Received: 26 December 2021 | Revised: 13 February 2022 | Accepted: 24 February 2022 | Published on: 13 March 2022

DOI: 10.1002/spy2.225

RESEARCH ARTICLE

WILEY

From data flows to privacy-benefit trade-offs: A user-centric semantic model

Yang Lu[1] | **Shujun Li**[2]

[1]School of Science, Technology and Health, York St John University, York, UK

[2]School of Computing, University of Kent, Kent, UK

Correspondence
Yang Lu, School of Science, Technology and Health, York St John University, York, UK.
Email: y.lu@yorksj.ac.uk

Funding information
Engineering and Physical Sciences Research Council, Grant/Award Number: EP/R033749/1

Abstract

In today's highly connected cyber-physical world, people are constantly disclosing personal and sensitive data to different organizations and other people through the use of online and physical services. This is because sharing personal information can bring various benefits for themselves and others. However, data disclosure activities can lead to unexpected privacy issues, and there is a general lack of tools that help to improve users' awareness of the subtle privacy-benefit trade-offs and to make more informed decisions on their data disclosure activities in wider contexts. To fill this gap, this paper presents a novel user-centric, data-flow graph based semantic model, which can show how a given user's personal and sensitive data have been disclosed to different entities and what benefits the user gained through such data disclosures. The model allows automatic analysis of privacy-benefit trade-offs around a target user's data sharing activities, therefore it can support development of user-centric software tools for people to better manage their data disclosure activities to achieve a better balance between privacy and benefits in the cyber-physical world.

KEYWORDS

consumer behavior, cost benefit analysis, cyber-physical systems, data privacy, information security, network theory (graphs), semantic web, social network services, ubiquitous computing

1 | INTRODUCTION

Living in a highly digitized and networked world and the wider cyber-physical space, people are interacting with organizations and other people more and more frequently via different kinds of online and offline (physical) services and products. For instance, through using travel agencies (eg, Agoda and Booking.com) online or via physical means, people can arrange flight tickets, hotel rooms, transportation choices and tourist activities. In addition to providing basic services, it is a common practice for service providers to share customers' personal data with other third-party organizations, such as advertisers, insurers and relevant governmental bodies, due to legal requirements or some business reasons (eg, to offer more personalized services). Furthermore, many people actively share information about their lives online with other people, for example, on online social networks (OSNs) and web forums,[1,2] which further extends the scale of data sharing. While submitting personal information to a service application, people often follow a process known as privacy calculus in the literature to make decisions on data disclosures.[3] Particularly, the privacy calculus refers to "a cost-benefit trade-off analysis that accounts for inhibitors and drivers that influence the decision on whether to disclose information or not."[4]

It has been observed that individuals often act irrationally when facing privacy sensitive decisions in real world. This effect is known as privacy paradox in the literature,[5-7] which is about the deviation of people's actual decisions from their intentions toward disclosure of personal data. For instance, people may be motivated to disclose more personal information for certain additional benefits, forgetting about or ignoring their concerns on privacy. This has been reported widely in the literature, for example, Lee et al.[8] found that monetary rewards could encourage consumers to share more personal data for personalized services, and via an empirical study involving 259 subjects Krasnova et al.[9] showed that people were motivated to disclose personal data on OSNs mainly for benefits such as maintaining and developing relationships and platform enjoyment, which was perceived as conflicting with the need for privacy.

All such data sharing activities can lead to different kinds of privacy issues, caused by personal data flowing from the user (ie, the data owner) and devices to different entities in the cyber-physical world, directly or indirectly.[10-12] Past work was mostly designed to address "known events" such as decisions on data collection, access, and processing, however insufficient work has been done towards privacy issues related to data flows unknown to users. To help identify what self-disclosure activities cause privacy issues, it is necessary to keep users aware of data flows that can lead to possible privacy issues. In this context, many researchers have proposed to use a privacy related ontology or other conceptual models to systematically formalize knowledge about privacy by "explicit concepts and relations," in order to discover "implicit facts" (ie, privacy issues or risks).[13] In addition, some researchers advocated the use of transparency-enhanced technologies (TETs)[14] and digital privacy nudges[15] for providing effective awareness notices. With enhanced awareness, further privacy enhancement mechanisms can be adopted to help people manage such privacy risks, for example, adjusting access control or privacy policies, removing unused data, switching to more privacy-friendly services, and using privacy software tools to automatically block unwanted data disclosures.

Most past theoretical work on privacy ontologies and conceptual modeling focuses either on high-level concepts or a narrow aspect or application domain (eg, privacy policies, OSNs). Except our preliminary work on the data-flow graph model[1] and how such a model can be used to build a user-centric privacy protection solution,[2] which form the basis of this extended work, we have not seen any work focusing on user-centric data flows across different types of data consumers (services, organizations, other people, etc.). In addition, added values (ie, benefits) gained from using online services should be explored as the supplement. By further extending our preliminary work mentioned above, this paper fills this gap by describing a novel user-centric and graph-based model for formalizing personal data and value flows that can allow joint analysis of privacy-benefit trade-offs. The model is generic enough to cover a wide range of data disclosure activities of people in the cyber-physical world. The model can be seen as an privacy-oriented data disclosure ontology, allowing manual and automatic analysis of known and unknown privacy issues represented as special topological patterns on a directed graph. Besides, the model also supports conceptualization and automatic inference of benefits generated by data disclosures, therefore allowing automatic analysis of the privacy-benefit trade-offs. The model lays the theoretical foundation of software tools that can be used by individual users themselves (ie, data owners rather than organizations and researchers) to monitor their data disclosure activities and help provide opportunities to adapt their behaviors toward a better trade-off between privacy protection and benefits gained through data disclosures.

The rest of the paper is organized as follows. Section 2 defines the proposed model in details. A number of case studies in two application categories are discussed in Section 3, in order to demonstrate how the proposed model can be used to identify different types of privacy issues. In Section 4, we discuss how automated semantic reasoning can be done based on the proposed model, which can be implemented with existing web ontology tools. Other related works and possible future directions are discussed in Sections 6 and 7, respectively.

2 | THE PROPOSED MODEL

In this section, we first give two example scenarios about the privacy-benefit trade-offs related to data disclosures, to illustrate real-world problems the proposed model aims at addressing. Then, we formally explain basic concepts behind the proposed model. Finally, we show how privacy issues and benefits (especially added values generated by sharing personal data) can be identified by analyzing different types of edges in the proposed graph-based model.

2.1 | Example scenarios

As stated, the proposed model aims at empowering users with more knowledge (ie, awareness) on their real-world data disclosure activities in the cyber-physical world, and offering them computational tools to balance their needs for privacy and benefits gained from such activities. The following two example scenarios help illustrate what privacy-benefit trade-offs users may face to and what helps the proposed model can provide.

Scenario 1: Online bookings with travel service providers (Data disclosed to organizations).

As an experienced Internet user, Alice is planning to use different online travel services to arrange her next trip to China. She knows that she has to share certain personal information with such services in order to make the bookings, but as a very privacy-aware customer she would like to have a better understanding of what organizations behind those services actually see the data she will disclose. Having some good knowledge of how organizations work together to provide online services, she worries that some of her sensitive personal data may end up with some organizations she does not trust without her knowledge. She also worries that some organizations may get too much data about her so that her geo-locations can be tracked. While being worried about her privacy, Alice also wants to make sure she gets the maximum benefits by optimizing the decision she makes upon such bookings since different online services offer different types of rewards to attract new customers and retain existing customers. Now she hopes to balance the need of privacy and the benefits she will receive from such bookings.

Scenario 2: Using online social networks (Data disclosed to other people).

Now Alice is in China after making all the bookings needed. Since she has seen so many interesting things and met so many interesting people in China, she is keen to share such experiences with her family, friends, colleagues and other people in her home country who have not got a chance to visit China. As an active user of OSNs such as Facebook, Twitter, and Instagram, she would like to share her travel experience on such platforms which allow her to socialize with other people more effectively and efficiently. Sharing her happiness on OSNs makes herself enjoy the travel even more, and she would also like to receive some new likes and followers on her OSN accounts (probably from China!). Since she is a very privacy-aware person, she would like to avoid over-sharing her geo-locations in real time through her text messages, photos and videos, and would like to track how personal information she posts on OSNs spread beyond people she knows well. This is important for her to make better decisions on what information to share, on what platform(s), with whom, when and how in future.

2.2 | The model: Basic concepts

At a higher level of conceptualization, our proposed model can be formalized as a directed graph describing how personal data of a of people can *possibly* flow through (ie, may be disclosed to) different types of entities in a cyber-physical world, as shown in Figure 1.[*] Mathematically, such a graph can be denoted by $\mathscr{G} = (\mathscr{V}, \mathcal{E})$, where $\mathscr{V} = \{\mathscr{V}_i\}_{i=1}^{M}$ is a set of M nodes and each node $\mathscr{V}_i$ represents a specific type of entities with the same semantic meaning in our model (depicted by ellipses), and $\mathcal{E} = \left\{\mathcal{E}_j\right\}_{j=1}^{N}$ is a set of N edges and each edge $\mathcal{E}_j$ represents a specific type of relations[†] between two entity types. Edges in $\mathscr{G}$ can be categorized into three different groups: edges representing *semantic relations*, *data flows*, and *value flows*, which are depicted by solid, black and gray dashed arrows in Figure 1, respectively. Note that in Figure 1, when there is "…" included in the textual label of an edge there should actually be multiple edges (only one is shown for the sake of simplicity) due to the existence of multiple semantic relations between the two corresponding entity types (eg, a service is provided by a company but owned by another, which have different implications on data flows). In the current model, there are $M = 8$ different entity types and a number of edge types between different entity types. These numbers can be increased if the model is extended further.

The *entity type level* graph $\mathscr{G}$ can only show entity types and *possible* relations between different entities, but not the actual entities and relations (eg, concrete data flows between two organizations/people) that are what we need to work with for detecting and analyzing privacy issues. To this end, we will need *entity level* graphs. Each of such graphs is a *different* directed graph $\boldsymbol{G} = (\mathbb{V}, \mathbb{E})$, where $\mathbb{V} = \{v | v \in \mathscr{V}_i, 1 \leq i \leq M\}$ is a set of nodes each representing an entity (ie, an instance of a specific entity type/node in $\mathscr{G}$) and $\mathbb{E} = \left\{e | e \in \mathcal{E}_j, 1 \leq j \leq N\right\}$ is a set of edges each representing a relation (ie, an instance of a specific relation type/edge in $\mathscr{G}$). Some concrete examples of such entity level models/graphs will be given in Section 3.

The entity types can be categorized into three groups: (1) physical entities that exist only in the physical world; (2) cyber entities that exist only in the cyber world (from user's perspective); (3) hybrid entities that may exist in both the cyber and physical worlds. In Figure 1, the eight different entity types are colored differently to show which group(s) each entity type belongs to (gray: physical, white: cyber, gradient: hybrid). In the following we explain what these types represent.

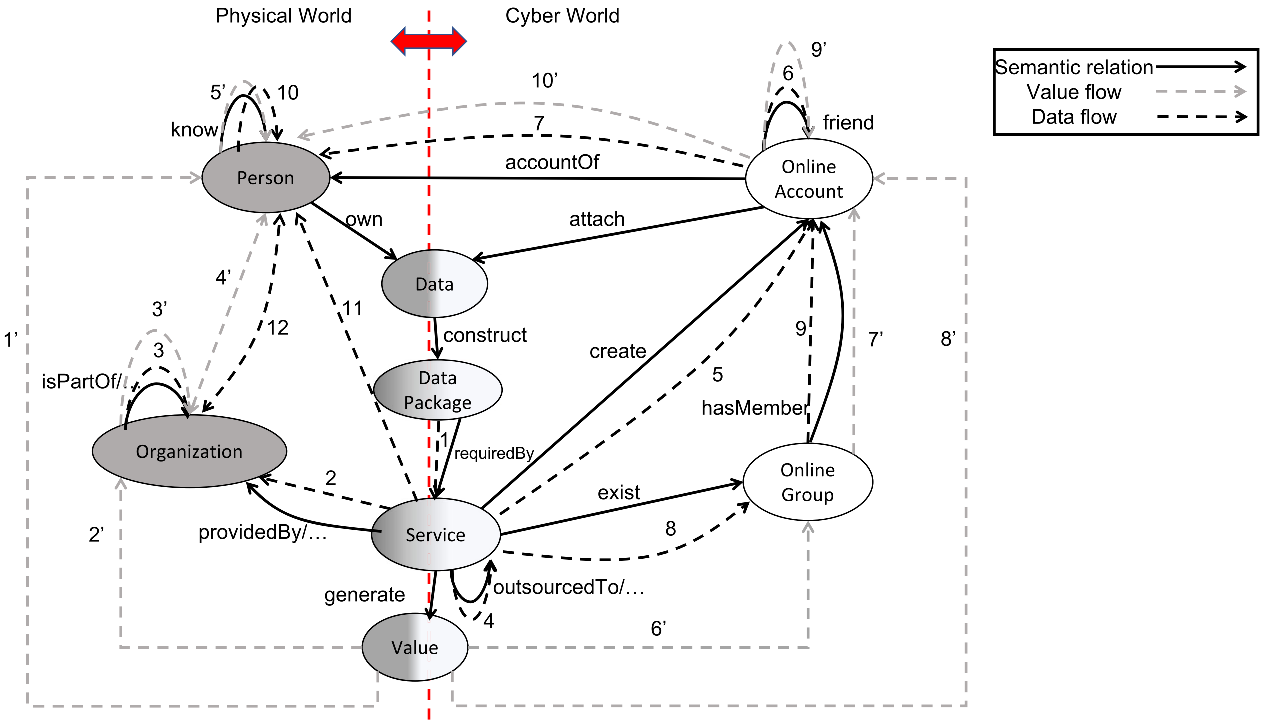


FIGURE 1 The entity-type graph of proposed model

Person (P) stands for natural people in the physical world. The model is *user-centric*, that is, about a special P entity "me" - the user for whom the model is built. The model will include other people as well because privacy issues of "me" can occur due to the data flowing to other people who interact directly or indirectly with "me."

Data (D) refers to *atomic* data items about "me" (eg, "my name"). Data entities may be by nature in the physical world, or in the cyber world, or in both worlds.

Value (V) refers to different types of "benefits" generated in the process of people using online services. Such benefits can be received by users or organizations, and can be transferred between different entities. For the purpose of privacy-benefit trade-off analysis, we focus more on benefits received by the target user "me," but it is necessary to model the concept of benefits more broadly to capture more complicated case of benefit transfer. For instance, by sharing user generated contents containing personal data on an OSN, a user can bring new users and ultimately new incomes to the platform, and the organization running the platform can decide whether to transfer parts of such benefits back to the user to encourage his/her engagement on the platform (as what has been happening on C2C and P2P platforms such as eBay and YouTube).

Service (S) refers to different physical and online services that serve people for a specific purpose (eg, a travel agent helping people to book flights).

Data Package (DP) refers to specific combinations of data entities required by one or more services. In this model, DP entities can be seen as encapsulated data disclosed in a single transaction.

Organization (O) refers to organizations that relate to one or more services (eg, service providers).

Online Account (OA) refers to "virtual identities" existing on online services.

Online Group (OG) refers to "virtual groups" of online accounts that exist on a specific online service.

2.3 The model: Edges

As stated before, each edge (ie, relation type) in the entity level graph $\mathscr{G}$, and hence each edge (ie, relation of a specific type) in an entity level graph $\boldsymbol{G}$, belongs to one of two groups of edges (relations). We explain these two edge groups in greater details below:

Semantic relations. The first edge group is about semantic relations that may or may not relate directly to personal data flows. For instance, the edge connecting entity types P and D means that the special P entity "me" owns some personal data items. Unlike the second group of edges that can cause immediate privacy impacts, the first group of edges help modeling the "evidence" about how and why data may flow among these entities.

Data flows. The second edge group is about data flows from a source entity to a destination entity. Most edges in this group are accompanied by semantic relation edges in the first group because the latter constructs the reason why a data flow can possibly occur.

Value flows. The third edge group is about value flows, generated from a destination entity and flowing to a source entity owning data items (P) or running services (O). On the graph, they can be found way back of edges in the second edge group. Likewise, value flows can be reasoned based on data flows and semantic relations.

To facilitate future discussions on data and value flows, we introduce more loosely defined concepts, "data flow edge type" and "value flow edge type" (and simply "edge type" when ambiguity or confusion will not arise) denoted by E_j and $E_{j'}$, the set of *all* data and value flow edges between a specific pair of entity types labeled by the same number j and j' in Figure 1. Accordingly, we use e_{j-k} (or $e_{j'-k}$) to denote the k-th edge of the loose edge type E_j (or $E_{j'}$) in an entity level graph $\boldsymbol{G}$, in order to give each individual edge in $\boldsymbol{G}$ a unique label. Note that E_j and $E_{j'}$ can cover multiple edges in $\mathscr{G}$ and $\boldsymbol{G}$ (eg, data flows between S and O entities) and it conceptually differs from $\mathcal{E}_j$ as the latter refers to different types of edges and also covers edges without a numeric edge label (eg, edges between P and D entity types in Figure 1).

The first data flow edge normally happens between DP and S entities, denoted by E_1. This is because before a data package is submitted to a service, no privacy issue can occur. Data shared with an online service can flow further to an organization via a business relationship such as *providedBy*, which is described by the edge type E_2. Further data flows can happen between different organizations, denoted by the edge type E_3 via business relationships such as *isPartOf*, *invest* and *collabrateWith*. In addition, there can be data flows between different services, donated by the edge type E_4, via business relationships such as *suppliedBy*, *poweredBy*, and *outsourcedTo*. The edge types E_5 and E_8 refer to data flows from an S entity to an OA or an OG entity (eg, data a user disclosed on an OSN is read by online accounts or people having read privileges in some online groups). The edge type E_6 refers to data flows between OA entities who are friends. The edge type E_7 refers to data flows from an OA to a P entity (ie, a human user of an online account). The edge type E_9 refers to data flows from OG to OA entities via a membership relationship. The edge type E_{10} refers to data flows caused by social relationships among people (eg, friendship and familial ties). The edge type E_{11} refers to data flows from an S entity directly to a person (ie, not via an OA entity), for example, a person can see public tweets on Twitter. The edge type E_{12} refers to potential *bidirectional* data flows between P and O entities, mapped to different types of semantic relations between P and O entities, for example, a person owns a company.

With personal data flowing to service providers and other people, corresponding benefits can be generated as the rewards (ie, returned values) for data subjects. In addition to semantic relations and potential data flows, Figure 1 shows that some types of values may be sent among different entity groups. While using online services and disclosing data to service providers, users can be offered benefits such as *cash return*, *discounts* when participating in their loyalty programs, which can be denoted by $E_{1'}$. Meanwhile, service providers may also receive benefits such as *advertising revenue* due to the frequent (service) usage and website visits. This value type is represented by $E_{2'}$. Similar to the data flows of E_3 type, information shared with business partners often helps to make better marketing strategies, which creates the values flowing along the same traces ($E_{3'}$). Following the edge types E_{12} and E_5, certain bi-directional value flows can be expected between P and O, as well as P and P entities ($E_{4'}$ and $E_{5'}$, respectively). Through posting data on OSNs, web forums, people can disclose their personal information for more attentions, higher reputations, enjoyments, and self-satisfactions. These are often passed in the forms of *reward/membership points*, *levels*, *likes*, *followers*, or *comments* in the cyber space. While participate in online activities, such benefits can also flow from an OG to OA entities included, and then to the P entities who own accounts, depicted as $E_{6'}$, $E_{7'}$, $E_{8'}$, and $E_{10'}$, respectively. Finally, values can also be transferred between different online accounts ($E_{9'}$) when they are communicating with each other in the cyber world.

The semantic relations, data, and value flows represented by edges between people (P), services (S), and organizations (O) can be complicated in real world, depending on how the business world works and how people and organizations are connected and interact with each other. Particularly, in Figure 1 on each edge (between S and O, from S to S and from O to O) there can be multiple different semantic relations, data or value flows, for example, a service is provided by an organization (ie, a service provider), a service is *outsourced to*, *supplied by* or *powered by* another service, an organization is *part of*, *in partnership with* or *invested by* another organization. In this work we do not intend to cover a complete list of such complicated business relations, but focus on the conceptual abstraction needed to capture all such relations.

Unlike privacy issues caused by data collection activities of services, privacy issues of online communities (such as OSNs) are mostly related to how well users manage the visibility of personal data.[16] For instance, with "friends only" and "members only" as privacy settings, contents shared on private spaces can be viewed by friends and group members only.

Such data flow edges are caused by semantic relations, for example, a person owns an online account; an online account is befriended with another account; a person is a friend of another person; and an online account is a member of an online group. Depending on what values (types) can be generated by the services (denoted by *generate*), users who start the data flows can achieve certain benefits from his/her data disclosure activities. In our proposed model, the edges between OA, OG, and P entities ($E_5, \dots, E_{10}$) and ($E_{6'}$,..., $E_{10'}$) describe how personal data and added value can possibly flow between such entities.

The coverage of the physical world part is very important because personal data and values can go from the cyber world to the physical world, and vice versa. In most cases, data going into the physical world are often less visible to the user and the user tends to have much less control over such data. With a fullerunderstanding of the cyber-physical data flow graph, a user can proactively do more to balance data privacy and benefits gained, in both cyber and physical worlds.

2.4 | "Topological" privacy issues

For a given user "me," if we can construct an *entity level* graph ***G***, which shows relevant entities, semantic relations, data and value flows, we will be able to identify different types of privacy issues concerning this given user, for example, if the user is disclosing too much information to a single service or organization, if the user has disclosed too much personal information to other people or the general public. Even when the graph ***G*** is incomplete, which is likely the case for most scenarios due to the lack of complete details about the user, some privacy issues may still be identified so such a model is still useful for the user. Meanwhile, a variety of added values flowing to "me" can be assessed in the context of privacy issues to inform the user about the privacy-benefit trade-offs. Such analysis can be automatically done by formalizing the relationships between data items/packages and benefits, which can be predicted based on the type of data shared and the business models of the relevant services and organizations.

Within the proposed model, we can define an important concept: a "data-flow path" is a sequence of consecutive data flows (edges in an entity level graph ***G***). This concept allows us to map different "privacy issues" to certain *topological* patterns that are formed by one or more data-flow paths. Different privacy issues may share the same topological pattern but follow different edges or different edge types, for example, one privacy issue may be related to one organization while another to a different organization. Beyond using the model to detect privacy issues, we can also try to quantify the risk of a given privacy issue and provide possible solutions to the user. Some concrete examples about such privacy issues will be discussed in the next section with a number of imaginary but realistic case studies. In addition to investigating privacy issues, it deserves mentioning that the proposed model can also find applications in other contexts, for example, studying how personal data are consumed by online services (even if there are no privacy issue for any particular user).

3 | CASE STUDIES

In this section, we use realistic examples in two broad categories to illustrate how entity level graphs can be built based on our proposed model and how privacy issues can be possibly identified. Note that in the graphs included in this section we will not show the whole entity level graph but only those elements needed, in order to better highlight the relevant data and value flows without complicating the look of the graphs. For instance, the P entity "me" may not be shown if a graph involves only "me" as a P entity, the edges between "me" and all D entities may not be shown if we focus on data flows from a single target user "me" only (which can be extended to cover multiple target users in our future work), and D entities may not be shown if the focus is more on DP entities. We also highlight P and O entities by filling the nodes with dark gray so that they stand out better in the graphs.

3.1 | Privacy issues related to service providers

Figure 2 shows the simplest model involving P, S, O, and V entities. An online service <service 1> connects to a service provider <provider 1> by a semantic relation edge *providedBy*, denoted by *providedBy*(service 1, provider 1). Through sending the data package <item 1>, the data owner <user 1> can use the service for certain purposes. For instance, an E_1 flow e_{1-1} at the beginning could cause an E_2 flow e_{2-1} from <service 1> to <provider 1>, denoted as e_{1-1}(item 1, service 1) and e_{2-1}(service 1, provider 1), respectively. Then, an $E_{1'}$ flow $e_{1'-1}$(value 1, user 1) will be generated from <service 1>

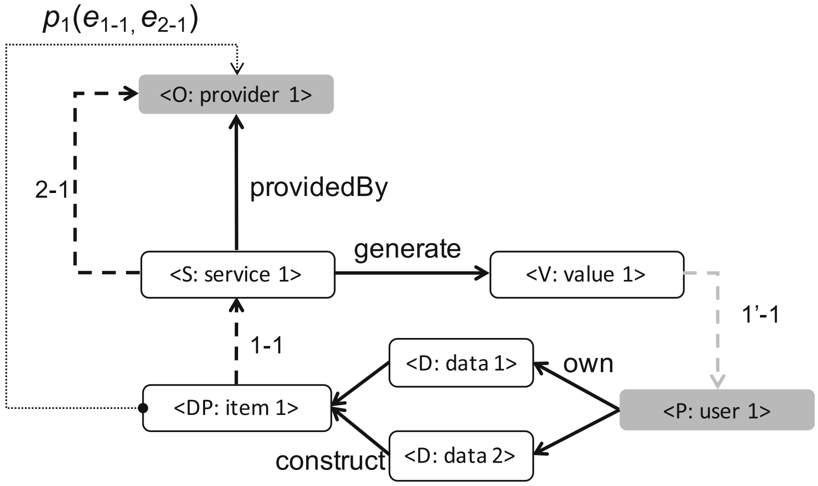


FIGURE 2 Example entity graph showing a data flow and a value flow

to <user 1>, triggered by e_{1-1}. As a result, there is only one path $p_1 = (e_{1-1}, e_{2-1})$ found from the source data <item 1> to the service provider <provider 1> in the physical world.[‡] Such a simple path does not normally lead to any privacy issue since it merely describes what data items are needed for a service to happen. In the following examples, we will show how non-trivial real privacy issues can be identified on more complicated data flow graphs.

In the real world, data flows can take place within a corporate family (connected by the semantic relation *isPartOf*). Therefore, it may be the case that different data items flow among multiple service providers and aggregate at a single organization, which may be unknown to the user thus leading to an unexpected and undesired privacy issue. For instance, in Figure 3, as <item 1> and <item 2> flow to <service 1> and <service 2> separately, E_2 flows e_{2-1} (service 1, provider 1) and e_{2-2} (service 2, provider 2) take place. Then, E_3 flows follow such as e_{3-1} (provider 1, provider 2), e_{3-2} (provider 1, provider 3), e_{3-3} (provider 2, provider 1) and e_{3-4} (provider 2, provider 3). Paths can be found from data packages <item 1> and <item 2> to service providers, <provider 1>, <provider 2>, and <provider 3>, such as $p_1 = (e_{1-1}, e_{2-1})$ and $p_6 = (e_{1-2}, e_{2-2}, e_{3-4})$. Inspecting the data flow graph, we see both data packages flow to the organization <provider a>, which may cause unknown disclosures of personal data. Assume the *added* value <value 1> is only generated by the data flow e_{2-1}(service 1, provider 1) and will flow to <me>, the owner of the data points <d1>. The edges $e_{1'-1}$ (value 1, me) can be inferred based on the data flows and semantic relations *generate* (service 1, value 1), *own* (me, d1) as well as *construct* (d1, item1). This allows us to study value flows from services to customers, and potentially study the motivation of sharing data their personal details with organizations. Note that all services normally generate some benefits for the service providers, and for the purpose of our proposed model we focus on those benefits that are relevant for the target user to explore the privacy-benefit trade-offs.

Complex business models exist in the real world. Figure 4 shows data flows among some business partners who jointly support online services. As shown in Figure 4A, an E_4 data flow e_{4-1} (service a, service b) can be found among the business partners connected by an *outsourcedTo* semantic relation edge. Based on an E_2 flow e_{2-1} (service b, provider b) and the service ownership expressed with the semantic relation edge *belongTo*, an E_3 flow e_{3-1} (provider b, provider a) can be identified. Similarly, Figure 4B shows E_4 flows that would incur due to the semantic relation edge *poweredBy* between online services, for example, e_{4-1} (service a, service 1) and e_{4-2} (service a, service 2), while in Figure 4C, the only E_4 flow e_{4-1} (service 1, service 2) is due to the semantic relation edge *suppliedBy* in between. If any of business relations between S and O entities are unknown, privacy concerns can arise at the user side.

To further illustrate how data flows in an entity level graph can be used to identify privacy issues, Figure 5 shows a scenario where a customer (a P entity) books flight tickets and hotels via online services provided by organizations Booking.com and Agoda. Privacy restrictions may be given to data items on pre-defined labels, such as *sensitive data items are not allowed to share with more than five organizations*. For this purpose, data entities are categorized in the following groups: **Profile** (Name, Age, Gender, and Email), **Event** (Itinerary, Companion, Dates, and Spending), **Location** (Destination, Landmark), **Sensitive** (Health), and **Entertainment** (Tour, Food). Sensitive data such as medical certificates may be required and shared with third-party suppliers, in case travelers need special medical assistance during travel. As a result, data package <item 1> will flow to 11 service providers along with paths p_1 to p_{11}. For instance,

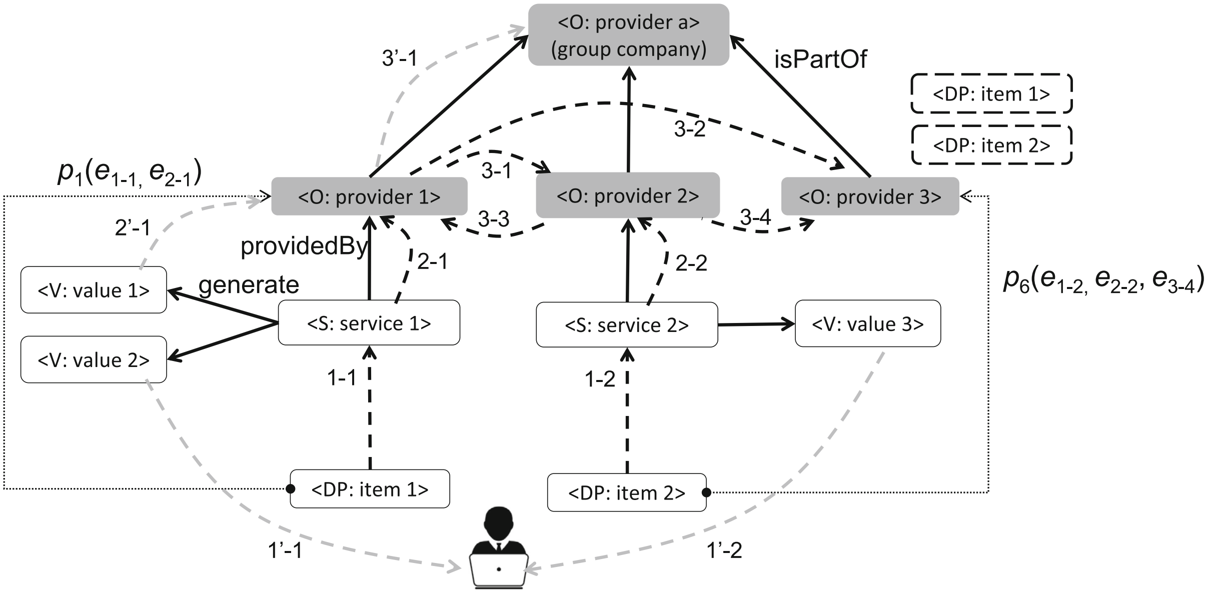


FIGURE 3 Example entity graph in provider hierarchies

paths $p_1 = (e_{1-1}, e_{4-1}, e_{2-1})$, $p_2 = (e_{1-1}, e_{4-1}, e_{2-1}, e_{3-1})$, and $p_{10} = (e_{1-1}, e_{4-1}, e_{2-1}, e_{3-9})$ can respectively lead data package <item 1> to <GoToGate>, <Booking>, and <SuperSaver>. Besides, the Agoda hotel booking service may incur data flows to seven service providers (led by paths p_{12} to p_{18}), such as $p_{12} = (e_{1-2}, e_{2-2})$ and $p_{13} = (e_{1-2}, e_{2-2}, e_{3-11})$ running to <Agoda> and <Kayak>. This may cause location privacy leakage if an O entity has the access to the user's <name> and <destination> simultaneously.

To guide the user to make decisions based on a trade-off analysis between potential privacy issues and benefits, it is essential to identify "what benefits can be achieved by using certain services and giving the consent to sharing personal data." For instance, Figure 6 shows different types of benefits offered by four different travel websites, Hotels.com, GoToGate, Booking.com, and Agoda with the same information <item 2> flowing to their accommodation booking services. Particularly, the Hotels.com allows the customers to collect their night stamps and redeem a reward one with 10 stamps. As a member of it, the user is also allowed to search within a Secret Prices system for lower prices.[§] Triggered by data flow edge e_{1-1}, certain value types <hotels_secretprice> and <hotels_hotelreward> can flow back through the edge $e_{1'-1}$. Likewise, the value edge $e_{1'-3}$ can take the value <booking_genius> to users who choose Booking.com. Its loyalty program Genius allows users to earn the benefits ranging from discounts to free breakfasts, as their transaction records unlock Genius levels.[¶] Through using the same service (disclosure), Agoda provides three types of benefits via the value edge $e_{1'-4}$, which can be categorized as **Monetary** (<agoda_cash>[**]) and **Non-monetary** (<agoda_VIP>[††] and <agoda_pointsmax>[‡‡]). The hotel-booking service of GoToGate is powered by Hotels.com, but Hotels.com does not provide rewards to users of GoToGate so disclosures on the edge e_{1-2} do not generate any added value.

3.2 | Unwanted disclosures to other people

In addition to privacy issues raised from data collection by service providers and data shared among services and organizations, online privacy issues may also be caused by unwanted data disclosures to other people, for example, on OSNs. Figure 7 is an entity level graph showing how the P entity <me> connects with other people through online and offline relations. Based on the friend relations between <fb_abc> and <ig_abc>, E_6 data flows such as e_{6-4} (fb_abc, fb_edward), e_{6-5} (ig_abc, ig_ed1989) could take place in the cyber space when "I" use Facebook and Instagram services and generate data flows e_{1-1}, e_{5-1}, e_{1-2}, and e_{5-2}. Given the account ownership, E_7 flows such as e_{7-4} (fb_edward, edward) and e_{7-5} (ig_ed1989, edward) will follow. Along with paths $p_4 = (e_{1-1}, e_{5-1}, e_{6-4}, e_{7-4})$ and $p_5 = (e_{1-2}, e_{5-2}, e_{6-5}, e_{7-5})$, it shows that both data packages <item 1> and <item 2> will be disclosed to <edward>. As a result, "my" current location may be inferred from the itinerary post on Facebook and landmark photos shared on Instagram during the trip. Meanwhile, data flows e_{5-1} and e_{5-2} to OSN services, <facebook> and <instagram> can bring certain benefits to "me."

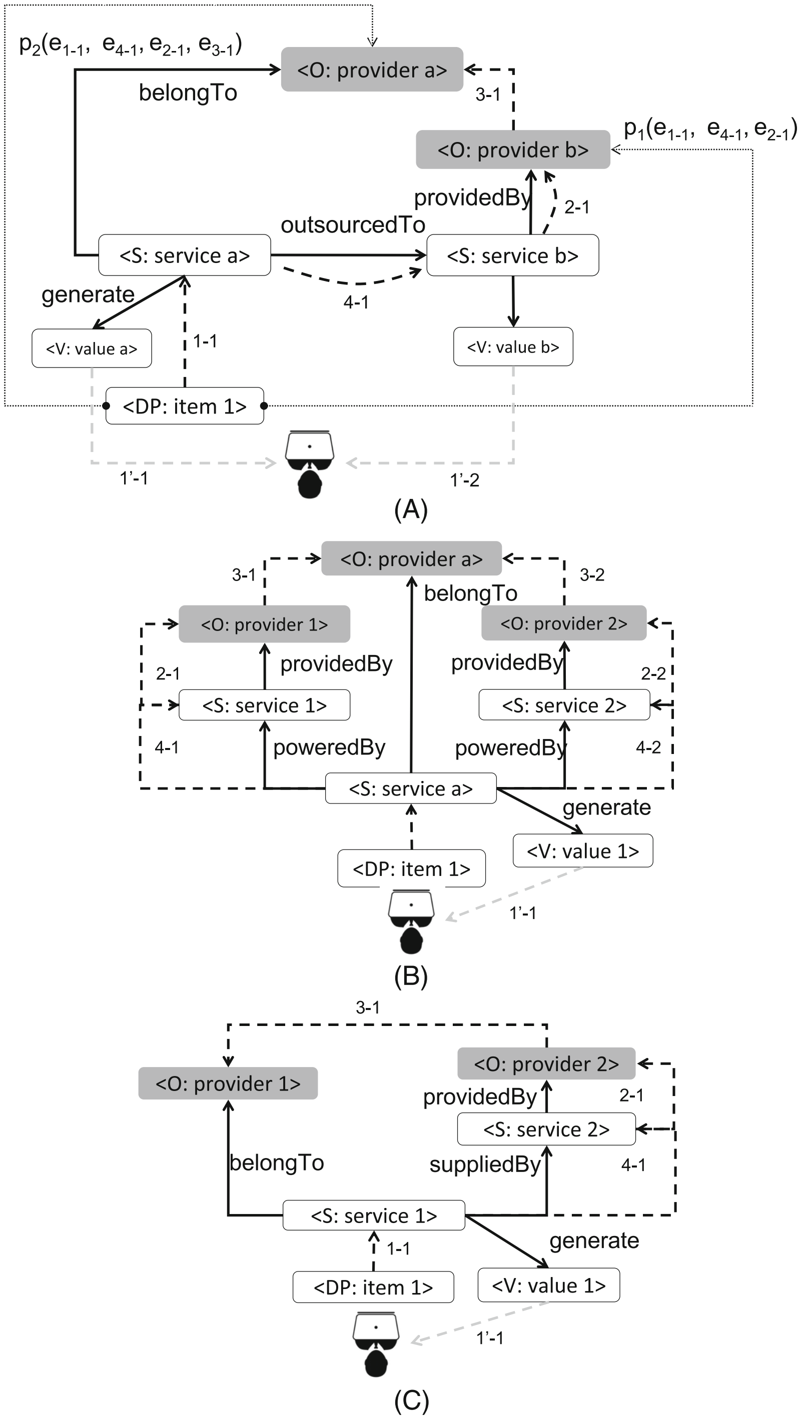


FIGURE 4 Example entity graphs of supply chains

Given the semantic model, such benefits can be inferred from the semantic relations *generate* (between S and V entities) and *accountOf* (between OA and P entities). For instance, given the relations *generate*(facebook, facebook_follow), *generate* (facebook, facebook_like), *accountOf* (fb_abc, me), as well as the data flows along with e_{5-1}, benefits such as <facebook_follow> and <facebook_like> can be automatically inferred. Such benefits refer to new Facebook users following "my" account <fb_abc>, or some Facebook followers clicking "Like" under "my" new posts. Similar inferences can be done for Instagram as new posts are created, that is, the value flow $e_{8'-2}$ is triggered by the data flow e_{5-2}.

Data visibility can be managed by privacy policies related to online friendships and memberships. As a result, privacy leakage could be caused when "I" permit unwanted access requests. Figure 8 shows a scenario where online data are

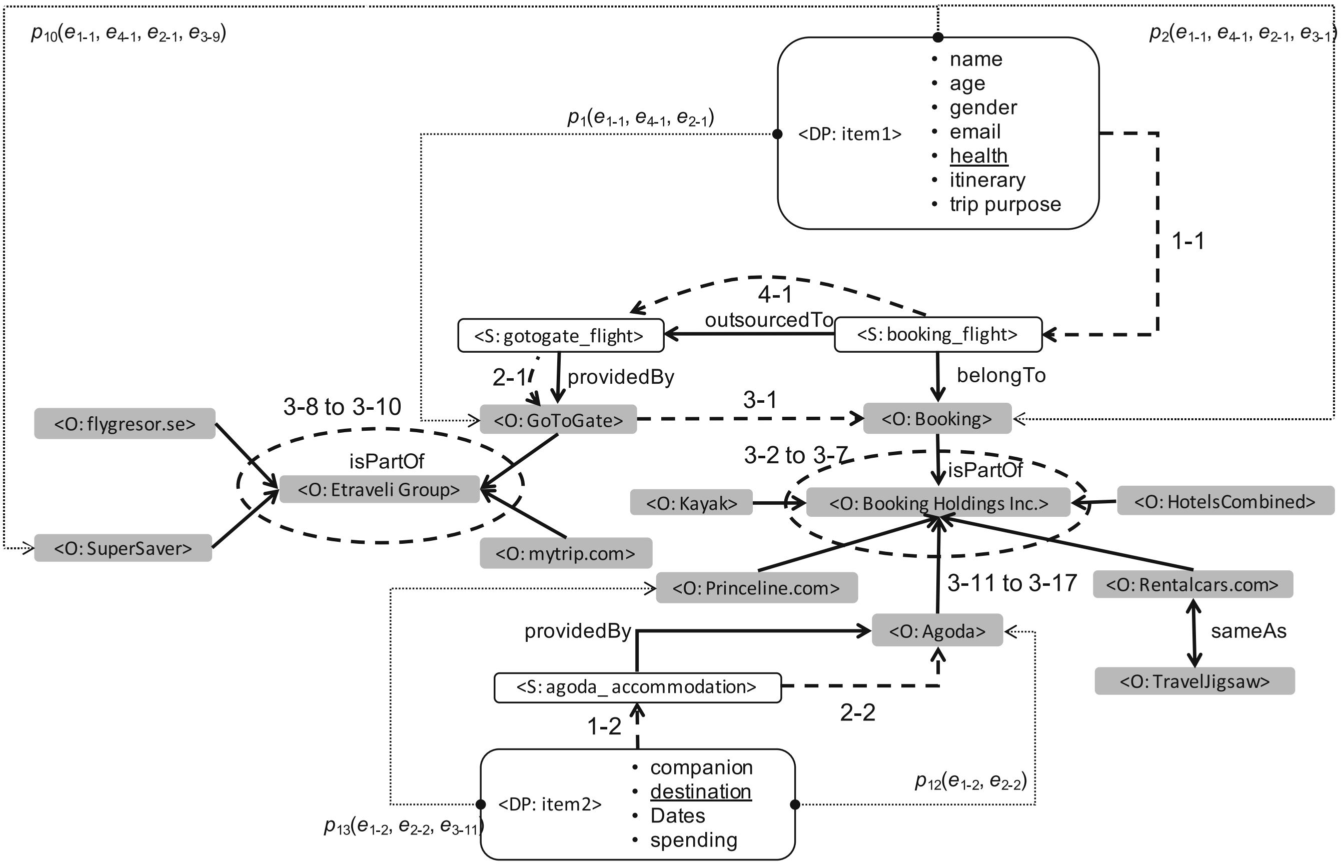


FIGURE 5 An example entity graph about data sharing in the travel context

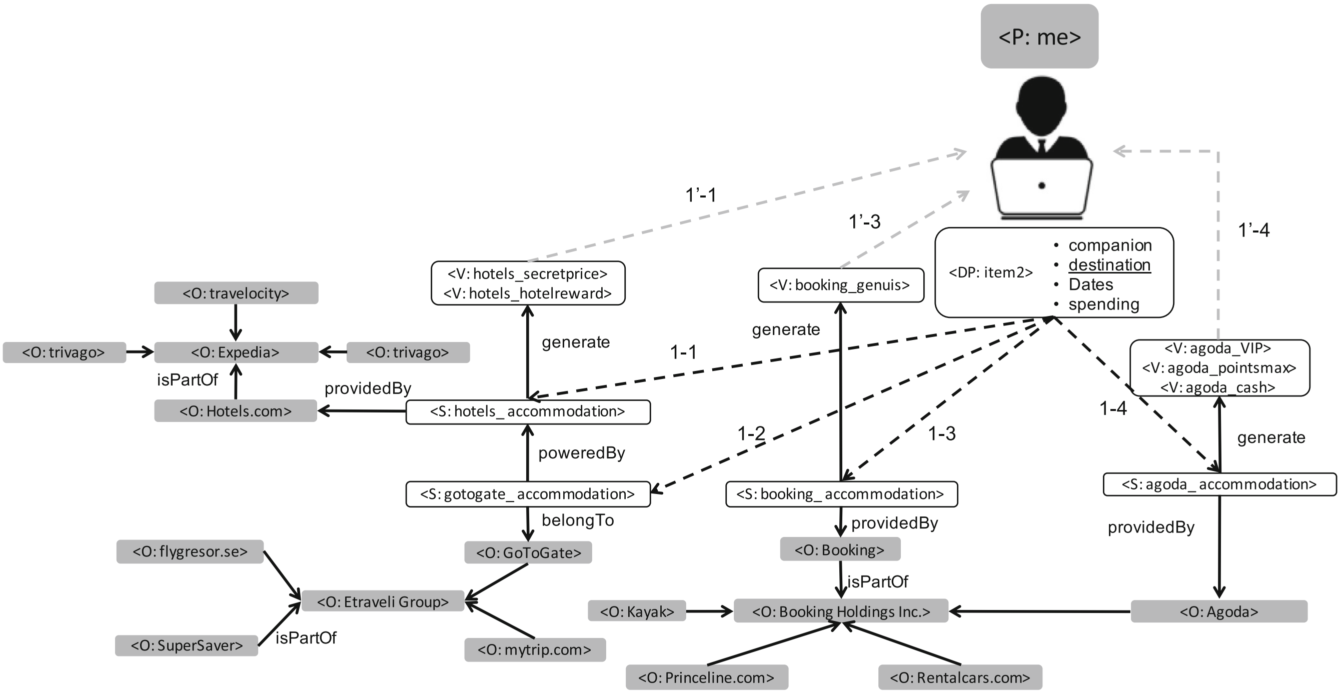


FIGURE 6 Value flows generated in hotel-booking services (the entity with a user icon indicates the special P entity "me" - the target user the model is serving; the same hereinafter)

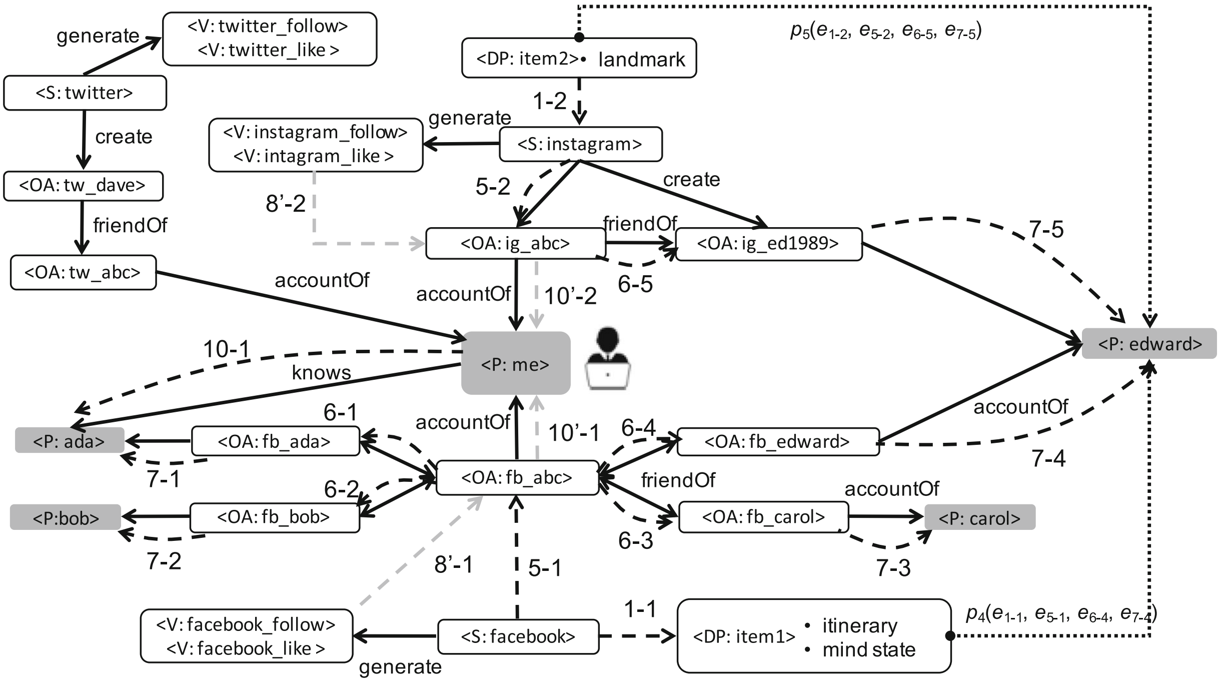


FIGURE 7 An example entity graph showing unwanted data disclosure on OSNs

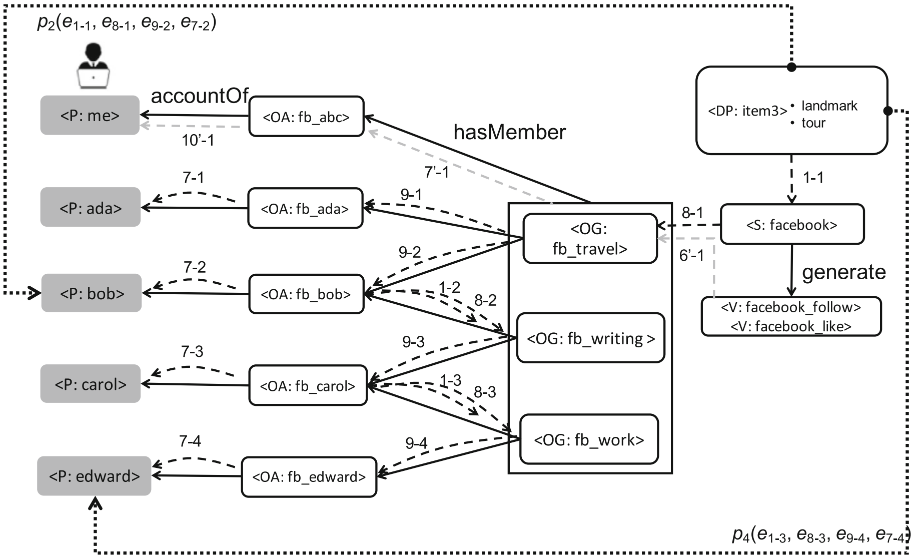


FIGURE 8 An example entity graph of cross-group data disclosure

propagated across groups that have members in common. Through E_9 flows e_{9-1} (fb_travel, fb_alice) and e_{9-2} (fb_travel, fb_bob), Alice and Bob can view <item 3> once "I" send it to a Facebook travel group the three people participate. In some situations, <item 3> can be resent to other groups and therefore generate some E_8 data flows, such e_{8-2} (fb_bob, fb_writing) and e_{8-3} (fb_carol, fb_work). Through the following E_9 and E_7 flows, <item 3> may be disclosed to unwanted members in the working group through the data-flow paths, $p_3 = (e_{1-3}, e_{8-2}, e_{9-3}, e_{7-3})$ $p_4 = (e_{1-3}, e_{8-3}, e_{9-4}, e_{7-4})$. Posting within a virtual group can lead to a wide range of data disclosures over Facebook. As a consequence, it is possible that shared posts get more "Likes" given by "strangers" accounts. Meanwhile, some of these accounts (like <fb_edward>) may start to follow <fb_abc> on Facebook. As shown in Figure 8, the E_8 flows to a virtual group will trigger the $E_{6'}$ flow $e_{6'-1}$, which taking the value <Facebook_Follow> and <Facebook_Like> to the group <fb_travel>. Instead of flowing to every group member, added values only flow to the online account of data subjects. The inference is defined in Section 4.

4 | AUTOMATED REASONING OF PRIVACY ISSUES AND BENEFITS

Web ontology language (OWL) and semantic web rule language (SWRL) have been widely utilized in specifying security and privacy policy constraints on data usage.[17-20] In this section, we use OWL and SWRL to formalize our model and show how reasoning can be done to detect privacy issues *automatically*. For the sake of simplicity, in this section we will focus on a subset of the entity types and relations.

4.1 | Semantic formalization

Following OWL and SWRL, different components in the proposed model can be defined as classes, predicates (with domains and values) and instances, as shown in Table 1. With the ontology and semantic rules (Rules 1–16) developed in Protégé 4.0 we can implement an automated semantic reasoning engine. Through running the reasoner Pellet[21] and description logic (DL) queries[22] on the knowledge base, implicit relations (ie, data flows) could be identified for privacy assessment and decision making purposes. Assuming that data flows to physical entities are likely causing privacy issues, privacy questions can be made to look for *finalFlowTo* (or *receive*) in the result sets. In the discussion on benefits as the returned values of utilizing online services (disclosures), we can also use *flowTo* to define the value flows and aggregate the benefits from *finalFlowTo* (or *receive*) clauses.

4.2 | Examples

In dealing with scenarios related to service providers, DL queries are utilized to answer the following questions: *"where the sensitive information flows to?"* and *"who can access the user profile and location at the same time?"* Through reasoning Rules 1–6 on the semantic graph of Figure 5, the engine shows that the number of service providers can be reduced by changing <flight_booking> to <flight_agoda> as the sensitive item <item 1> will be shared with one single corporate group, as shown in Figure 9. In a scenario about purchasing travel service packages, Figure 10 shows the result of comparing two service packages by running queries to answer *"who can access the user profile and location at the same time?"* Given the demand for booking "flights + hotels," the result sets show that adopting Package 2 can better control the privacy risks. In this case, query services can enhance user privacy by splitting personal details contained in data flows. By semantic Rule 7 we define how certain added values are triggered by a data flow to the online service, and then returned to the service user. In the example scenario of Figure 6, the user submits the same personal details to book a hotel from four different websites. According to query results to *"what added values can I achieve by booking from the website?"* it is clear to see that <accommodation_agoda> offers their customers more value classes (in Figure 11A) than other two services as the reward of usage (in Figures 11B,C).

Towards the privacy requirements in the scenarios concerning unwanted data disclosures to other people, Rules 8–11 and DL queries can be applied to check things such as if someone else can access certain data combinations or if entertainment-related messages are disclosed to colleagues. As illustrated in Figure 12, through querying on recipients *"who can access two data types during the same period,"* the system is expected to provide privacy suggestions such as *"blocking Facebook account fb_edward so as to stop such disclosure to Edward in the real world"* (see Figure 7). Similarly, a DL query can be made to check if certain data will flow to unwanted groups (recipients), such as *"Is there any post having entertainment related contents is visible to my colleagues?"* As shown in Figure 13, it shows <item 3> has breached personal privacy and thus demands for extra modification, like removing entertainment information from the Facebook post to <fb_travel>. In addition, the benefits from posting trip experiences on different OSN platforms/groups can be learned by reasoning Rules 12–14. For instance, through querying "what types of added values can be gained?" in the scenario of Figure 7, the added values on Facebook and Instagram can be offered to two different accounts and then the person "me," as shown in Figure 14. In addition to the value types, the model can be extended to calculate "to what extents the value can be achieved" in different use cases. Suppose that "my" Facebook friends may share my posts on their personal pages. Therefore, the number of viewers can depend on how many followers they have. Particularly, Rule 15 is to reason how many viewers on Facebook to the account which receives the value, Facebook_Followers. Suppose the Facebook friends Alice, Bob, Carol and Edward have 6, 3, 7, 19 followers respectively. As shown in Figure 15A, the reasoning results are denoted by *hasViewers_fb* (facebook:abc, INTEGER). In addition, the Facebook account levels up once any viewer group size is found larger than the original follower group. With the use of W3C built-in function *greaterThan*,[23] we specified

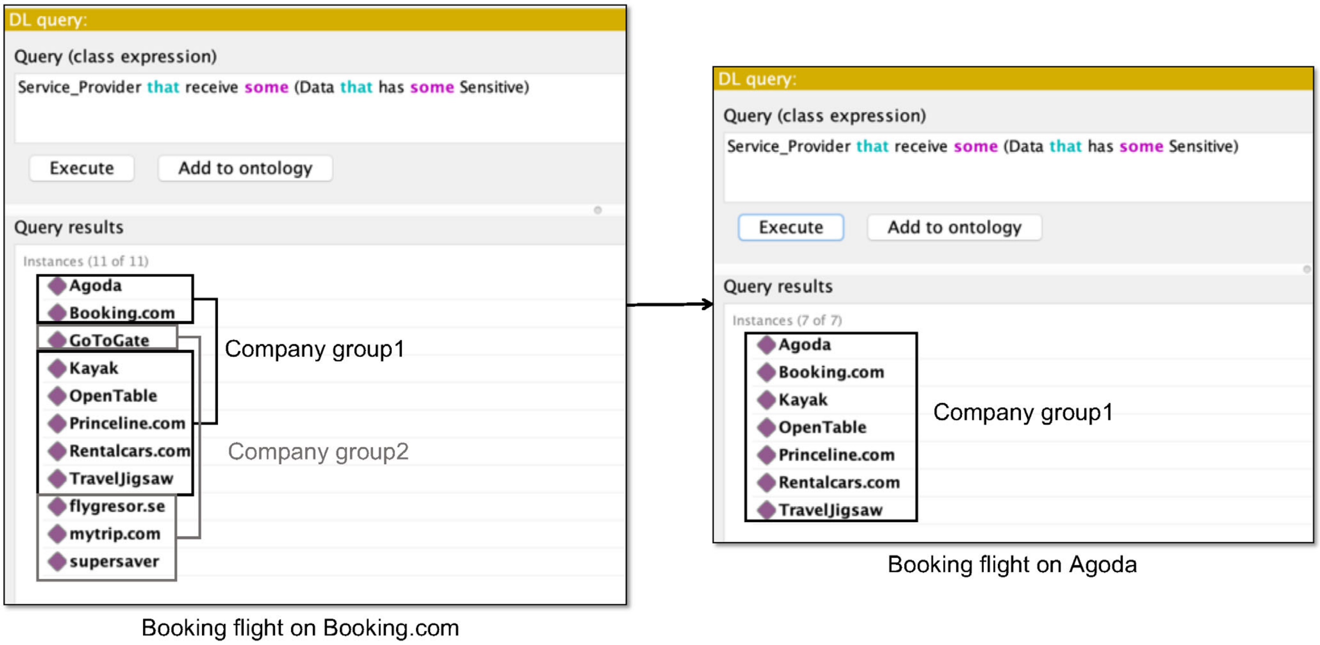
TABLE 1 Definitions of classes, predicates, and instances to represent different components of the proposed model

Class (domain)	Predicate	Range	Instance
Data_Package (DP)	*flowTo*	OA, OG, OS	item1, item2, item3, …
	finalFlowTo	P, SP	
	has	D	
Data (D)	*construct* (↔ *has*)	DP	itinerary, email, name, date_of_birth, …
Online_Account (OA)	*accountOf*	P	fb_ada, tw_dave, ig_ed1989, …
	friendOf	OA	
	memberOf	OG	
	hasFollowers_x[a]	INTEGER	
	hasViewers_x[b]	INTEGER	
	upgrade_x[c]	BOOLEAN	
Online_Group (OG)	*hasMember* (↔ *memberOf*)	OA	fb_travel, fb_writing, fb_work, …
Online_Service (OS)	*belongTo*	SP	flight_booking, accommodation_agoda, facebook, twitter, instagram, …
	providedBy	SP	
	outsourcedTo	OS	
	poweredBy	OS	
	suppliedBy	OS	
	create	OA	
	exist	OG	
	generate	V	
Service_Provider (SP)	*isPartOf*	SP	Booking, Agoda, TripAdvisor, …
	receive (↔ *finalFlowTo*)	DP	
Person (P)	*know*	P	ada, bob, me, dave, edward, …
	receive (↔ *finalFlowTo*)	DP	
	own	D	
Value (V)	*flowTo*	OA, OG	Agoda_Cash, Booking_Genius, Hotels_HotelReward, Facebook_Like, Instagram_Followers, …
	finalFlowTo	P	

[a] hasFollowers_x is a data property which describes "how many followers (accounts) the account owns on the platform x." Based on this, sub-properties can be defined for specific service provider, for example, hasFollowers_fb. Therefore, the range must be of the integer type.
[b] hasViewers_x is a data property which describes "how many strangers (accounts) can find the user account on the platform x." Based on this, sub-properties can be defined for specific service provider, for example, hasViewers_fb. Therefore, the range must be of the integer type.
[c] upgrade_x is a Data property which describes "whether the account levels up or not." Therefore, the range must be of the Boolean type (True/False).

the Rule 16 for checking if *upgrade_fb* (facebook:abc, true) can establish, meaning that the account facebook:abc gets upgraded. As shown in Figure 15B, querying *"which online accounts level up"* in this example scenario results in the name of "my" online account. This is to say, regardless of the privacy issues found from data flows to Facebook and Instagram, certain benefits such as upgrading some accounts can be achieved.

1. DP(?dp), flowTo(?dp, ?s), providedBy(?s, ?sp) → finalFlowTo(?dp, ?sp).
2. DP(?dp), flowTo(?dp, ?s), outsourcedTo(?s, ?s1), providedBy(?s1, ?sp) → finalFlowTo(?dp, ?sp).
3. DP(?dp), flowTo(?dp, ?s), poweredBy(?s, ?s1), providedBy(?s1, ?sp) → finalFlowTo(?dp, ?sp).
4. DP(?dp), flowTo(?dp, ?s), suppliedBy(?s, ?s1), providedBy(?s1, ?sp) → finalFlowTo(?dp, ?sp).
5. DP(?dp), flowTo(?dp, ?s), finalFlowTo(?dp, ?sp), belongTo(?s, ?sp1) → finalFlowTo(?dp, ?sp1).
6. SP(?sp), isPartOf(?sp, ?sp1), isPartOf(?sp2, ?sp1), finalFlowTo(?dp, ?sp) → finalFlowTo(?dp, ?sp2).



DL query:

Query (class expression)

Service_Provider that receive some (Data that has some Sensitive)

Query results

Instances (11 of 11)

- Agoda
- Booking.com
- GoToGate
- Kayak
- OpenTable
- Princeline.com
- Rentalcars.com
- TravelJigsaw
- flygresor.se
- mytrip.com
- supersaver

Booking flight on Booking.com

DL query:

Query (class expression)

Service_Provider that receive some (Data that has some Sensitive)

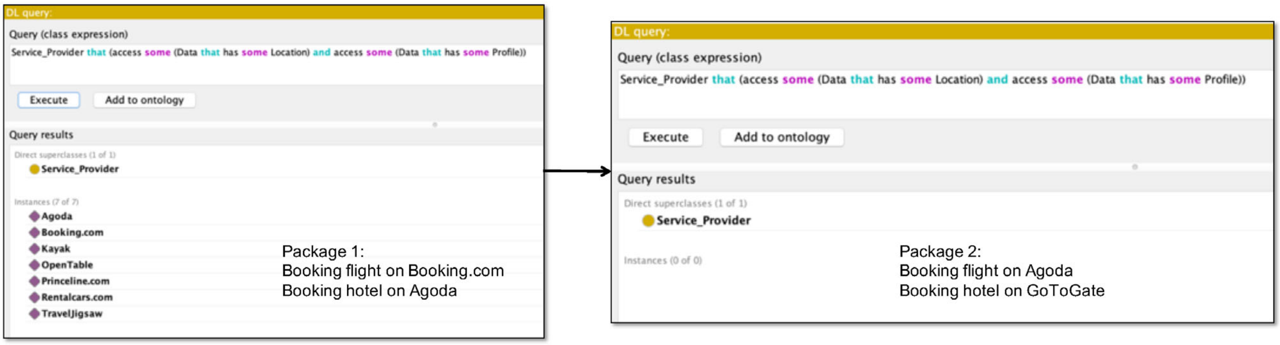
Query results

Instances (7 of 7)

- Agoda
- Booking.com
- Kayak
- OpenTable
- Princeline.com
- Rentalcars.com
- TravelJigsaw

Booking flight on Agoda

FIGURE 9 Example query on sensitive data disclosures



DL query:

Query (class expression)

Service_Provider that (access some (Data that has some Location) and access some (Data that has some Profile))

Query results

Direct superclasses (1 of 1)

- Service_Provider

Instances (7 of 7)

- Agoda
- Booking.com
- Kayak
- OpenTable
- Princeline.com
- Rentalcars.com
- TravelJigsaw

Package 1:
Booking flight on Booking.com
Booking hotel on Agoda

DL query:

Query (class expression)

Service_Provider that (access some (Data that has some Location) and access some (Data that has some Profile))

Query results

Direct superclasses (1 of 1)

- Service_Provider

Instances (0 of 0)

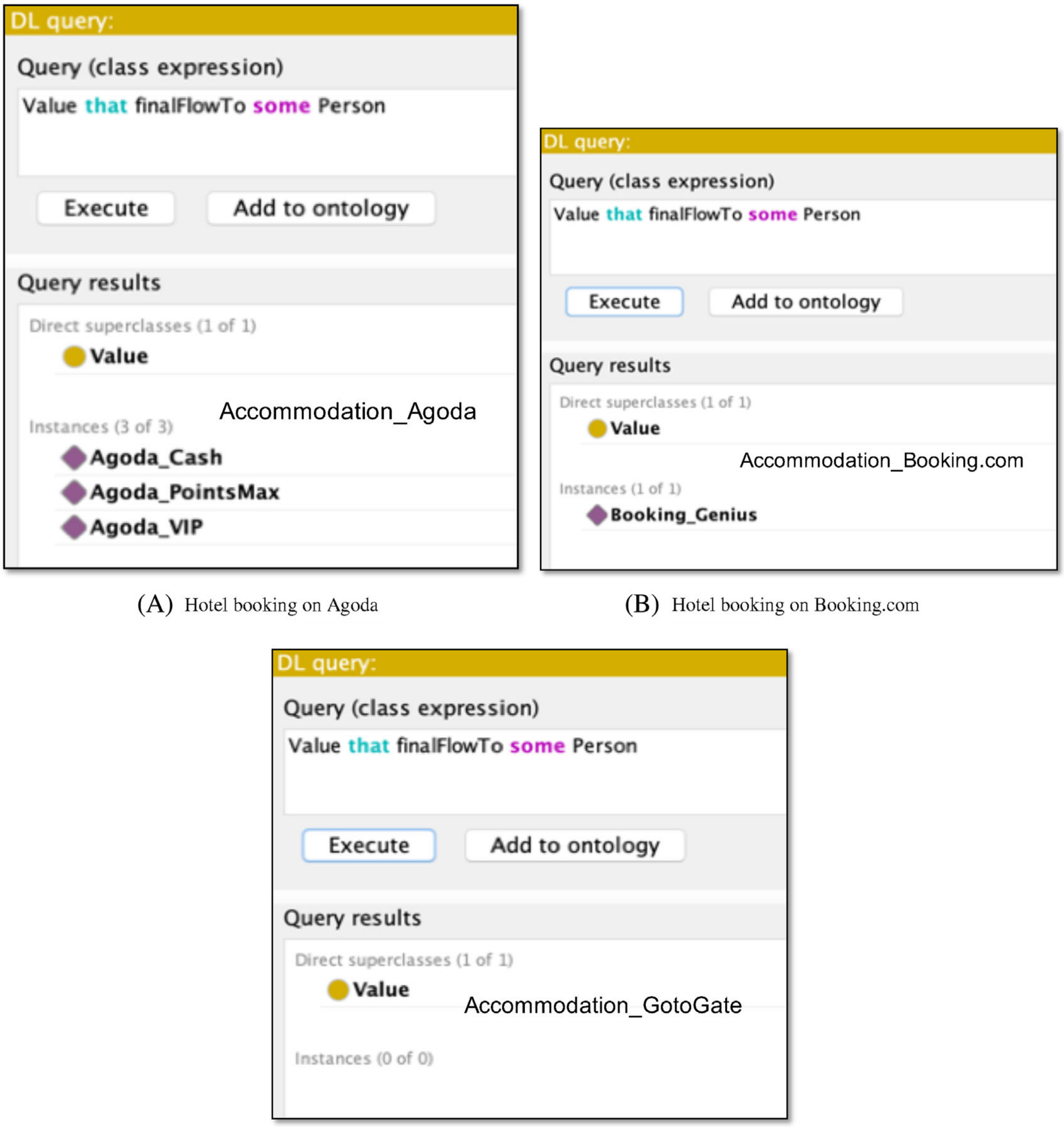
Package 2:
Booking flight on Agoda
Booking hotel on GoToGate

FIGURE 10 Example query on combined data disclosures

7. Person(?p), own(?p, ?d), construct(?d, ?dp), flowTo(?dp, ?s), generate(?s, ?v) → finalFlowTo(?v, ?p).
8. DP(?dp), flowTo(?dp, ?s), create(?s, ?a), friendOf(?a, ?a1) → flowTo(?dp, ?a1).
9. OG(?dp), flowTo(?dp, ?g), hasMember(?g, ?a) → flowTo(?dp, ?a).
10. DP(?dp), flowTo(?dp, ?g), hasMember(?a, ?g) → flowTo(?dp, ?a).
11. DP(?dp), flowTo(?dp, ?a), accountOf(?a, ?p) → finalFlowTo(?dp, ?p).
12. DP(?dp), flowTo(?dp, ?s), generate(?s, ?v), create(?s, ?a) → flowTo(?v, ?a).
13. DP(?dp), flowTo(?dp, ?g), exist(?s, ?g), create(?s, ?a), generate(?s, ?v) → flowTo(?v, ?a).
14. V(?v), flowTo(?v, ?a), accountOf(?a, ?p) → finalFlowTo(?v, ?p).
15. OA(?a), flowTo(Facebook_Followers, ?a), friendOf(?a, ?a1), hasFollowers_fb(?a1, ?n) → hasViewers_fb(?a, ?n).
16. OA(?a), hasFollowers_fb(?a, ?n1), hasViewers_fb(?a, ?n2), greaterThan(?n2, ?n1) → upgrade_fb(?a, true).
17. SP(?sp), isPartOf(?sp, ?sp1), finalFlowTo(?dp, ?sp) → finalFlowTo(?dp, ?sp1).

5 | FUTURE WORK

The proposed model is generic enough to cover a wide range of applications and privacy issues. Due to the limited time and the lack of available data sets, the proposed approach has not been tested against real-world graphs. As a consequence,



Value that finalFlowTo some Person

Direct superclasses (1 of 1)
Value

Accommodation_Agoda

Instances (3 of 3)
Agoda_Cash
Agoda_PointsMax
Agoda_VIP

(A) Hotel booking on Agoda

Value that finalFlowTo some Person

Direct superclasses (1 of 1)
Value

Accommodation_Booking.com

Instances (1 of 1)
Booking_Genius

(B) Hotel booking on Booking.com

Value that finalFlowTo some Person

Direct superclasses (1 of 1)
Value

Accommodation_GotoGate

Instances (0 of 0)

(C) Hotel booking on GoToGate

FIGURE 11 Example query on received values from hotel-booking services

this study does not include performance analysis on the operational efficiency and effectiveness of privacy protection. To extend the work, we identified a number of key areas for further development of the proposed model and its application.

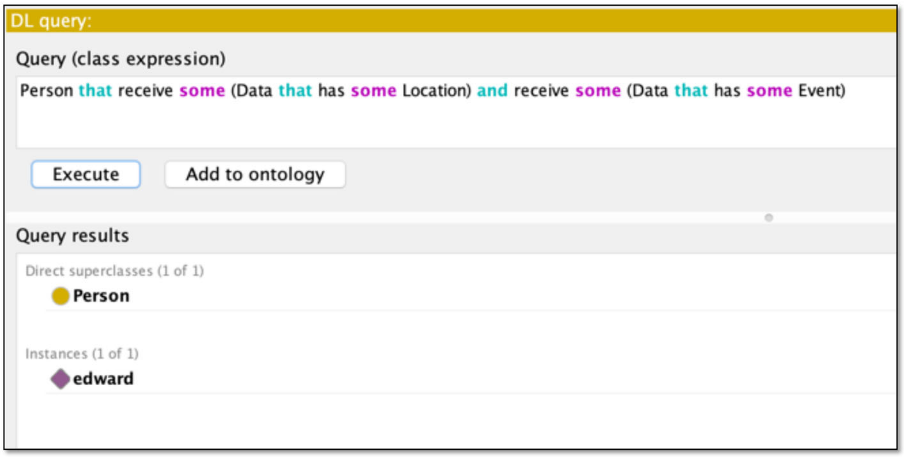
5.1 | Extending the proposed model

The model described in this paper is just the first step and it can be further enhanced in many ways. Below we list some such areas.

More entity types and relations. Our proposed model currently covers 8 entity types and a number of relations between them. There are other entity types we may add, for example, physical groups of people and groups of organizations.

More complicated business models. As mentioned before, the business world is actually very complicated and we have considered only some simple business relations between services and organizations. Therefore, graphs should be built based on more complicated, real-world business models, and related data flows.

More complicated inter-personal relations. Similar to the above, there can be more complicated relationships among people as well. Therefore, current relations to person (P) entities need to be refined to capture more semantic information



Query (class expression)

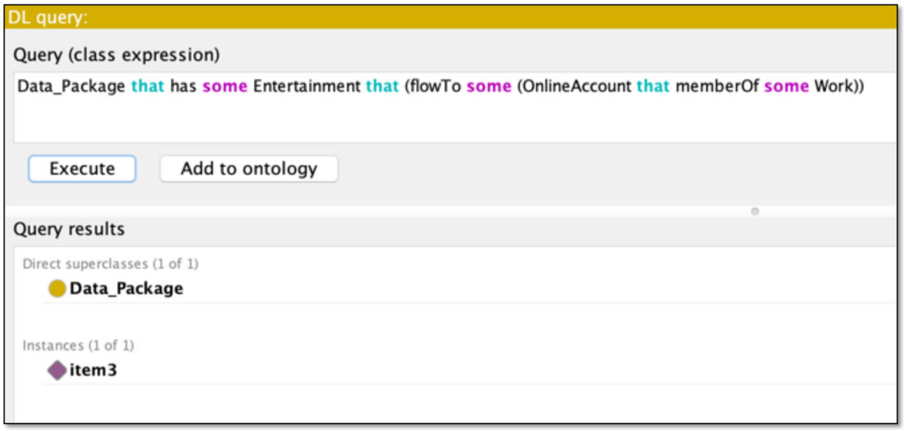
Person that receive some (Data that has some Location) and receive some (Data that has some Event)

Query results

Direct superclasses (1 of 1)
● Person

Instances (1 of 1)
◆ edward

FIGURE 12 Example query on combined data disclosures



Query (class expression)

Data_Package that has some Entertainment that (flowTo some (OnlineAccount that memberOf some Work))

Query results

Direct superclasses (1 of 1)
● Data_Package

Instances (1 of 1)
◆ item3

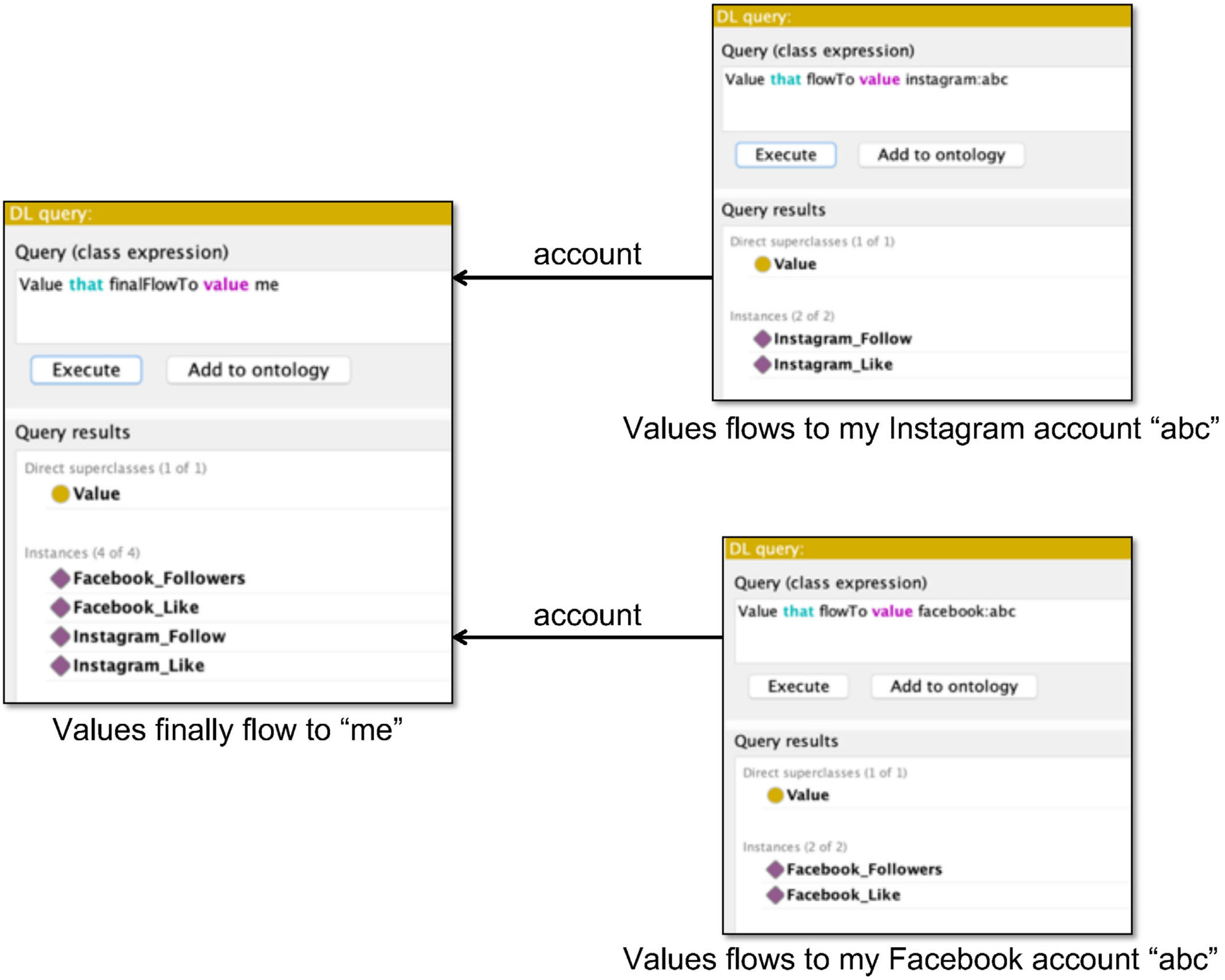
FIGURE 13 Example query on unintended disclosures

from real-world human relations, for example, family, friends, colleagues, carers. According to the semantic relations between P entities, data flows can be differentiated by quantities and thus improve the accuracy of detection results. For instance, to avoid potential privacy issues caused by other people, the central user can exchange recorded data flows with "friends" to see if s/he has overly disclosed data to them.

More complicated data structures for D and DP entities. Our current model abstracts data using Data (D) and Data Package (DP) entity types related with *construct*. In reality, many data entities often include complicated attributes, which may be important for analyzing privacy issues as well. For instance, a travel itinerary contains multiple destinations visited at specific times, transportation types, points of interest, and so forth. Similar issues exist in email, date of birthday, and so forth.

More complicated data structures for V entities. The proposed model currently follows a flat and relatively simple data structure for value (V) entities. In many real world applications, such a flat structure will not work well since there are benefits composed of other more atomic ones (eg, a booking website may offer a package of benefits to their customers for a high-value booking). In addition, our current model assumes a benefit can be more accurately defined and linked with one or more specific data flows, however in reality some benefits are difficult to define but too important to be ignored or simplified, for example, the overall travel experience of a whole trip and the overall happiness gained from positively interacting with other people, groups, services and organizations. Enriching the data structure requires more inter-disciplinary research, as a better understanding on human psychology, economics, business models, and social lives of people.

Re-purposing the model to serve organizations. The proposed model currently focuses more on protecting a target user. The model can however be re-purposed to support organizations in a similar manner. For instance, it can focus more on



Query (class expression)
Value that flowTo value instagram:abc

Query results
Direct superclasses (1 of 1)
Value
Instances (2 of 2)
Instagram_Follow
Instagram_Like

Values flows to my Instagram account "abc"

Query (class expression)
Value that finalFlowTo value me

Query results
Direct superclasses (1 of 1)
Value
Instances (4 of 4)
Facebook_Followers
Facebook_Like
Instagram_Follow
Instagram_Like

Values finally flow to "me"

Query (class expression)
Value that flowTo value facebook:abc

Query results
Direct superclasses (1 of 1)
Value
Instances (2 of 2)
Facebook_Followers
Facebook_Like

Values flows to my Facebook account "abc"

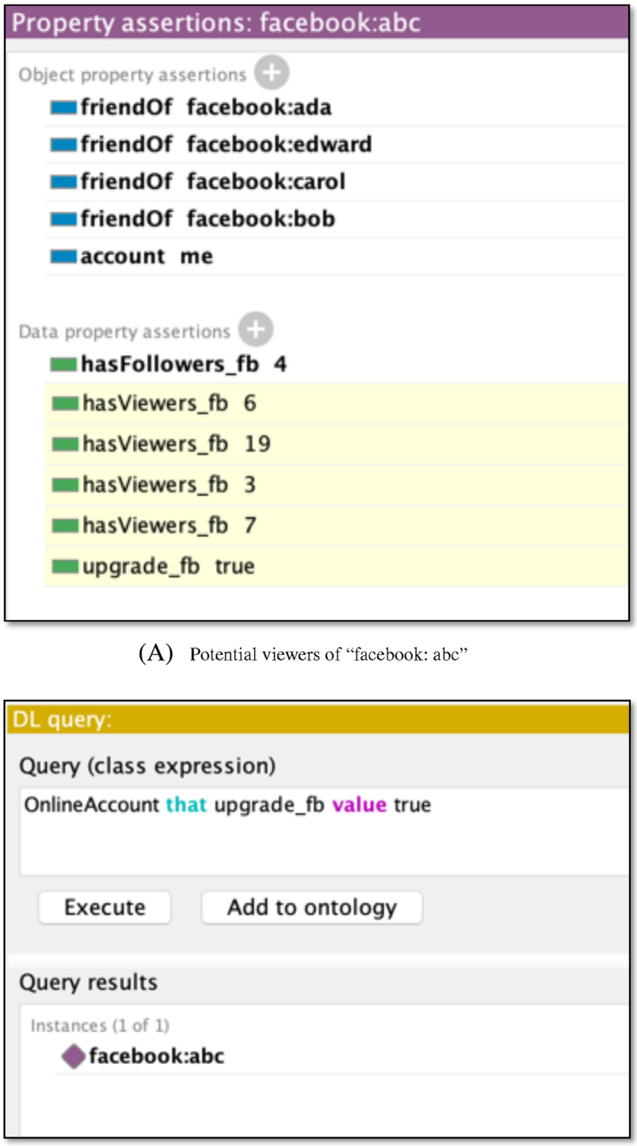
FIGURE 14 Example query regarding value types received from using OSNs

benefits service providers receive from their customers, in order to help such commercial organizations to refine business strategies of running their business, for example, offering customers certain benefits in order to collect more useful data for improving their services.

Introducing negative benefits (ie, harms). Our current model assumes all benefits are added values, that is, they are all positive. In the real world, negative values or harms beyond privacy issues can also be generated from data disclosures. For instance, dislikes on OSNs and downvotes on a P2P system are directly or indirectly caused by data disclosures. Such negative "benefits" can co-exist with positive ones, so there is a different type of trade-offs between these two types of benefits.

Invisible or implicit data flows. This work mainly focuses on data flows caused by visible data sharing, that is, all data flows are explicit and visible to the user concerned. However, it is necessary to monitor invisible or implicit data disclosures that can happen without users' explicit knowledge. For instance, a user's IP address is often disclosed to service providers without a separate explicit notice, which however can be captured as invisible and implicit data flows by extending our model.

Legal framework for data protection and privacy laws. The proposed model can be further enhanced by including a legal framework regarding legality, consequences, and users' rights as data subjects. This can be added as attributes and constraints to data flows and relations. There has been some related work on formalizing such legal frameworks, for example, on the EU GDPR (General Data Protection Regulations).[24]



Property assertions: facebook:abc

Object property assertions

- friendOf facebook:ada
- friendOf facebook:edward
- friendOf facebook:carol
- friendOf facebook:bob
- account me

Data property assertions

- hasFollowers_fb 4
- hasViewers_fb 6
- hasViewers_fb 19
- hasViewers_fb 3
- hasViewers_fb 7
- upgrade_fb true

(A) Potential viewers of "facebook: abc"

DL query:

Query (class expression)

OnlineAccount that upgrade_fb value true

Query results

Instances (1 of 1)

facebook:abc

(B) Query on Level-up of "facebook: abc"

FIGURE 15 Example inference and query about value details

Connecting multiple models together. As a user-centric model, the entity level graph has a special entity "me" at the center of everything. Given a number of users, it is possible to connect their user-centric graphs to form a larger graph showing how privacy issues change from person to person, which will help study larger-scale privacy issues, for example, how privacy issues of one user propagate to his/her friends on OSNs.

5.2 | Developing useful tools and databases

There are also some useful tools that will make it easier to use the proposed model. We give some examples below.

Automatic and dynamic building of the entity level graph. The cyber-physical system (CPS) is not static itself and by nature large scale. Therefore, it is unsuitable to build semantic-based graphs by manual generation. To better manage the ever-changing world and to cover as many relevant information as possible, an automation tool is necessary for building large-scale graphs where data flows are produced all the time. Such dynamically and automatically detected data flows about users' data disclosure activities lay the foundation of privacy risk assessment.

Interactive visualization of data flow paths and privacy issues. As mentioned before, each privacy issue can be represented by a specific topological pattern involving one or more data flow paths. It will be helpful to develop some visualization tools to show such paths and topological patterns, possibly with animation. Such tools can support users to make more informed decisions on their data disclosure activities, with an enhanced level of system transparency.

Automatic comparison of data disclosure options. Given a data flow graph and a number of options for data disclosure, we can automatically check all such options to compare them and determine which options provide better privacy protection. After benefit/value returns are added, such comparison can be done to balance two main objectives: privacy and utility. This particularly will help people achieve a better understanding of privacy protection options in their daily life, particularly among people who are unaware or less aware of privacy issues.

Automatic discovery of OSN accounts that belong to the same organization or individual. More potential privacy issues can be detected if we have more information about the physical entities (organizations or people) behind OSN accounts. Some automatic tools can be developed to detect OSN accounts belonging to the same organization or person to allow a more complete data flow graph related to such accounts, therefore exposing more potential privacy issues.

Building a database of privacy issues. Since different privacy issues can have different topological (and temporal once we add the time dimension in) patterns, it will be interesting to look into more use cases and scenarios contexts to build a database of different privacy issues in different application contexts. Such a database could find applications in training machine learning models for automatic identification of privacy risks.

Building machine learning models to automatic classify privacy issues. With a database of different privacy issues as different topological patterns, it will be possible to train a machine learning model to classify such issues given any privacy-benefit graph. Such a classifier can be used to dynamically detect or predict real-time privacy issues and alert users about necessary actions.

Building a database of business relationships. As shown before the proposed model includes an important part about business relationships among services and organizations. To support a complete coverage of such dynamic business relationships in the real world, it will be very helpful to develop some automatic tools to harvest public data about such relationships and construct a database to support construction of the privacy-benefit graphs. A human and machine readable language will be necessary to allow automatic data storage and manual addition of business relationships that cannot be automatically harvested online for various reasons. Such a database will be useful for other research topics, for example, studying how online services are operated in the dynamically evolving cyber-physical world.

5.3 | Using the proposed model

The proposed model can help support privacy enhancing tools that increase users' awareness of privacy issues and guide them to make more informed decisions by balancing their privacy preferences and expected benefits from data disclosures. As part of an ongoing research project, we are in the process of building a mobile app incorporating the proposed conceptual model as a core component for raising privacy awareness enhancement and for building a privacy-benefit recommender system. We will use behavioral nudging for the privacy recommender system, and the app will be developed to be personalized and independent of any third-party services. The overarching framework we will follow to develop the app can be found in a related paper we published.[2]

6 | RELATED WORK

The most related area is privacy ontologies, which often involve a graph-based model. Most work on this topic mainly focuses on specifying conditions of data access by the controllers. For instance, ontological models can be built to incorporate privacy causes, impacts, and contextual factors. Sacco and Passant (2011)[25] proposed a privacy preference ontology (PPO) to allow users specify fine-grained conditions of using of their RDF data. To effectively combine data (or knowledge) of different sources in the cyber security domain, Iannacone (2015)[26] built a knowledge graph STUCCO with data from 13 structured sources. To ensure privacy criteria of different stakeholders are properly implemented, Kost et al. (2011)[27] integrated an ontology into privacy policy specifications and the evaluation of privacy constraints. Michael et al. (2008)[13] proposed a privacy ontology to support the provision of privacy and derive the privacy levels associated with e-commerce transactions and applications. To guarantee business processes are performed securely, Ioana et al. (2011)[28] designed a semantic annotation tool to assist users in specifying security and privacy constraints onto different business process models. As far as we know, no existing ontologies consider how likely privacy issues are caused from user-centric data flows like what we report in this paper. In Reference 29, Adam et al. (2006) presented a logical framework for expressing and reasoning about norms of transmission of personal information. This formal model covers some central ideas of contextual integrity, a conceptual framework for understanding privacy expectations that has been developed in the literature on law and public policy. Differing from other formal verification models, Kafalŷ et al. (2016)[30] proposed Revani, a solution that incorporates the social dimension and thereby provides a computational basis to regulate interactions among agents. Toward the privacy requirement specification, Gharib et al. (2016)[31] reported that a process to elicit, classify, prioritize and validate privacy requirements for the VisiOn Privacy Platform.§§ As companies could collect and process personal information on their supply chains, it is necessary for stakeholders to follow the privacy and security requirements that cover their practices. Specially, Breaux et al. (2013)[32] derived the methodology from an exploratory case study of the Facebook platform policy and an extended case study using privacy policies from Zynga and AOL Advertising. Many mobile applications collect a significant amount of privacy (sometimes sensitive) data from their users' devices. Although the organization that develops an app has a legal obligation to declare what data are being collected in the app's privacy policy and through permission requests, there are demands about mechanisms for checking the consistency between the privacy policy and the app's actual data collection behavior.[33] To meet such demands, some researchers have proposed solutions. For instance, Slavin et al. (2016)[34] proposed a semi-automated framework that aims at detecting privacy policy violations via code analysis of Android apps. Their evaluation showed that, out of 477 Android apps tested, 341 had potential privacy policy violations.

Reasoning from background knowledge on human relationships, content types, and contextual factors can support decision making on authorization and privacy preservation. Passant et al. (2009)[35] utilized semantic vocabularies such as FOAF (friend of a friend) and SIOC (Semantically Interlinked Online Communities) to establish a trust and privacy layer to restrict publishing, sharing, or browsing data by various social behaviors. By categorizing privacy violations of OSNs as endogenous and exogenous information disclosures in a direct or an indirect way, Kökciyan and Yolum (2016)[36] proposed an agent-based representation based on users' privacy requirements on their generated contents. Considering that limited privacy requirements can be expressed through access control policies, semantic data models have been suggested to assist in authorization to reduce leakage risks.[37] To anonymize e-health records with statistical disclosure control (SDC) methods, Martínez (2013)[38] proposed to incorporate the healthcare terminology SNOMED CT[39] into a privacy ontology to mask categorical attributes and to preserve information utility. To help designers understand security mechanisms and how well they are aligned with corporate missions, Massacci et al. (2018)[40] also considered modeling the ontology around information systems and settings on permission, delegation, and trust at the organizational level.

Another closely related research area is OSN (structural) anonymity. Focusing on OSN data protection, Qian et al. (2017)[41] proposed individual network snapshots. In case sensitive attributes are inferred by attackers, distance between published data and background knowledge needs to be controlled in a safe range. Noticing that anonymized graphs may incur identification attacks, Peng et al. (2014)[42] developed a two-staged algorithm: constructing a sub-graph of users (seed) and connecting to the rest (grow) to show the feasibility. User similarities are shared among "neighbors." As a result, Zhou and Pei (2008)[43] showed that knowing neighbor nodes and attached attributes can increase the probability of identification central users. In addition to static relations, Srivatsa and Hicks (2012)[44] formalized "contact graphs" with contextual factors in mobility. Similarly, Bhagat (2009)[45] argued that graph representations storing user interactions over OSNs should be protected against privacy attacks. Singh and Zhan (2007)[46] analyzed the vulnerability to identity attacks based on topological properties. Instead of modeling network graphs, Li et al. (2016)[47] converted tabular data in data graphs, including original data sets, anonymity data sets and background knowledge of attackers. Instead of direct anonymity on graphs, our goal is to offer users a knowledge graph about data flows to reflect their data disclosure activities in the wider business world (online and offline). Since our approach effectively combines the ontological formalization about data and value flows, graph-based structures of service providers and people as well as a knowledge base with semantic meanings to support automatic reasoning on potential issues individual users care about, we believe that this model can support further development of user-centric privacy-enhancement applications on personal devices, for the purposes such as monitoring data-related activities through different mobile apps.

Studies have shown there are (privacy) costs and benefits while sharing personal data in eCommerce, Web communities, and so forth.[11] For instance, convenience, automation, personalization, and price premiums are generally seen as the benefits for people participating in eCommerce activities.[48-51] Recent studies about innovative business models have shown extra values can be provided. Hamari et al. (2016)[52] added that collaborative consumption can foster sustainable marketplaces, where participants receive economic incomes, satisfactions and enjoyments in business activities. In addition, some research has shown that the online peer-to-peer exchanges can allow better resource allocation and utilization,[53] so that environmental benefits can result from the act of sharing.[54] For instance, from the Q and A forums such as TripAdvisor, tourists seek the information about destinations can obtain the answers from local residents or travelers have visited it in the past.[55] The user-generated reviews on rating systems can help consumers avoiding bad sellers, while boosting the sales of retailers with higher reputations.[56] Reversely, it was also found strong associations exist between consumers' reviews and hotel performance.[57] By sharing media data (photos, texts, music, videos) on OSNs, benefits can be identified from collaboration, relationships, social capitals,[58-60] well-beings and the engagements in offline activities.[61] Previous studies on modeling and joint analysis of such costs and benefits did not conceptualize the associations between benefits and data flows from self-disclosures.[60,62-65] As unexpected data flows can take place while getting benefits, technical solutions will be needed to help users identify and balance the privacy costs and added values. In addition to representing and discovering privacy issues as data-flow graphs, the value component is built within our proposed model to support inferences on both privacy issues and value enhancement. As far as we know, this is the first time that such joint modeling of data and value flows are formally addressed in the context of a privacy ontology.

7 | CONCLUSIONS

In this paper, we propose a user-centric, graph-based semantic model to identify data flows produced from a given user's online and offline activities that can potentially lead to privacy issues. In the conceptual model, privacy issues concerning the given user can be represented as specific topological patterns involving one or more data-flow paths. The model is generic enough to be applied to a wider range of scenarios, some of which were given in this paper to illustrate how it can be used. We also demonstrate that the model can be easily implemented using OWL tools to enable automatic semantic reasoning of privacy issues.

ACKNOWLEDGMENTS

The author is deeply grateful to the anonymous reviewers for their valuable suggestions and comments on the first version of this paper. The authors' work was supported by the research project, PRIvacy-aware personal data management and Value Enhancement for Leisure Travellers (PriVELT, https://privelt.ac.uk/), funded by the EPSRC (Engineering and Physical Sciences Research Council) in the UK, under grant number EP/R033749/1.

ENDNOTES

[*] Names of edges in Figure 1 are not actually part of the conceptual model. They are used for enhancing readability and for informing naming of predicates in Table 1. The dashed edges are numbered to help discuss data and value flows in the rest of the paper.

[†] Terminology wise, both "relation" and "relationship" are used in the research literature. We chose to use the word "relation" because it is the one used in Web Ontology Language (OWL), which we used to implement the automatic reasoning part of the model in Section 4.

[‡] The path is shown as a dotted line in Figure 2 from the source to the destination, ignoring the entities in the middle. The same hereinafter for other figures.

[§] Hotels.com rewards terms and conditions - Terms and Conditions: https://uk.hotels.com/customer_care/terms_conditions.html

[¶] Genius - Booking.com's loyalty programme: https://www.booking.com/genius.html

[**] AgodaCash Rewards Terms and Conditions: https://www.agoda.com/info/agoda-policies.html#10

[††] Agoda introduces AgodaVIP program to boost sales for hotel partners: https://www.agoda.com/press/agoda-introduces-agodavip-program-to-boost-sales-for-hotel-partners?cid=1844104

[‡‡] Agoda PointsMAX: https://www.agoda.com/pointsmax.html

[§§] https://www.visioneuproject.eu/

ORCID

Yang Lu https://orcid.org/0000-0002-0583-2688

REFERENCES

1. Lu Y, Li S. From data flows to privacy issues: a user-centric semantic model for representing and discovering privacy issues. *Proceedings of the 53rd Hawaii International Conference on System Sciences*. USA: University of Hawai'i at Mānoa; 2020:6528-6537. doi:10.24251/HICSS.2020.799
2. Lu Y, Li S, Ioannou A, Tussyadiah I. From data disclosure to privacy nudges: a privacy-aware and user-centric personal data management framework. *Dependability in Sensor, Cloud, and Big Data Systems and Applications: 5th International Conference, DependSys 2019, Guangzhou, China, November 12–15, 2019, Proceedings; Vol. 1123 of Communications in Computer and Information Science*. Singapore: Springer; 2019:262-276. doi:10.1007/978-981-15-1304-6&uscore;21

3. Culnan MJ, Bies RJ. Consumer privacy: balancing economic and justice considerations. *J Soc Issues*. 2003;59(2):323-342. doi:10.1111/1540-4560.00067
4. Li H, Sarathy R, Xu H. Understanding situational online information disclosure as a privacy calculus. *J Comput Inf Syst*. 2010;51(1):62-71. doi:10.1080/08874417.2010.11645450
5. Acquisti A. Privacy in electronic commerce and the economics of immediate gratification. *Proceedings of the 5th ACM Conference on Electronic Commerce*. New York: ACM; 2004:21-29. doi:10.1145/988772.988777
6. Norberg PA, Horne DR, Horne DA. The privacy paradox: personal information disclosure intentions versus behaviors. *J Consum Aff*. 2007;41(1):100-126. doi:10.1111/j.1745-6606.2006.00070.x
7. Bhatia J, Breaux TD. Empirical measurement of perceived privacy risk. *ACM Trans Comput-Hum Interact*. 2018;25(6):34-47. doi:10.1145/3267808
8. Lee H, Lim D, Kim H, Zo H, Ciganek AP. Compensation paradox: the influence of monetary rewards on user behaviour. *Behav Inform Technol*. 2015;34(1):45-56. doi:10.1080/0144929X.2013.805244
9. Krasnova H, Spiekermann S, Koroleva K, Hildebrand T. Online social networks: why we disclose. *J Inf Technol*. 2010;25(2):109-125. doi:10.1057/jit.2010.6
10. Ge J, Peng J, Chen Z. Your privacy information are leaking when you surfing on the social networks: a survey of the degree of online self-disclosure (DOSD). *Proceedings of the 2014 IEEE 13th International Conference on Cognitive Informatics and Cognitive Computing*. Manhattan, New York, U.S.: IEEE; 2014:329-336. doi:10.1109/ICCI-CC.2014.6921479
11. Pötzsch S. Privacy awareness: a means to solve the privacy paradox. *The Future of Identity in the Information Society: 4th IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School, Brno, Czech Republic, September 1–7, 2008, Revised Selected Papers; Vol. 298 of IFIP Advances in Information and Communication Technology*. New York, United States: Springer; 2008:226-236. doi:10.1007/978-3-642-03315-5&uscore;17
12. Wang X, Qin X, Hosseini MB, Slavin R, Breaux TD, Niu J. GUILeak: tracing privacy policy claims on user input data for android applications. *Proceedings of the 40th International Conference on Software Engineering*. New York, NY, United States: ACM; 2018:37-47. doi:10.1145/3180155.3180196
13. Hecker M, Dillon TS, Chang E. Privacy ontology support for E-commerce. *IEEE Internet Comput*. 2008;12(2):54-61. doi:10.1109/MIC.2008.41
14. Hedbom H. A survey on transparency tools for enhancing privacy. *The Future of Identity in the Information Society: 4th IFIP WG 9.2, 9.6/11.6, 11.7/FIDIS International Summer School, Brno, Czech Republic, September 1–7, 2008, Revised Selected Papers; Vol. 298 of IFIP Advances in Information and Communication Technology*. New York, United States: Springer; 2008:67-82. doi:10.1007/978-3-642-03315-5&uscore;5
15. Almuhimedi H, Schaub F, Sadeh N, et al. Your location has been shared 5,398 times! A field study on Mobile app privacy nudging. *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. New York, NY, United States: ACM; 2015:787-796. doi:10.1145/2702123.2702210
16. Hu H, Ahn GJ, Jorgensen J. Multiparty access control for online social networks: model and mechanisms. *IEEE Trans Knowl Data Eng*. 2013;25(7):1614-1627. doi:10.1109/TKDE.2012.97
17. Finin T, Joshi A, Kagal L, et al. R*OWL*BAC - representing role based access control in *OWL*. *Proceedings of the 13th ACM Symposium on Access Control Models and Technologies*. New York, NY, United States: ACM; 2008:73-82. doi:10.1145/1377836.1377849
18. Muhleisen H, Kost M, Freytag JC. SWRL-based access policies for linked data. *Proceedings of the 2nd Workshop on Trust and Privacy on the Social and Semantic Web; CEUR-WS.org*. Vol 1; online: CEUR-WS.org; 2010:12. http://ceur-ws.org/Vol-576/paper1.pdf
19. Lu Y, Sinnott RO. Semantic security for e-health: a case study in enhanced access control. *Proceedings of the 2015 IEEE 12th International Conference on Ubiquitous Intelligence and Computing, the 2015 IEEE 12th International Conference on Autonomic and Trusted Computing and the 2015 IEEE 15th International Conference on Scalable Computing and Communications and its Associated Workshops*. Manhattan, New York, U.S.: IEEE; 2015:407-414. doi:10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.90
20. Lu Y, Sinnott RO. Semantic-based privacy protection of electronic health Records for Collaborative Research. *Proceedings of the 2016 15th IEEE International Conference on Trust, Security and Privacy in Computing and Communications*. Piscataway, NJ, US: IEEE; 2016:519-526. doi:10.1109/TrustCom.2016.0105
21. Sirin E, Parsia B, Grau BC, Kalyanpur A, Katz Y. Pellet: a practical OWL-DL reasoner. *J Web Semant*. 2007;5(2):51-53. doi:10.1016/j.websem.2007.03.004
22. Calvanese D, De Giacomo G, Lembo D, Lenzerini M, Rosati R. DL-lite: tractable description logics for ontologies. *Proceedings of the 20th National Conference on Artificial Intelligence*. Vol 5. Palo Alto, California, U.S.: AAAI; 2005:602-607. http://www.aaai.org/Library/AAAI/2005/aaai05-094.php
23. Horrocks I, Patel-Schneider PF, Boley H, Tabet S, Grosof B, Dean M. SWRL: A semantic web rule language combining OWL and RuleML. W3C Member submission; 2004. http://www.w3.org/Submission/SWRL/.
24. Bartolini C, Robaldo L. PrOnto: privacy ontology for legal reasoning. *Electronic Government and the Information Systems Perspective: 7th International Conference, EGOVIS 2018, Regensburg, Germany, September 3–5, 2018, Proceedings; Vol. 11032 of Lecture Notes in Computer Science*. Germany: Springer; 2018:139-152. doi:10.1007/978-3-319-98349-3&uscore;11
25. Sacco O, Passant A. A privacy preference ontology (PPO) for linked data. *Proceedings of WWW 2011 Workshop on Linked Data on the Web; CEUR-WS.org*. Vol 1; Online: CEUR-WS.org; 2011:5. http://ceur-ws.org/Vol-813/ldow2011-paper01.pdf
26. Iannacone MD, Bohn S, Nakamura G, et al. Developing an ontology for cyber security knowledge graphs. *Proceedings of the 10th Annual Cyber and Information Security Research Conference*. Vol 12. New York, NY, United States: ACM; 2015:12. doi:10.1145/2746266.2746278

27. Kost M, Freytag JC, Kargl F, Kung A. Privacy verification using ontologies. *Proceedings of the 2011 6th International Conference on Availability, Reliability and Security*. 1730 Massachusetts Ave., NW Washington, DC, United States: IEEE Computer Society; 2011:627-632. doi:10.1109/ARES.2011.97
28. Ciuciu I, Zhao G, Mülle J, et al. Semantic support for security-annotated business process models. *Enterprise, Business-Process and Information Systems Modeling: 12th International Conference, BPMDS 2011, and 16th International Conference, EMMSAD 2011, Held at CAiSE 2011, London, UK, June 20–21, 2011. Proceedings; Vol. 81 of Lecture Notes in Business Information Processing*. Germany: Springer; 2011:284-298. doi:10.1007/978-3-642-21759-3&uscore;21
29. Barth A, Datta A, Mitchell JC, Nissenbaum H. Privacy and contextual integrity: framework and applications. *Proceedings of the 2006 IEEE Symposium on Security and Privacy*. Manhattan, New York, U.S.: IEEE; 2006. doi:10.1109/SP.2006.32
30. Kafalỳ Ö, Ajmeri N, Singh MP. Revani: revising and verifying normative specifications for privacy. *IEEE Intell Syst*. 2016;31(5):8-15. doi:10.1109/MIS.2016.89
31. Gharib M, Salnitri M, Paja E, et al. Privacy requirements: findings and lessons learned in developing a privacy platform. *Proceedings of the 2016 IEEE 24th International Requirements Engineering Conference*. Manhattan, New York, U.S.: IEEE; 2016:256-265. doi:10.1109/RE.2016.13
32. Breaux TD, Hibshi H, Rao A. Eddy, a formal language for specifying and analyzing data flow specifications for conflicting privacy requirements. *Requir Eng*. 2014;19(3):281-307. doi:10.1007/s00766-013-0190-7
33. Zhang X, Wang X, Slavin R, Breaux T, Niu J. How does misconfiguration of analytic services compromise mobile privacy? *Proceedings of the 2020 IEEE/ACM 42nd International Conference on Software Engineering*. New York, NY, United States: ACM; 2020:1572-1583. doi:10.1145/3377811.3380401
34. Slavin R, Wang X, Hosseini MB, et al. Toward a framework for detecting privacy policy violations in android application code. *Proceedings of the 38th International Conference on Software Engineering*. New York, NY, United States: ACM; 2016:25-36. doi:10.1145/2884781.2884855
35. Passant A, Kärger P, Hausenblas M, Olmedilla D, Polleres A, Decker S. Enabling trust and privacy on the social web. *Proceedings of the 2009 W3C Workshop on the Future of Social Networking*. Online: W3C; 2009:15-16. https://www.w3.org/2008/09/msnws/papers/trustprivacy.html
36. Kökciyan N, Yolum P. PriGuard: a semantic approach to detect privacy violations in online social networks. *IEEE Trans Knowl Data Eng*. 2016;28(10):2724-2737. doi:10.1109/TKDE.2016.2583425
37. Paci F, Zannone N. Preventing information inference in access control. *Proceedings of the 20th ACM Symposium on Access Control Models and Technologies*. New York, NY, United States: ACM; 2015:87-97. doi:10.1145/2752952.2752971
38. Martínez S, Sánchez D, Valls A. A semantic framework to protect the privacy of electronic health records with nonnumerical attributes. *J Biomed Inform*. 2013;46(2):294-303. doi:10.1016/j.jbi.2012.11.005
39. SNOMED International. SNOMED-5-Step Briefing. Web page; 2020. http://www.snomed.org/snomed-ct/five-step-briefing.
40. Massacci F, Mylopoulos J, Zannone N. An ontology for secure socio-technical systems. *Handbook of Ontologies for Business Interaction*. Pennsylvania, United States: IGI Global; 2008:188-206. doi:10.4018/978-1-59904-660-0.ch011
41. Qian J, Li XY, Zhang C, Chen L, Jung T, Han J. Social network De-anonymization and privacy inference with knowledge graph model. *IEEE Trans Dependable Secure Comput*. 2017;16(4):679-692. doi:10.1109/TDSC.2017.2697854
42. Peng W, Li F, Zou X, Wu J. A two-stage deanonymization attack against anonymized social networks. *IEEE Trans Comput*. 2014;63(2):290-303. doi:10.1109/TC.2012.202
43. Zhou B, Pei J. Preserving privacy in social networks against neighborhood attacks. *Proceedings of the 2008 IEEE 24th International Conference on Data Engineering*. Manhattan, New York, U.S.: IEEE; 2008:506-515. doi:10.1109/ICDE.2008.4497459
44. Srivatsa M, Hicks M. Deanonymizing mobility traces: using social network as a Side-Channel. *Proceedings of the 2012 ACM Conference on Computer and Communications Security*. New York, NY, United States: ACM; 2012:628-637. doi:10.1145/2382196.2382262
45. Bhagat S, Cormode G, Krishnamurthy B, Srivastava D. Class-based graph anonymization for social network data. *Proc VLDB Endow*. 2009;2:766-777. https://www.vldb.org/pvldb/vol2/vldb09-pvldb26.pdf
46. Singh L, Zhan J. Measuring topological anonymity in social networks. *Proceedings of the 2007 IEEE International Conference on Granular Computing*. Manhattan, New York, U.S.: IEEE; 2007:770-774. doi:10.1109/GrC.2007.31
47. Li XY, Zhang C, Jung T, Qian J, Chen L. Graph-based privacy-preserving data publication. *Proceedings of the 2016 35th Annual IEEE International Conference on Computer Communications*. Manhattan, New York, U.S.: IEEE; 2016. doi:10.1109/INFOCOM.2016.7524584
48. Murae Y, Ho BQ, Hara T, Okada Y. Two aspects of customer participation behaviors and the different effects in service delivery: evidence from home delivery services. *JMDC*. 2019;13(1):45-58. doi:10.33423/jmdc.v13i1.681
49. Zeithaml VA, Parasuraman A, Malhotra A. A conceptual framework for understanding e-service quality: implications for future research and managerial practice. Working paper MSI WP 00-115. Marketing Science Institute. 2000. https://www.msi.org/wp-content/uploads/2020/06/MSI_WP_00-115.pdf
50. Kaynama SA, Black CI. A proposal to assess the service quality of online travel agencies: an exploratory study. *J Prof Serv Mark*. 2000;21(1):63-88. doi:10.1300/J090v21n01&uscore;05
51. Yang Z, Peterson RT, Huang L. Taking the pulse of internet pharmacies. *Mark Health Serv*. 2001;21(2):4-10.
52. Hamari J, Sjöklint M, Ukkonen A. The sharing economy: why people participate in collaborative consumption. *J Assoc Inf Sci Technol*. 2016;67(9):2047-2059. doi:10.1002/asi.23552
53. Quattrone G, Proserpio D, Quercia D, Capra L, Musolesi M. Who benefits from the "sharing" economy of Airbnb? *Proceedings of the 25th International Conference on World Wide Web*. Republic and Canton of Geneva, Switzerland: ACM; 2016:1385-1394. doi:10.1145/2872427.2874815

54. Böcker L, Meelen T. Sharing for people, planet or profit? Analysing motivations for intended sharing economy participation. *Environ Innov Soc Trans*. 2017;23:28-39. doi:10.1016/j.eist.2016.09.004
55. Gal-Tzur A, Rechavi A, Beimel D, Freund S. An improved methodology for extracting information required for transportrelated decisions from Q&A forums: a case study of TripAdvisor. *Travel Behav Soc*. 2018;10:1-9. doi:10.1016/j.tbs.2017.08.001
56. Resnick P, Zeckhauser R, Swanson J, Lockwood K. The value of reputation on eBay: a controlled experiment. *Exp Econ*. 2006;9(2):79-101. doi:10.1007/s10683-006-4309-2
57. Xie KL, Zhang Z, Zhang Z. The business value of online consumer reviews and management response to hotel performance. *Int J Hosp Manag*. 2014;43:1-12. doi:10.1016/j.ijhm.2014.07.007
58. Beldad AD, Hegner SM. More photos from me to thee: factors influencing the intention to continue sharing personal photos on an online social networking (OSN) site among young adults in The Netherlands. *Int J Hum Comput Interact*. 2017;33(5):410-422. doi:10.1080/10447318.2016.1254890
59. Krasnova H, Veltri NF, Günther O. Self-disclosure and privacy calculus on social networking sites: the role of culture. *Bus Inf Syst Eng*. 2012;4(3):127-135. doi:10.1007/s12599-012-0216-6
60. Heravi A, Mubarak S, Choo KKR. Information privacy in online social networks: uses and gratification perspective. *Comput Hum Behav*. 2018;84:441-459. doi:10.1016/j.chb.2018.03.016
61. Pendry LF, Salvatore J. Individual and social benefits of online discussion forums. *Comput Hum Behav*. 2015;50:211-220. doi:10.1016/j.chb.2015.03.067
62. De SJ, Imine A. Enabling users to balance social benefit and privacy in online social networks. *Proceedings of the 2018 16th Annual Conference on Privacy, Security and Trust*. Manhattan, New York, U.S.: IEEE; 2018. doi:10.1109/PST.2018.8514202
63. De SJ, Imine A. To reveal or not to reveal: balancing user-centric social benefit and privacy in online social networks. *Proceedings of the 33rd Annual ACM Symposium on Applied Computing*. New York, NY, United States: ACM; 2018:1157-1164. doi:10.1145/3167132.3167258
64. Schweitzer F, Mavrodiev P, Seufert AM, Garcia D. Modeling user reputation in online social networks: the role of costs, benefits, and reciprocity. *Entropy*. 2020;22(10):1073. doi:10.3390/e22101073
65. Awad NF, Krishnan MS. The personalization privacy paradox: an empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Q*. 2006;30(1):13-28. doi:10.2307/25148715

How to cite this article: Lu Y, Li S. From data flows to privacy-benefit trade-offs: A user-centric semantic model. *Security and Privacy*. 2022;5(4):e225. doi: 10.1002/spy2.225