building service provider.

4.2 Antecedents of Perceived Benefits

New services are the subject of perceived benefits. For example, in the context of adopting new mobile applications, people would show a positive attitude towards disclosing their data to service providers if they could get personalized services in return (e.g., a birthday coupon or book recommendation) [25]. For building operators, we hypothesize that obtaining smart building services are positively related to perceived benefits.

We consider two types of smart building services: (1) building operation services and (2) building occupant services. Building operation services improve building operation, i.e., they help building operators save effort or cost in operating building. For example, the BLF, COP and FDD in §2 are of this type. Building occupant services improve the comfort and productivity of building occupants. For example, the HAR and TMD (also in §2) are of this type. Our hypotheses about these two types of smart building services are:

H3. Building operation services are positively related to building operators' perceived benefits.

H4. Building occupants services are positively related to building operators' perceived benefits.

It has been found that the ecosystem of a service can boost the quality of a service and also increase users' loyalty to the service [42]. An example of this is the Apple ecosystem established through iTunes and the App Store. We note that smart buildings co-evolve with smart city, smart governance, smart grid, etc., e.g., it has been shown that a smart grid can help to improve the energy management system of a building [33]. We address people's belief in the smart ecosystem as a driving force behind the perceived benefits. We propose the following hypothesis:

H5. A smart ecosystem is positively related to building operators' perceived benefits.

4.3 Antecedents of Perceived Risks

Building operators will estimate the consequences of the use of building data and perceive the potential risks involved [32]. Perceived severity refers to the trigger of privacy protection behavior [23]. For example, a building operator considers that the water temperature data of a chiller may not trigger a threat, yet the CO_2 density data may trigger a threat to the location privacy of occupants. It is empirically shown that in the context of the adoption of technology, perceived severity could raise people's concerns on the adoption [32]. We propose the following hypothesis:

H6. The perceived severity of sharing building data with a smart building services provider is positively related to the building operator's perceived risks.

If trust is established, people will lower their awareness of the potential risks and be more inclined to disclose data [8]. In a study on people's concerns about privacy on the Internet, trust was addressed and it was found that people believe that a company is dependable if it protects personal information [27]. Following studies investigated people's trust in terms of the closeness of the relationships of the company [7] and people's knowledge showing of the company (e.g., brand reputation [3]). We thus have:

H7. Trust is negatively related to a building operator's perceived risks.

Previous studies also showed that the effect of trust could even overwhelm people's concerns about privacy and directly influence the decision to disclose information [18]. Thus, in addition to the effect of trust on perceived risks, we also hypothesize the effect of trust on intention to disclose directly:

H8. Trust is positively related to a building operator's intention to share building data with a smart building services provider.

5 RESEARCH METHODOLOGIES

5.1 Questionnaire Design

Our questionnaire consists of two parts, a background survey and main questions. Our questionnaire is public accessible[2].

The first part of the questionnaire focuses on the background of the participants, including their demographic profile and their professional knowledge profile.

The second part contains the main questions. A summary is given in Table 3: Column 1 shows the eight constructs; Column 2 shows the measurement items, i.e., the questions that we developed[3], organized according to constructs; (note, however, that in our questionnaire we neither inform the participants of the constructs, nor tell them that the questions are organized according to constructs); and Column 3 shows the references we consulted when developing the measurement items; many constructs were investigated in other studies, and we chose our measurement items (questions) primarily by adopting the items from them:

Perceived Severity: we consulted the studies on social network [32], where the measurement items represent the people's perceived severity of different types of data. Consequently, we design PSEV1 for the perceived severity on overall data and PSEV2 to PSEV4 on three major categories of smart buildings data.

Trust: we referred to the study [8, 21] on the trust between the data owner and e-commerce retailer, where the measurement items show how the data owner determines that the service provider is trustworthy. We designed TRST1, which is about the reputation of the service provider, and TRST2 which is about the business relationship between the building operator and the service provider.

Building Operation Services and Building Occupant Services: we followed a study on location-based services [38], where the measurement items list different types of services that require the data

[2] https://github.com/KaruBios/PISB

[3] In the privacy calculus theory, measurement items is the formal term used to refer to questions related to the constructs that are to be measured. A questionnaire includes other questions, such as those on demographic profiles.

owner to disclose their data in exchange. We designed BOS1-BOS3 for three types of building operation services and OPS1-OPS2 for two types of building occupant services.

Smart Ecosystem: we followed the literatures on investigating IoT ecosystems [14, 28]. We noted that Smart City is one type of Smart Ecosystem. We thus designed SECO1 to investigate the smart city development in the cities of the participants. We designed SECO2 to investigate the popularity of smart buildings in the participants' cities, where the popularity of a service/product represents the development of an ecosystem [28].

Perceived Risks: we used the measurement items from [25, 44] as our PR1-PR3 to investigate the perceived risks of disclosing building data. PR1-PR3 requires the participants to give answers based on their intuition, knowledge and estimations.

Perceived Benefits: we adapted the measurement items from [14, 25, 44] to investigate the perceived benefits on disclosing building data in exchange for smart building services. We designed PB1 for the overall benefits and PB2-PB4 for three specific types of benefits.

Disclosure Intention: we followed [38], and designed DI1 to investigate people's willingness to disclose building data. We designed DI2-DI3 to investigate whether people's disclosure intention differs according to different types of building data involved.

We measure the constructs with items assessed using a 5-point Likert scale, with 1 indicating ("Strongly Disagree") and 5 indicating ("Strongly Agree"). A summary of the items is shown in Table 3.

5.2 Characteristics of the Sample

To recruit participants for this study, we sought professional services from an independent marketing company. We requested to recruit building operators (e.g., the facility operators in a commercial building) as well as engineers from smart building service providers (e.g., application engineers from Johnson Control) as our target participant group. We held multiple rounds of discussions with the company to ensure that they understand our requests.

Eventually, we have 95 responses with 46 from the building operators and 49 from engineers from smart building service providers. The background statistics on the respondents are reported in Table 4. The respondents were diverse in terms of their level of experiences, the type of building they operated, and their speciality.

6 DATA ANALYSIS

We analyze the measurement model and structural model. A measurement model shows the relationship between the measurement items and the constructs. It indicates whether the overall design of our measurement items is successful, e.g., whether the respondents can understand our questions in the way we expected. A structural model shows the relationship between the constructs; e.g., whether trust has an impact on the disclosure intention (Hypothesis 5), and if yes, how much. It shows the results in relation to our hypotheses.

We conduct the analysis using the Structural Equation Modeling (SEM) methods. SEM methods are a group of methods used in social and behavioral sciences for a *model* representing some observable or theoretical phenomenon where a phenomenon is theorized to be related to one another with a *structure* [9].

There are a number of SEM methods, e.g., covariance-based structural equation modeling (CB-SEM) and Partial Least Square structural equation modeling (PLS-SEM). CB-SEM is used in studies with a large number of respondents (e.g., more than 200); otherwise, there may be non-convergence model fitting problems. In this paper, we adopted PLS-SEM, which is designed for experiment with small sample size. PLS-SEM can simultaneously evaluate a measurement model and a structural model. We used SamrtPLS 3.0 to analyze our collected data and fit our research model.

6.1 Measurement Model Results

We first present how successful our measurement model is. We follow the standard procedure and evaluate the results from an intra-construct perspective and an inter-construct perspective:

Intra-construct evaluation. For a measurement item, e.g., Perceived Severity, we determine whether the responses of the participants, e.g., to PSEV1-PSEV4, are sufficiently consistent with this construct. We evaluate the (1) the convergent validity of the proposed constructs; i.e., whether the scale setting of different measurement items in the same construct are consistent or not; and (2) the discriminant validity of the proposed constructs, which verifies the consistency of the question's design (e.g., expression and wording) in the same construct. We use four metrics: Factor Loading, Average Variance Extraction (AVE), Composite Reliability (CR), and Cronbach's α. The results are shown in Table 5.

Factor loading refers to the correlation coefficients between the measurement items and the construct, e.g., the correlation between PSEV1 (sharing data raises serious problems) and Perceived Severity. The factor loading of PSEV1 to perceived severity is 0.828, which indicates that the correlation between PSEV1 and Perceived Severity is significantly strong. The threshold for factor loading showing that the design of measurement model is successful is 0.6 [11]. We can see that all factor loadings are greater than the threshold.

AVE is a measure of the measurement errors introduced by the design of the questions and the collecting of data [19]. A high AVE value indicates that the corresponding construct has good discriminant validity. The recommended threshold for AVE is 0.5 [11]. We can see that all AVEs are greater than the threshold.

Composite Reliability (CR) and Cronbach's α are the metrics used to evaluate measurement errors introduced by the scale setting of the measurement items. A high CR value and a high Cronbach's α value represent good convergent reliability of the corresponding construct. The recommended Cronbach's α and CR thresholds are both greater than 0.7 [13]. We can see that the Cronbach's α and CR values are all above the thresholds, with the exception of the value of Cronbach's α for the Smart Ecosystem. As shown in [39], this will not overturn our results. Therefore, the results indicate that the measurement model is of sufficient reliability.

Inter-construct evaluation. We evaluate the discriminant validity of our proposed constructs (convergent validity cannot apply to an inter-construct evaluation). Basically we need to check the root of the AVE value of each construct and the correlation coefficients with other constructs. As shown in Table 6, the root of the AVE value for all constructs is greater than the correlation coefficients with the other constructs, which confirms the discriminant validity according to [11].

6.2 Structural Model Analysis

Now we present the test results on our structural model, i.e., to determine whether our proposed hypotheses are supported or rejected. We use the PLS statistics for the evaluation and show the

Table 3: Measurement Items

Construct	Measurement Items	Ref.
1. Perceived Severity	PSEV1: Sharing the data collected from my building with service provider would raise serious problems.	[32]
	PSEV2: Sharing energy consumption data of my building with service provider would raise serious problems.	
	PSEV3: Sharing the mechanical data of the devices in my building with service provider would raise serious problems.	
	PSEV4: Sharing the data related to the occupants in my building with service provider would raise serious problems.	
2. Trust	TRST1: The service provider is a well-known company in industry.	[8, 21]
	TRST2: The service provider has a closed and stable collaboration relationship with us.	
3. Building Operation Services	BOS1: The smart building service provider could provide us energy saving service.	[38]
	BOS2: The service provider could provide us predictive maintenance system	
	BOS3: The service provider could provide us smart security system.	
4. Building Occupant Services	OPS1: The service provider could provide us smart workplace management system.	[38]
	OPS2: The service provider could provide us smart human-centric lighting system.	
5. Smart Ecosystem	SECO1: Our city is a well-developed smart city.	[14, 28]
	SECO2: Many buildings have been adapted into smart building in our city.	
6. Perceived Risks	PR1: Adopting smart building services to my building would involve many unexpected problems.	[25, 44]
	PR2: Adopting smart building services to my building would be risky.	
	PR3: The potential for loss in adopting smart building services to my building would be high.	
7. Perceived Benefits	PB1: I believe that smart building services could bring benefits to us and our building.	[25, 44]
	PB2: I believe that using smart building service can improve the asset value of the building.	[14]
	PB3: I believe that smart building services could help my building meet the requirements of green norms or sustainable development released by the government or other related authorities of our city.	
8. Disclosure Intention	DI1: I will allow the service provider to have fully access to the data of my building.	[21, 38]
	DI2: I will allow the service provider to access the data of my building regardless the category of the data.	
	DI3: I will allow the service provider to access the data of my building regardless the time period of the data.	

Table 4: Characteristics of the sample

Variables	Levels	Frequency	Percent
Building Operators (N1=46)			
Building Operation Experience (years)	<=5	14	30.4%
	5-10	20	43.5%
	10-15	7	15.2%
	>15	2	4.3%
	X	3	6.5%
Building Type	Commercial	20	43.5%
	Residential	20	43.5%
	Mixed	4	8.7%
	X	2	4.3%
Engineers from Service Provider (N2=49)			
Work Experience (years)	<=5	25	51.0%
	6-10	16	32.7%
	>10	3	6.1%
	X	5	10.2%
Speciality	Computer Science	14	28.6%
	Civil Engineering	20	40.8%
	Others	10	20.4%
	X	5	10.2%

"X" represents missing item of input

Table 5: Measurement Model Statistics

Construct	Item	Factor Loading	AVE	CR	Cronbach's α
Perceived Severity	PSEV1	0.828	0.785	0.936	0.908
	PSEV2	0.915			
	PSEV3	0.918			
	PSEV4	0.901			
Trust	TRST1	0.784	0.739	0.849	0.666
	TRST2	0.930			
Building Operation Services	BOS1	0.809	0.677	0.863	0.762
	BOS2	0.794			
	BOS3	0.864			
Building Occupant Services	OPS1	0.833	0.729	0.843	0.637
	OPS2	0.769			
Smart Ecosystem	SECO1	0.865	0.705	0.827	0.583
	SECO2	0.814			
Perceived Benefits	PB1	0.768	0.689	0.869	0.773
	PB2	0.851			
	PB3	0.867			
Perceived Risks	PR1	0.857	0.668	0.858	0.755
	PR2	0.843			
	PR3	0.748			
Disclosure Intention	DI1	0.893	0.795	0.921	0.871
	DI2	0.905			
	DI3	0.878			

results in Fig. 2. We check the path coefficients (β) between the constructs and the p value (p) to verify whether our proposed hypotheses are supported. Path coefficients represent the statistical linkage between the independent factor and the dependent factor (e.g., in Perceived Severity → Perceived Risks, Perceived Severity is the independent factor and Perceived Risks is the dependent factor). A path coefficient with a positive value indicates that the independent factor has a positive effect on the dependent factor, and a path coefficient with a higher absolute value represents a stronger effect. The p value is the statistical indicator showing whether the effect

Table 6: Discriminant Validity

*Diagonal numbers indicate the square roots AVEs, the other numbers indicate the correlation coefficients between constructs.

	DI	BOS	OPS	PB	PR	PSEV	SECO	TRST
DI	**0.892**							
BOS	0.483	**0.823**						
OPS	0.428	0.684	**0.854**					
PB	0.464	0.675	0.764	**0.830**				
PR	-0.381	-0.351	-0.312	-0.234	**0.818**			
PSEV	-0.371	-0.324	-0.291	-0.248	0.817	**0.886**		
SECO	0.530	0.375	0.392	0.533	-0.050	-0.124	**0.839**	
TRST	0.556	0.521	0.531	0.533	-0.419	-0.393	0.488	**0.860**

of the independent factor on the dependent factor is significant, i.e., whether the linkage is acceptably real. A small p value represents great significance. An acceptable p value should be less than 0.1 for experiments with a small sample size.

The results show that Perceived Risks has a negative impact ($\beta = -0.178$, $p = 0.079$) on the willingness of building operators to disclose their building data, meanwhile Perceived Benefits has positive and statistically significant effect ($\beta = 0.233$, $p = 0.031$) on building operators' disclosure intention. Therefore, in our study, both **H1** and **H2** are supported. In terms of the antecedents of perceived benefits, the positive effect of both Building Operation Services ($\beta = 0.237$, $p < 0.05$) and Building Occupant Services ($\beta = 0.505$, $p < 0.001$) on the building operators' Perceived Benefits are significant. This means both **H3** and **H4** are supported. In addition, the positive effect of a Smart Ecosystem on Perceived Benefits is also significant ($\beta = 0.246$, $p < 0.001$). Therefore, **H5** is also supported in our study. In terms of the antecedents of perceived risks, Perceived Severity has a significant positive effect ($\beta = 0.772$, $p < 0.001$) on the building operators' Perceived Risks, which leads to support for **H6**. We also see that Trust does not have a significant effect on Perceived Risks ($\beta = -0.116$, $p = 0.205$). This means that **H7** is rejected. Intuitively, this means that whether or not there is trust, trust cannot be used to predict whether or not building operators have Perceived Risks. However, the positive effect of Trust on the building operators' disclosure intention was found to be significant ($\beta = 0.358$, $p < 0.001$). Therefore, **H8** is supported.

6.3 Insight and Practical Implications

We present three insights that can be drawn from our results.

First, we observe that the magnitude of perceived benefits is significantly greater than the perceived risks for 130.90% (0.233 over 0.178). This reflects that one can expect more gains by putting more efforts on perceived benefits than perceived risks. This echoes the main theme of this paper, and implies that the willingness of sharing building data not only depends on perceived risks but significantly more on perceived benefits.

Second, we observe that the magnitude of building occupant services is twice of building operation services for 213.08% (0.505 over 0.237). We believe that this reflects the reality. Building operators weigh more on occupant services since occupant services may directly increase their rental fees; which may be more significant than operations services. One implication is to synthesize operation services (e.g., energy conserving) into occupant services such as Environmental, Social, and Governance (ESG) services; this can increase the willingness of sharing building data more significantly.

Third, we observe that the magnitude of trust is significantly greater than perceived benefits for 153.65% (0.358 over 0.233). This asks building data consumers to pay more attention to trust. One potential solution is to develop a trust model with different levels. With differentiable levels, building data owners have more choices in sharing different parts of building data besides a 0-or-1 choice.

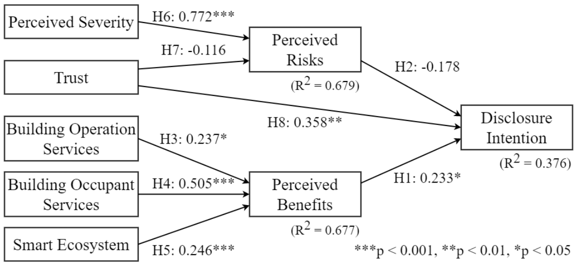

Figure 2: Structural Model

7 CASE STUDY: TRUST MODEL-ENHANCED PAD

We present a case to show how our study can be used to increase data sharing, which may also be used to improve the performance of systems. Specifically, we study a privacy-preserving building data publishing framework, PAD [37]. PAD computes k-anonymity for building data records. It has high computational workloads when the amount of data to be anonymized is large. In this paper, we observe that *trust* can increase the intention to disclose. We argue that if there exist a trust model with different levels on the trust of building service providers, building operators will have a mechanism to increase data sharing. This can reduce the amount of data to be anonymized. In what follows, we briefly introduce PAD and then develop a new trust model-enhanced PAD.

7.1 PAD: a Privacy-preserving Building Data Publishing Framework.

PAD was developed to protect building data privacy through k-anonymity of the published data while maintaining good data utility. Specifically, PAD performs anonymization on the building data by *aggregating* the data records with *pre-training* a measuring method on distance metric. For example, the number of occupants in a zone may expose the location privacy of occupants, e.g., it can disclose the working hours of the occupants located in a specific zone. With PAD, zones can be k-anonymized, i.e., one can only tell the number of occupants in an aggregate of k-zones and it becomes difficult to infer the number of occupants in a specific zone.

Similar to other anonymization methods [30], PAD has relatively high computational workloads when the amount of data is large. Specially, let N be the total number of data collection sensors, T be the total period of the collection time, n_{ij} be the amount of data collected by sensor i in time period j. Let M be the total amount of data. We have $M = \sum_{i=1}^{N} \sum_{j=1}^{T} n_{ij}$. Let k be the anonymity level. With the pre-training measuring method and the aggregation algorithm of the PAD, the computation workloads of PAD is determined by the number of sensors and the amount of data (which is affected by the time period) which can be expressed as:

$$O(PAD) = N^2 M + (\frac{N}{k} - 1)(\frac{N}{2} + k - 2). \tag{1}$$

7.2 Trust model-enhanced PAD

We now develop a new trust model. This trust model allows building operators to share different amount of data given different levels of trust. For example, a building operator may give an academician data to develop a COP analytics application. If he trusts that, different to a commercial company, the academician will neither have the incentive nor have the capability to develop personalized advertisement to building occupants, the amount of data to be anonymized can be reduced. In another example, if the building operator gives the data to a long-term collaborator, and he trusts that the building service provider will seek consent from him when using data different from the COP analytics application [27], the building operator can share even more data, e.g., only zones that are most sensitive with the senior admins need to be anonymized.

We propose a simple *trust model.* Specifically, suppose each data record has a *sensitivity level* $l \in \{1, 2, \ldots, \mathcal{L}\}$, where a greater value of l indicates a higher sensitivity. Let $L \in \mathcal{L}$ be the *trust level*; i.e., a data owner trusts a data consumer in L means that raw data records with sensitivity level less than or equal to L ($l \leq L$) can be published to the data consumer without anoynimization. We materialize the sensitivity levels and trust levels as follows.

Sensitivity levels: It has been observed that some sensors (e.g., flowrate sensors of the HVAC system) and some time periods (e.g., arrival/departure periods) are sensitive in inferring the occupancy [16]. We develop spatial and temporal sensitivity levels for both the sensors and the time periods; these are also the two core factors to bring about the computation workloads of PAD in Eq. (1):

- Sensor sensitive (SS) levels: It is known that the flowrate sensors can reflect the real-time cooling load and can be used to predict the occupancy levels; yet the flowrate sensors on different chillers contribute differently in the occupancy prediction [10]. Accordingly, we develop three SS-levels: (1) Low-level: the flowrate sensors installed on back-up chillers; (2) Middle-level: the flowrate sensors installed on the chillers for daily use; (3) High-level: the flowrate sensors installed on main chiller pipes; high-level sensors can infer the overall cooling load whereas middle-level sensors can infer the cooling load of a specific chiller.
- Time sensitive (TS) levels: It is known that different time periods have different sensitivity [16]. Accordingly, we develop three TS-levels: (1) Low-level: the closing hours and Work From Home (WFH) period due to Covid-19; (2) Middle-level: normal office hours; (3) High-level: the arrival/departure periods (8am-10am/5pm-7pm); the location privacy are more sensitive.

Trust levels: building operators can classify building service providers. We study four trust levels: (1) no trust; all data records need to be anonymized; (2) moderate trust, e.g., building service providers will use the data for non-commercial activities. The data records of SS and TS at the middle-level and high-level need to be anonymized; (3) strong trust, building service providers will inform the building operators of how he uses the data [27], (e.g., TRST2 in our questionnaire). Only the data records of SS and TS at the high-level need to be anonymized; and (4) full trust.

We apply our trust model in a real-world case of the WKGO Building of Hong Kong. The total number of the flowrate sensors of WKGO is 596, and the collection period was November 2019 to May 2021. With PAD, the computation time to anonymize all data is 16.42 hours in a state-of-the-art computer. With our trust model, the amount of data to be anonymized decreases. For example, at the moderate trust level and the strong trust level, the amount of data to be anonymized can decrease to 45.3% and 15.2% and the computational time reduces to 2.48 hours and 0.82 minutes, which is 15.1% and 0.02% to that of anonymizing all data.

8 DISCUSSIONS

8.1 Theoretical and Practical Contributions

In this study, we analyze a number of social research theoretical frameworks. The main theoretical contribution of this study is that we put forward the first privacy calculus model to study the data disclosure behavior of building operators.

In addition to what have been discussed in §6.2, our study exhibits several practical implications with regard to increasing the willingness of building operators to disclose their data. First, the main theme of the paper was supported, i.e., the intention to disclose is affected not only by perceived risks but also by perceived benefits. We think one current fact is that building operators are still not familiar with (and not confident about) the benefits brought about by smart building services. Towards this end, we suggest that smart building service providers and academicians strive to better articulate, quantify, and standardize building services. Second, a smart ecosystem can also promote perceived benefits. We suggest that the government promote smart buildings and smart cities. Recently, we saw a few AI competitions in smart buildings and smart cities [2] led by the government, and data disclosures followed suit. Third, we observe that trust has a significant effect on the disclosure intention. Therefore, we suggest that service providers engage in satisfactory business arrangement with building operators and also provide systems with better security protection.

8.2 Limitations

We present three limitations of our work. First, the number of participants in our study was on the low side. In social research, the recommended number of participants is around 10 times the number of measure items (i.e., questions). We had 22 technical questions (Table 3) but only 95 participants. This is because we needed participants with a certain level of experience in buildings. Recruiting such participants by professional companies has high costs. Second, there are many stakeholders in the context of smart buildings, such as building occupants, property owners, information security officer, etc. They represent diverse data owners and their concerns may not be the same. Previous studies in other contexts did categorize data owners, e.g., age and ethnic groups. We omitted these to control the complexity of our research model. Third, there are many privacy factors in the context of smart buildings. For example, the government could be involved as a driving force in increasing the willingness of building operators to disclose their data. Previous studies in other contexts did involve government regulations as a factor in this endeavor. Similarly, building operators would be willing to make their building smart in order to fulfill their social responsibilities and requirements. For example, Environmental, Social, and Governance (ESG) factors have become extremely important in obtaining investment. These factors can become future work to understand specific topics in depth.

9 CONCLUSION AND FUTURE WORK

Data sharing is quite uncommon in the smart building industry. Existing studies (e.g., k-anonymity, differential privacy) have focused on the aspects of algorithms in protecting data while maintaining good data utility. The root causes of why people are or are not willing to share building data are unclear. This work demonstrated, through a social research study based on the privacy calculus theory, that the intention to disclose is related not only to perceived risks, but also to perceived benefits. We studied six antecedent influential factors, such as trust, building services, smart ecosystems, and others. We developed a trust model allowing building data to have differentiable levels. Thus, building operators have choices to share different parts of the data and this increases data sharing and reduces the workloads of anonymization systems.

Our study provides an initial understanding of the decision-making process on building data sharing. Future works can involve three aspects. First, this paper adopted the privacy calculus theory. We believe that other theoretical frameworks are also worth considering. Second, we considered a limited number of stakeholders and six basic influential factors. Clearly, there are other stakeholders and factors involved in the decision-making process on building data sharing. Third, we urge to study differentiable levels for all factors; aiming at de facto practices and consensuses. This can facilitate overall data sharing.

ACKNOWLEDGMENTS

The authors would like to thank Dr. Arun Vishwanath for his contributions in the early draft of this paper. The authors also would like to thank the reviewers and shepherd of this paper for their constructive comments, which improved the paper significantly. This work is supported by RGC GRF 15210119, 15209220, 15200321, 15201322, ITF ITSP-ITS/070/19FP, C5018-20G of Hong Kong.

REFERENCES

[1] 2019. Bee'ah selects Johnson Controls and Microsoft for its Office of the Future. https://www.johnsoncontrols.com/media-center/news/press-releases/2019/04/23/middle-east-sustainability-pioneer-bee-ah-selects-johnson-controls-microsoft-for-office-of-future. (2019).

[2] 2021. Global AI Challange. https://www.globalaichallenge.com/en/home. (2021).

[3] H. Afzal, M. A. Khan, K. ur Rehman, I. Ali, and S. Wajahat. 2010. Consumer's trust in the brand: Can it be built through brand reputation, brand competence and brand predictability. *International business research* 3, 1 (2010), 43.

[4] B. Brown. 2001. Studying the internet experience. *HP laboratories technical report HPL* 49 (2001).

[5] J. J. Bute and M. Brann. 2015. Co-ownership of private information in the miscarriage context. *J. Applied Communication Research* 43, 1 (2015), 23–43.

[6] D. Cook, M. Schmitter-Edgecombe, A. Crandall, C. Sanders, and B. Thomas. 2009. Collecting and disseminating smart home sensor data in the CASAS project. In *Proc. ACM CHI'09*. 1–7.

[7] K. De Wulf, G. Odekerken-Schröder, and D. Iacobucci. 2001. Investments in consumer relationships: A cross-country and cross-industry exploration. *Journal of marketing* 65, 4 (2001), 33–50.

[8] T. Dinev and P. Hart. 2006. An extended privacy calculus model for e-commerce transactions. *Information systems research* 17, 1 (2006), 61–80.

[9] O. D. Duncan. 1975. Introduction to structural equation models. (1975).

[10] T. Ekwevugbe, N. Brown, and D. Fan. 2012. A design model for building occupancy detection using sensor fusion. In *Proc. IEEE DEST'12*. 1–6.

[11] C. Fornell and D. F. Larcker. 1981. Evaluating structural equation models with unobservable variables and measurement error. *Journal of marketing research* 18, 1 (1981), 39–50.

[12] J. Granderson, G. Lin, A. Harding, P. Im, and Y. Chen. 2020. Building fault detection data to aid diagnostic algorithm creation and performance testing. *Scientific data* 7, 1 (2020), 1–14.

[13] J. F. Hair. 2009. Multivariate data analysis. (2009).

[14] C. Hsu and J. C. Lin. 2016. An empirical examination of consumer adoption of Internet of Things services: Network externalities and concern for information privacy perspectives. *Computers in Human Behavior* 62 (2016), 516–527.

[15] E. D. Jaspers and E. Pearson. 2022. Consumers' acceptance of domestic Internet-of-Things: The role of trust and privacy concerns. *J' Bus Res* 142 (2022).

[16] R. Jia, R. Dong, S. S. Sastry, and C. J Sapnos. 2017. Privacy-enhanced architecture for occupancy-based HVAC control. In *Proc. ACM/IEEE ICCPS'17*. IEEE.

[17] D. Kahneman and A. Tversky. 1979. Prospect Theory: An Analysis of Decision under Risk. *Econometrica* 47, 2 (1979), 263–292.

[18] D. Kim, K. Park, Y. Park, and j. Ahn. 2019. Willingness to provide personal information: Perspective of privacy calculus in IoT services. *Computers in Human Behavior* 92 (2019), 273–281.

[19] R. B. Kline. 2015. *Principles and practice of structural equation modeling*. Guilford publications.

[20] N. Kordzadeh, J. Warren, and A. Seifi. 2016. Antecedents of privacy calculus components in virtual health communities. *International Journal of Information Management* 36, 5 (2016), 724–734.

[21] T. Kowatsch and W. Maass. 2012. Critical privacy factors of internet of things services: An empirical investigation with domain experts. In *Proc. MCIS'12*. Springer.

[22] T. Kriechbaumer and H. Jacobsen. 2018. BLOND, a building-level office environment dataset of typical electrical appliances. *Scientific data* 5, 1 (2018), 1–14.

[23] R. LaRose, N. Rifon, S. Liu, and D. Lee. 2005. Online safety strategies: a content analysis and theoretical assessment. In *Proc. ICA'05*.

[24] R. S. Laufer and M. Wolfe. 1977. Privacy as a concept and a social issue: A multidimensional developmental theory. *J. social Issues* 33, 3 (1977).

[25] Y. Li. 2014. The impact of disposition to privacy, website reputation and website familiarity on information privacy concerns. *Decision support systems* 57 (2014).

[26] M. A Lisovich, D. K. Mulligan, and S. B. Wicker. 2010. Inferring personal information from demand-response systems. *IEEE S.&P.* 8, 1 (2010), 11–20.

[27] N. K. Malhotra, S. S. Kim, and J. Agarwal. 2004. Internet users' information privacy concerns (IUIPC): The construct, the scale, and a causal model. *Information systems research* 15, 4 (2004), 336–355.

[28] O. Mazhelis, E. Luoma, and Warma. 2012. Defining an internet-of-things ecosystem. In *Internet of things, smart spaces & next generation N.* Springer, 1–14.

[29] C. J. Meinrenken, N. Rauschkolb, S. Abrol, T. Chakrabarty, V. C. Decalf, C. Hidey, K. McKeown, A. Mehmani, V. Modi, and P. J. Culligan. 2020. MFRED, 10 second interval real and reactive power for groups of 390 US apartments of varying size and vintage. *Scientific Data* 7, 1 (2020), 1–9.

[30] A. Meyerson and R. Williams. 2004. On the complexity of optimal k-anonymity. In *Proc. ACM PODS'04*. 223–228.

[31] C. Miller and F. Meggers. 2017. The Building Data Genome Project: An open, public data set from non-residential building electrical meters. *Energy Procedia* 122 (2017), 439–444.

[32] N Mohamed and I. H. Ahmad. 2012. Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from Malaysia. *Computers in Human Behavior* 28, 6 (2012), 2366–2375.

[33] B. Morvaj, L. Lugaric, and S. Krajcar. 2011. Demonstrating smart buildings and smart grid features in a smart energy city. In *Proc. IYCE'11*. IEEE.

[34] H. Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004).

[35] M. Pipattanasomporn, G. Chitalia, J. Songsiri, C. Aswakul, W. Pora, S. Suwankawin, K. Audomvongseree, and N. Hoonchareon. 2020. CU-BEMS, smart building electricity consumption and indoor environmental sensor datasets. *Scientific Data* 7, 1 (2020), 1–14.

[36] J. Reyes-Campos, G. Alor-Hernández, I. Machorro-Cano, J. O. Olmedo-Aguirre, J. L. Sánchez-Cervantes, and L. Rodríguez-Mazahua. 2021. Discovery of resident behavior patterns using machine learning techniques and IoT paradigm. *Mathematics* 9, 3 (2021), 219.

[37] F. C. Sangogboye, R. Jia, T. Hong, C. Spanos, and M. B. Kjærgaard. 2018. A framework for privacy-preserving data publishing with enhanced utility for cyber-physical systems. *ACM TOSN* 14, 3-4 (2018), 1–22.

[38] Y. Sun, N. Wang, X. Shen, and J. X. Zhang. 2015. Location information disclosure in location-based social network services: Privacy calculus, benefit structure, and gender differences. *Computers in Human Behavior* 52 (2015), 278–292.

[39] K. S. Taber. 2018. The use of Cronbach's alpha when developing and reporting research instruments in science education. *Res Sci Educ* 48, 6 (2018), 1273–1296.

[40] T. Wang, T. D. Duong, and C. C. Chen. 2016. Intention to disclose personal information via mobile applications: A privacy calculus perspective. *International journal of information management* 36, 4 (2016), 531–542.

[41] X. Wang and P. Tague. 2014. Non-invasive user tracking via passive sensing: Privacy risks of time-series occupancy measurement. In *Proc. AISec '14*.

[42] C. Weiller and A. Neely. 2013. Business model design in an ecosystem context. *University of Cambridge, Cambridge Service Alliance* (2013), 1–21.

[43] A. F. Westin. 1968. Privacy and freedom. *Washington and Lee Law Review* 25, 1 (1968), 166.

[44] H. Xu, X. R. Luo, J. M. Carroll, and M. B. Rosson. 2011. The personalization privacy paradox: An exploratory study of decision making process for location-aware marketing. *Decision support systems* 51, 1 (2011), 42–52.

[45] Z. Zheng, Q. Chen, C. Fan, N. Guan, A. Vishwanath, D. Wang, and F. Liu. 2018. Data driven chiller sequencing for reducing hvac electricity consumption in commercial buildings. In *Proc. ACM e-Energy'18*.