Towards Cost Effective Privacy Provision for Typed Resources in IoT Environment

Yucong Duan*[1,2], Zhengyang Song[1,2], Xiaoxian Yang[3], Quan Zou[4], Xiaobing Sun[5], Xinyue Zhang[1,2]
[1,2]State Key Laboratory of Marine Resource Utilization in the South China Sea,College of Information Science and Technology,
Hainan University, 570228 Haikou, China
[3]School of Computer and Information Engineering, Shanghai Polytechnic University, 201209 Shanghai, P.R. China
[4]College of Computer Science, Tianjin University, China
[5]School of Information Engineering, Yangzhou University, China
Email: duanyucong@hotmail.com, 1464626602@qq.com, xxyang@sspu.edu.cn, zouquan@tju.edu.com,
sundomore@163.com , yuexinaai@163.com

***Abstract*—We present privacy resources in IoT as data, information, and knowledge. We construct a privacy protection architecture on our previously proposed DIKW graphs: Data Graph, Information Graph, and Knowledge Graph. On this architecture, we search privacy protection target resources both as they appear explicitly in their original types and as they appear implicitly which means that they are expressed not in their original types. For a single privacy protection target, it may have various concrete compositions in various layers of DIKW Graph. It becomes more complex since the implementation of a privacy target might also be intertwined with the implementation of other privacy targets. We propose to protect target resources according to their types by either isolating the elements comprising an implementation, or weakening relationships among elements comprising an implement.To optimize among several choices of implementing a protection in a business environment, we introduced the tradeoff between customers' expectations/investment and privacy providers' expectation. Thereafter we proposed to prioritize implementation according to their ratio of cost/benefit.**

Keywords—Internet of Things; typed resources; privacy provision; Knowledge Graph

I. Introduction

We classify privacy resources into data privacy, information privacy, and knowledge privacy. Chaim [1] illustrated the concepts of defining data, information and knowledge. Duan et al. [3] clarified the architecture of Knowledge Graph in terms of data, information, knowledge and wisdom. In [4], the authors proposed to designate the form of Knowledge Graph as four basic forms including Data Graph, Information Graph, Knowledge Graph and Wisdom Graph. We have identified there are enormous potentials of security protection according to the difference of resource types in an investment driven manner [2]. We propose to protect privacy resources in a three-tier architecture consisting of Data Graph, Information Graph and Knowledge Graph. For a single privacy protection target, it may have various concrete compositions in various layers of DIKW Graph. For example, a piece of data might exist as a piece of data or a set of data in Data Graph explicitly, or it might take the implicit form of being expressed as a series of relationships in Information Graph. It becomes more complex since the implementation of a privacy target might also be intertwined with the implementation of other privacy targets.

In [5], the authors proposed a cost effective approach to satisfy performance requirements while minimizing dynamic power consumption. In [6], cost-effective data sharing with forward security improved efficiency is provided through reducing the computation and communication cost. In [7], the authors proposed to provide the I/O resources for specific workloads, while minimizing the total operating cost. We elaborate towards cost effective [8] information privacy provision approach on the basis of Data Graph, Information Graph, and Knowledge Graph.

The rest of this paper is organized as follows. Section II defines typed resources and presents a privacy protection architecture. Section III shows a running example. Section IV presents implementation procedures. Section V shows a simulation. We conclude in Section VI.

II. Privacy protection Architecture

A. *Definitions of Typed Resources*

We have reconstructed a DIKW system for modeling and implementation of resource identification and management[2], [3]. Data is not specified for a specific stakeholder or a machine. Data represents directly observed objects as isolation which only contains the shared common meaning of their necessary identifications. Information represents data or information which are observed or interacted directly or indirectly by human. Knowledge represents the abstracted data, information and knowledge which are taken in a limited or unlimited complete manner as a whole. Knowledge here can be roughly mapped to cover what Kant called Categories[9]. Knowledge is exploited to reason and predict unknown resources or not observed but happened relationships in terms of Data, Information and Knowledge. We are actually building "schemas"[9] for DIKW resources for privacy modeling and provision subsequently.

Definition 1. Typed resources. We define typed resources as a triad:

$$TR_{DIK}: = < D_{DIK}, I_{DIK}, K_{DIK} >$$

D represents Data, I represents Information and K represents Knowledge for convenient description. D_{DIK} is transformed to I_{DIK} through taking roles in real or imaginary scenarios by connecting to other D_{DIK} or I_{DIK} in term of time [9] or order. The associated D_{DIK} corresponds to I_{DIK}.

Definition 2. DIKWGraph. We specify the usually used concept of Knowledge Graph in three layers of Data Graph (DG_{DIK}), Information Graph(IG_{DIK}), and Knowledge Graph(KG_{DIK}) [3].

DOI reference number: 10.18293/SEKE2018-115

DIKWGraph: = < (DG_{DIK}), (IG_{DIK}), (KG_{DIK})>.

Definition 3. DG_{DIK}.

DG_{DIK}: = collection {array, list, stack, queue, tree, graph}.

DG_{DIK} is a collection of discrete elements expressed in the form of various data structures including arrays, lists, stacks, trees, graphs, and so on. DG_{DIK} can record basic structures of entities. Also, DG_{DIK} can record spacial and topological relationships with frequencies.

Definition 4. IG_{DIK}.

IG_{DIK}: = $\text{composition}_{\text{time}}$ { D_{DIK}}.

IG_{DIK} comprises of temporal relationships based on D_{DIK} with specific senarios. IG_{DIK} expresses the interaction and transformation of I_{DIK} between entities in the form of a directed graph. IG_{DIK} can record the interactions between entities including direct interaction and indirect interaction.

Definition 5. KG_{DIK}.

KG_{DIK}: = $\text{collection}_{\text{consistent}}\{\text{Rules}_{\text{Statistic OR Logical}}\}_{\text{category}}$.

KG_{DIK} consistently accommodates either empirical statistical experiences expressed in terms of categories which represent the underlying elements as a whole or completely.

B. Schemas for using DIKW Graphs

To utilize the graphs in DIKW Graphs, we need to mediate the bidirectional feasible transformations of resources among different types of Data, Information and Knowledge. By restricting the transformation with feasible, we mean that not all bidirectional transformations are deemed to be meaningful and practical.

Schemas[9] are proposed by Kant to cognitively mediate the cognitive objects/experiences mostly through logical reason and concretization in time dimension. We borrowed this term here for specifying the transformation among resources with a focus on the type level implementation.

Schema "Data-Resource(Data, Information)": Data are observed by observers from outside world or from inside categorization on a set of resources, structured or not, which are given the conceptual unity as an entity, or on abstraction of information expressions which are exposed as temporal association among elements. Since resource elements can be abstracted upward or decomposed downward, the expressions of specific **DG_{DIK}** and **IG_{DIK}** are therefore intertwined based on the overlapping of the elements and their relationships. We propose to justify and predict the semantic meaning and semantic associations corresponding to resource element expressions based on the reasoning and calculation in a bottom up manner from composing elements of **DG_{DIK}** and **IG_{DIK}**.

Schema "Knowledge-Resource(Data, Information, knowledge)": Knowledge here is either based on the probabilistic experience or based on reason on categories abstracted from directly observed resources or indirectly reasoned resources. A shared characteristic of both forms of knowledge is that they both demand a semantic identification of completeness regardless of whether the actual target resources which are the basis of conceptualization of related categories are limited or unlimited. The schema to enact knowledge on resources is either through temporally decomposing the content of the comprising categories in the knowledge expression as elements shared or can be related to elements in target resources, or through logical or probability reasoning first and decomposing and relating subsequently.

For construction of " Wisdom " related schemas, we adopt the intuition from Schopenhauer[10] to take wisdom as the balancing between reasoning and will for optimizing human long run goals. We omit the discussion on the schema of wisdom here.

C. Privacy Protection Architecture

I_{DIK} expresses interaction and collaboration between entities. Through classifying and abstracting interactive records or behavior records related to the dynamic behavior of entities, we obtain K_{DIK} in the form of statistical rules. We infer K_{DIK} from known resources and collect necessary I_{DIK} in the process of inference through appropriate techniques such as experiments, surveys, and so on. Transformation from I_{DIK} to D_{DIK} takes place either as a conceptualization process from relationships to an entity or as an abstraction which selectively maps the involved elements comprising the I_{DIK} to elements in a structure of a target elements of D_{DIK}. We obtain D_{DIK} through observing an object at a certain time in a static state. K_{DIK} elements are either associated with underlying fine grained instances within their categories such as sub-attributes or sub-operations, or connected through pure logical reason or mathematical computation. The top-down influence from K_{DIK} to I_{DIK} and D_{DIK} is realized through creatively decomposing of the content of K_{DIK} to I_{DIK} and D_{DIK} temporally.

Algorithm 1. Process of Privacy Protection Architecture

```
Input: Target Privacy Resources
Output: Final Elements
  SWITCH ( all target privacy resources)
    CASE 1: Data resources
        IF (they are explicit)
            Isolate or transform them;
        ELSE IF (they are implicit)
            Find out them on DIKW Graphs;
            Isolate or transform them;
    CASE2: Information resources
        IF (they are explicit)
            Isolate or transform them;
        ELSE IF (they are implicit)
            Find out them on DIKW Graphs;
            Isolate or transform them;
Return final elements;
```

Our process of privacy protection architecture is shown as Algorithm 1. When we protect target Data privacy resources, we search for them on the three-layer graphs firstly. There are two senarios. One is that these resources are explicit, we can succeed in searching them directly. The other is that we fail to search, then we analyze associated relationships between target Data privacy resources and other relative typed resources on DG_{DIK}, IG_{DIK}, and KG_{DIK}. We infer them through three associated relationships as associated Data resources infer target Data resources, associated Information resources infer target Data resources and associated Data and Information resources infer target Data resources.

When we protect target Information privacy resources, we find them on the three-layer graphs firstly. There are also two senarios the same as finding Data privacy resources. If these Information resources are explicit, we isolate them or transform them directly. If they are not, we find out them on DIKW Graphs. At last, we isolate these resources, or transform them into other typed resources and store these final elements into a security space, in which the elements will not be used, tampered with, lost and destroyed in unauthorized situations.

III. Running Example

We design a campus monitoring system shown in Figure 1, which consists of geographic location acquisition module, credit card consumption tracking module, video acquisition module, and resource analysis and processing module.

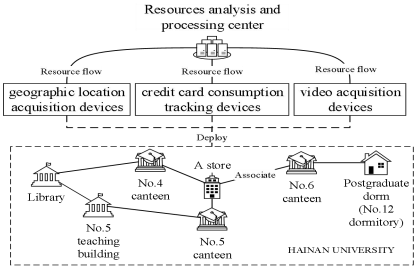
Resources analysis and processing center
Resource flow
Resource flow
Resource flow
geographic location acquisition devices
credit card consumption tracking devices
video acquisition devices
Deploy
Library
No.4 canteen
A store
Associate
No.6 canteen
Postgraduate dorm (No.12 dormitory)
No.5 teaching building
No.5 canteen
HAINAN UNIVERSITY

Figure 1. Campus monitoring system

We collected resources and construct DG_{DIK}, IG_{DIK}, and KG_{DIK}. When target privacy resources are implicit, there is a way that we find out these resources by analyzing associated resources in forms of specific I_{DIK}, such as linked and aggregative structure I_{DIK}. LSI denotes linked structure I_{DIK}. ASI denotes aggregative structure I_{DIK}.

Given an IG_{DIK} denoted as a directed graph G= (V, E), where V is a set of nodes, E is a set of edges connecting those nodes. VoN denotes the value of a node which has two attributes of the number and direction of edge connected to this node. We use t (f_1 (direction (N))*f_2 (number (N)) to evaluate VoN.

A. Protection strategies of LSI

1) Protection of linked I_{DIK} without branched structure in LSI (IUP). We present two strategies to protect IUP privacy which are illustrated as follows.

Binary Damage Method: We search the middle node of each IUP. Second, we continue to search two middle nodes of two IUP parts divided by previous middle node, respectively. And so on, we find out m nodes need protection.

Middle Centralized Damage Method: We search nodes of the middle location in IUP. In sequence, we find out m nodes need protection.

After searching, we isolate or transform these m nodes and store them into security space.

2) Protection of linked I_{DIK} with branched structure in LSI(IBP). We rank nodes of this link according to VoN. We find out n nodes in sequence. Then we isolate or transform those n nodes and store them into a security space.

B. Protection strategies of ASI

1) Protection of aggregative nodes with equal value (VEG). We find out h nodes according to depth-first algorithm or breadth-first algorithm, after isolating or transforming these I_{DIK} nodes, we store them in a security space.

2) Protection of aggregation nodes with unequal node value (VUG). Similar to protection of IBP, we determine the order of target privacy resources according to the rank of VoN. For example, in Figure 1, after analyzing, we know that a student uses his credit card in No.6 canteen frequently. But he rarely uses his credit card in No.4 canteen which is popular to students. Therefore we infer this student lives near No.6 canteen. Since No.6 canteen is near postgraduate dorms, we can infer this student is a postgraduate. Our proposed mechanism isolates or transforms these collected resources of No.6 dormitory according to the rank of VoN and store them into a given secure space.

IV. Framework of value driven Privacy Provision

To optimize among several choices of implementing a protection in a business environment, we introduced the tradeoff between customers' expectations/investment and privacy providers' expectation.

Definition 9. We denote target privacy resource collection as TPR. TRPC represents associated privacy resources which is the target resource processing collection. We define TRPC as a tuple:

TRPC: = < LSIC, ASIC >

LSIC is the set of LSI privacy resources and ASIC is the set of ASI privacy resources. LSIC= {IUPC, IBPC}, where IUPC is the set of IUP privacy resources, and IBPC is the set of IBP privacy resources. ASIC= {VEGC, VUGC}, where VEGC is the set of VEG privacy resources, and VUGC is the set of VUG privacy resources.

Definition 10. Security Space. We define security space denoted with SS as a tuple:

SS: = < SST, SSS >

SST is the type set of graph resources denoted with a triad < sst_D, sst_I, sst_K >. SSS is the scale of different kinds of graph resources represented by a triad < sss_D, sss_I, sss_K >. Each sss denotes the scale of resource in the form of sst.

A. Calculation of User Investment

*1) Cost of damaging nodes:*We assign that 1C is atomic cost of damaging each node in TRPC. In fact, it is necessary to analyze importance of relationship between nodes. But at present we only consider the number of nodes when computing cost. Cost of damaging nodes can be illustrated as

$$DeCost = (m + n + h +k)* 1C \qquad (1)$$

2) Cost of transforming TRPC into SS : SS is a security space. We convert resource types to optimize storage and computation, which makes it hard for unauthorized users to access protected resources. As shown in Table I, PUnitCost represents atomic cost of transforming unit resource into SS.

Cost of transforming resources is illustrated as

$$TrCost = \sum_{i \in \{D, I, K\}} SUnitCosti * sss_i + sssi' \qquad (2)$$

TABLE I. ATOMIC COST OF CONVERTING UNIT RESOURCE IN SS

	D_{DIK}	I_{DIK}	K_{DIK}
D_{DIK}	$SUnitCost_{D\text{-}D}$	$SUnitCost_{D\text{-}I}$	$SUnitCost_{I\text{-}K}$
I_{DIK}	$SUnitCost_{I\text{-}D}$	$SUnitCost_{I\text{-}I}$	$SUnitCost_{I\text{-}K}$
K_{DIK}	$SUnitCost_{K\text{-}D}$	$SUnitCost_{K\text{-}I}$	$SUnitCost_{K\text{-}K}$

3) *User investment computaion*

Costs of providing protection services for private resources consist of two parts as damaging nodes and transforming nodes and store them into SS. We calculate total cost of protecting target privacy, which is illustrated as

$$TotalCost = DeCost + TrCost \qquad (3)$$

μ denotes the unit investment of TotalCost that obtained through data training. Corresponding user investment can be illustrated as

$$UserCost = \mu * TotalCost \qquad (4)$$

B. *Privacy Level Computation*

Privacy level reflects the ability of protection service. The smaller privacy level is, the better protection ability the service has. We denote the privacy level as PL. M_m represents the total number of nodes in *m*th link. Privacy level of IUP I_{DIK} denoted with LPL_{UP} is illustrated as:

$$LPL_{UP} = m / Mm \qquad (5)$$

N_n represents the total number of nodes in *n*th link. α represents an adjusted parameter that is obtained through data mining. V_i represents node V_i that belongs to *n*th link. Privacy level of IBP I_{DIK} denoted with LPL_{BP} is illustrated as:

$$LPL_{BP} = (\sum_{i \le n} \alpha * VoN(V_i)) / Nn \qquad (6)$$

H_h denotes the total number of nodes in *h*th aggregation. Privacy level of VEG I_{DIK} denoted with APL_{EG} is illustrated as:

$$APL_{EG} = h / Hh \qquad (7)$$

K_k denotes the total number of nodes in *k*th aggregation. β represents an adjusted parameter that is obtained through data mining. V_i represents node V_i that belongs to *k*th link. Privacy level of VUG I_{DIK} denoted with APL_{UG} is illustrated as

$$APL_{UG} = (\sum_{i \le k} \beta * VoN(V_i)) / Kk \qquad (8)$$

Algorithm 2 describes a procedure of calculating PL according to user investments.

```
Algorithm 2. Calculating PL
Input: TRPC, SS, $UserCost_0$, $PL_0$
Output: The maximum $PL_0$
  FOR all TRPC DO
    Calculate DeCost;   Calculate TrCost;
    Calculate PL;       Calculate UserCost;
    IF ($UserCost < UserCost_0$ & $PL > PL_0$)
      $PL_0 = PL$;
    ELSE IF ($m \le M_m$ OR $n \le N_n$ OR $k \le K_k$)
      Next step;
Return $PL_0$;
```

V. SIMULATION

Following the scenario in section III, we set target privacy collection (TPR) and constructed DG_{DIK}, IG_{DIK}, and KG_{DIK} according to collected resources. We considered a constructed IG_{DIK} denoted with a directed graph G = (V, E). We assigned that expectant investment is 50 units and expectant privacy level is 0.6. For convenience, we assumed that there are same resource collections in all considered 18 nodes, which is {I, D, I}. Meanwhile, we assigned that sst_i = {D, I, K}, sss_i = {4, 2, 1}, μ = 0.4, 1C = 2 units. Figure 2 shows that the greater the number of nodes which need protection is, the better PL of each I_{DIK} structure performs.

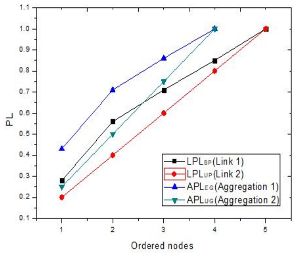

Figure 2. PL values of LSI and ASI.

VI. CONCLUSION

According to the types of various resources, we provide protection solutions based their state of being explicitly or implicitly expressed. During the implementation, we take into consideration of the tradeoff between the value expectation of customers and the value expectation of privacy providers to prioritize privacy targets and corresponding levels of protection.

ACKNOWLEDGMENT

We acknowledge Hainan Project No.ZDYF2017128, NSFC under Grant (No.61662021 and No. 61502294). *refers to corresponding author.

REFERENCES

[1] C. Zins, “Conceptual approaches for defining data, information, and knowledge,” JASIST, vol. 58, no. 4, pp. 479–493, 2007.

[2] L. Shao, Y. Duan, L. Cui, Q. Zou, and X. Sun, “A pay as you use resource security provision approach based on data graph, information graph and knowledge graph,” IDEAL 2017, pp. 444–451.

[3] Y. Duan, L. Shao, G. Hu, Z. Zhou, Q. Zou, and Z. Lin, “Specifying architecture of knowledge graph with data graph, information graph, knowledge graph and wisdom graph,” SERA 2017, pp. 327–332.

[4] L. Shao, Y. Duan, X. Sun, Q. Zou, R. Jing, and J. Lin, “Bidirectional value driven design between economical planning and technical implementation based on data graph, information graph and knowledge graph”, SERA 2017, pp. 339–344.

[5] J. Guerra, H. Pucha, J. S. Glider, W. Belluomini, and R. Rangaswami, “Cost effective storage using extent based dynamic tiering,” 9th USENIX Conference on File and Storage Technologies, 2011, pp. 273–286.

[6] X. Huang, J. K. Liu, S. Tang, Y. Xiang, K. Liang, L. Xu, and J. Zhou, “Cost-effective authentic and anonymous data sharing with forward security,” IEEE Trans. Computers, vol. 64, no. 4, pp. 971–983, 2015.

[7] N. Zhang, J. Tatemura, J. M. Patel, and H. Hacig¨um¨us, “Towards cost-effective storage provisioning for dbmss,” CoRR, vol. abs/1201.0226, 2012.

[8] M. A. Jabbar, G. S. Bopche, B. L. Deekshatulu, and B. M. Mehtre, “Diversity-aware, cost-effective network security hardening using attack graph,” SSCC 2017, pp. 1–15.

[9] Kant I. Critique of pure reason[M]. Cambridge University Press, 1998.

[10] Schopenhauer, The World as Will and Representation, Vol. I, Appendix, "Criticism of the Kantian Philosophy," p. 449 f.