2021 International Conference on Emerging Technology in Computing, Communication and Electronics (ETCCE)

2021 Emerging Technology in Computing, Communication and Electronics (ETCCE) | 978-1-6654-8364-3/21/$31.00 ©2021 IEEE | DOI: 10.1109/ETCCE54784.2021.9689791

A Conceptual Anonymity Model to Ensure Privacy for Sensitive Network Data

N H M Arafat*
Dept. of Computer Science and Technology
Henan Polytechnic University
454003, Jiaouo, Henan, P.R. China
arafat.nhm@gmail.com

Md Ileas Pramanik
Dept. of Computer Science and Engineering
Begum Rokeya University
5404, Rangpur, Bangladesh
ileas.cse@brur.ac.bd

Abu Jafar Md Muzahid
Faculty of Computing
Universiti Malaysia Pahang
26600, Pahang, Malaysia
mrumi98@gmail.com

Bibo Lu
Dept. of Computer Science and Technology
Henan Polytechnic University
454003, Jiaouo, Henan, P.R. China
lubibo@hpu.edu.cn

Sumaiya Jahan
Dept. of Computer Science and Technology
Henan Polytechnic University
454003, Jiaouo, Henan, P.R. China
sumaiya2110jahan@gmail.com

Saydul Akbar Murad
Faculty of Computing
Universiti Malaysia Pahang
26600, Pahang, Malaysia
saydulakbarmurad@gmail.com

***Abstract*—In today's world, a great amount of people, devices, and sensors are well connected through various online platforms, and the interactions between these entities produce massive amounts of useful information. This process of data production and sharing appears to be on the rise. The growing popularity of this industry, as well as the required development of data sharing tools and technology, pose major threats to an individual's sensitive information privacy. These privacy-related issues may elicit a regularly strong negative reaction and restrain further organizational invention. Researchers have identified the privacy implications of large data collections and contributed to the preservation of data from unauthorised exposure to solve the challenge of information privacy. However, the majority of privacy strategies concentrate solely on traditional data models, such as micro-data. The academe and industry are paying more attention to network data privacy challenges. In this paper, we offer (ℓ, k)-anonymity, a novel privacy paradigm for network data that focuses on maintaining the privacy of both node and link information. Here, original network data will turn to attribute generalization nodes through a complex process, where several algorithms, clustering, node generalization, link generalization and ℓ-diversification will be applied. As a result, (ℓ, k)-anonymous network will be generated and will filter original network data to ensure publishable (ℓ, k)-anonymize data. Hopefully, this anonymity model will have a stronger role against homogeneity attacks of intruders, which will prevent the unauthorized disclosure of sensitive network data for several areas, such as - health sector. This model will also be cost effective and data loss will be controlled using two different ways.**

***Index Terms*—Privacy, k-anonymity, Network-data, Data publishing, Information Loss.**

I. Introduction

With the advancement of computational technology, a massive amount of individual data is generated by various people and collected by numerous organisations on a regular basis, providing tremendous value to our society. These data generate a big data volume, with the majority of datasets being network data. Sociologists and computer scientists agree that social network communication has recently become a very frequent contact tool around the world, and that trend will continue in the future [1], [2]. Because of the emergence of various social networks, data sets have evolved from basic traditional data models to complex ones. Entities and relationships between them indicate social network users. Identifiers, quasi-identifiers, and sensitive information are used to represent entities or nodes, while relationships are used to indicate connections or links [3]. These large datasets also present severe privacy problems, resulting in a regulatory backlash and hindered organizational innovation. Furthermore, social network research focuses on uncovering structures that indicate user relationships [4], [5]. In the last two decades, researchers have looked into privacy preserving approaches to meet the difficulty of social network privacy. Furthermore, research on salable methods for preserving privacy using social network data is gaining traction in academia and industry [6]. The development of effective and efficient solutions for varied data models [7] is a growing trend in data privacy research. Because of the Health Insurance Portability and Accountability Act, the majority of present privacy work focuses on the healthcare domain [8]. However, privacy anxieties have been expanded