Privacy Disclosure in the Real World:
An Experimental Study

Siyu Wang, Beijing University of Technology, China
Nafei Zhu, Beijing University of Technology, China
Jingsha He, Beijing University of Technology, USA
Da Teng, Beijing University of Technology, China
Yue Yang, Beijing University of Technology, China

ABSTRACT

Privacy protection is a hot topic in network security. Many scholars are committed to evaluating privacy information disclosure by quantifying privacy, thereby protecting privacy and preventing telecommunications fraud. However, in the process of quantitative privacy, few people consider the reasoning relationship between privacy information, which leads to the underestimation of privacy disclosure and privacy disclosure caused by malicious reasoning. This paper completes an experiment on privacy information disclosure in the real world based on WordNet ontology. According to a privacy measurement algorithm, this experiment calculates the privacy disclosure of public figures in different fields and conducts horizontal and vertical analysis to obtain different privacy disclosure characteristics. The experiment not only shows the situation of privacy disclosure, but also gives suggestions and method to reduce privacy disclosure.

KEYWORDS

Information Protect, Ontology, Ontology Relationship, Path Traversing, Privacy Disclosed Experiment, Privacy Protection, Privacy Quantification, Quantization Algorithm

1. INTRODUCTION

In recent years, hack attack, virus, trojan and personal information disclosure are increasingly exposed to public view, network security has attracted much attention. The existing network security technologies have a broad definition of protected data, which does not divide the types of data. Most of the time, they treat the protected data equally. Network service providers don't think differently about different types of data when they grab network data, also they don't communicate with data providers(Vincent et al., 2011). In fact, different types of data need different degrees of protection, especially personal privacy data. For example, someone's gender is one privacy data that can be open to a lot of people, such as friends, colleagues and so on, passwprd of bank card is also kind of privacy data, but it is confidentiality to others, which can only be kept by himself.

DOI: 10.4018/IJISP.2022010101

Many network activities may cause personal privacy information disclosure, such as website registration, online shopping, clicking links and so on. What information is disclosed in process of network activity is difficult to capture, even if it can be captured, it is difficult to measure what impact it will have on the data owner after privacy information is disclosed. However, there is a certain connection between different types of personal privacy data(Omoronyia, 2016), more data can be derived from one or more exist data. However, these connections are not considered in network activities. Most existing privacy protection schemes can only match individual privacy information. The P3P platform(Cranor et al., 2002) which proposed in 2007 provides a method for personalized privacy protection, enabling users to define their privacy preferences. But P3P platform has not done any processing for the relationship between different privacy information, also has not used specific numbers to quantify these links.

Ontology is a good tool to find the relationship between different data. In an ontology library, the correlation between two separate data can be calculated. Because of the existence of reasoning relationship, most of the existing privacy protection methods will actually still cause privacy disclosure. In order to understand the situation of privacy disclosure in real world, this paper conducts a privacy disclosure measurement experiment on ten public figures in different fields, so as to understand what kind of privacy information is easily disclosed by people in different fields. By calculating of a privacy measurement algorithm based on WordNet ontology, the degree of disclosure of these public figures' privacy information is quantified with a specific number and analyzed comparatively.

The rest of this paper is organized as follows. Section 2 review related work include privacy protection, ontology based privacy protection and privacy measurement. Section 3 introduces the relationship and structure of WordNet. Section 4 builds a privacy disclosure model and defines the goal. Section 5 introduces the quantification algorithm, include the framework of algorithm, the relationship handling and the characteristics of algorithm. Section 6 is the experimental part and section 7 is the conclusion.

2. RELATED WORK

There are many hot topics in the field of privacy protection, such as location based privacy protection, data publishing based privacy protection, Internet of things based privacy protection, video privacy protection and so on. This paper mainly considers privacy protection in data dissemination. In this respect, the more common way is to restrict publication by defining rules or encrypting data. Huo et al.(2018) proposed a new logic that combines distributed logic with distributed temporal logic to enable all parties in social network to implement a balance strategy based on pre-defined rules and reached a contractual access control. Ulybyshev et al.(2017) designed a cloud enterprise framework to ensure that each service can only access authorized database fields by assigning keys in untrusted cloud platform. Chen et al.(2015) proposed a data distribution protocol called ReDD, which enables high quality information to be transmitted in a mobile social network by estimating the quality of messages, without revealing user's privacy information. Some scholars had begun privacy protection from known information, such as Zhang et al.(2018) considered a variety of medical knowledge, established a destination tree, and dynamically defined the access boundary of hospital information system based on context. These methods could all achieve privacy protection to a certain extent, but they didn't divide the protected privacy information based on meanings, or to say, the division is not detailed enough to connect privacy information which had relationships in terms of meaning.

With the application of ontology technology to the computer field, some scholars began to achieve privacy protection based on ontology. Tian et al.(2017) proposed an ontology based service rating framework, which was based on semantic similarity to match requirements and services, thereby reducing privacy disclosure. Chou et al.(2017) introduced ontology technology regarding privacy disclosure problems in the field of network visualization, implements a visual interface based on ontology, and assists with some privacy protection operations. Belaazi et al.(2015) proposed that