Enabling Secure Cross-Modal Retrieval Over Encrypted Heterogeneous IoT Databases With Collective Matrix Factorization

Cheng Guo, *Member, IEEE*, Jing Jia, Yingmo Jie, Charles Zhechao Liu, and Kim-Kwang Raymond Choo, *Senior Member, IEEE*

***Abstract*—Significant volume of information of a broad variety (or modalities, such as image, audio, video, and text) is sensed and collected [such as those by the Internet of Things (IoT) devices] regularly (e.g., hourly). Such information is then analyzed to inform decision making, such as clinical diagnosis and product recommendation. Data with different representations may have the same semantic information, and there have been considerable efforts devoted to designing efficient searching approaches on objects with different modalities. However, multimodal data carry sensitive information, and maintaining privacy is crucial in our privacy-aware and interconnected society. In this article, we combine both the collective matrix factorization (CMF) and homomorphic encryption (HE) to construct an efficient and accurate scheme to facilitate cross-modal retrieval, without the loss of any sensitive information. Our scheme identifies the unified feature vectors for every object in the training set with different modalities and obtains the mapping matrices for out-of-sample objects. After the encryption process, these matrices are stored on the remote cloud server (CS). Hence, the server can calculate the secure, unified features for any query. In this article, we also built a privacy-preserving index structure using locality-sensitive hashing (LSH), which provides both security and efficiency. Performance evaluations demonstrate the potential for our proposed scheme in the real-world IoT applications.**

***Index Terms*—Collective matrix factorization (CMF), homomorphic encryption (HE), locality-sensitive hashing (LSH), secure cross-modal retrieval (SCMR).**

Manuscript received October 4, 2019; revised November 23, 2019; accepted January 1, 2020. Date of publication January 8, 2020; date of current version April 14, 2020. This work was supported in part by the National Science Foundation of China under Grant 61501080, Grant 61871064, and Grant 61877007; in part by the Fundamental Research Funds for the Central Universities under Grant DUT19JC08; and in part by the Guangxi Key Laboratory of Trusted Software under Grant kx201903. The work of Kim-Kwang Raymond Choo was supported by the Cloud Technology Endowed Professorship and the National Science Foundation Centers of Research Excellence in Science and Technology (CREST) under Grant HRD-1736209. *(Corresponding author: Kim-Kwang Raymond Choo.)*

Cheng Guo, Jing Jia, and Yingmo Jie are with the School of Software Technology, Dalian University of Technology, Key Laboratory for Ubiquitous Network and Service Software of Liaoning Province, Dalian 116620, China, and also with the Guangxi Key Laboratory of Trusted Software, Guilin University of Electronic Technology, Guilin 541004, China (e-mail: guocheng@dlut.edu.cn; jiajing1995@163.com; jymsf2015@mail.dlut.edu.cn).

Charles Zhechao Liu and Kim-Kwang Raymond Choo are with the Department of Information Systems and Cyber Security, University of Texas at San Antonio, San Antonio, TX 78249 USA (e-mail: charles.liu@utsa.edu; raymond.choo@fulbrightmail.org).

Digital Object Identifier 10.1109/JIOT.2020.2964412

I. Introduction

INTERNET of Things (IoT) devices sense, collect, and transfer significant volume of data, and such data may exist in heterogeneous types (or modalities, such as image, audio, video, and text). Collectively, data from different sources can contain information of commercial and societal interests. Hence, there have been efforts to design cross-modal searches to facilitate the retrieval of heterogeneous data containing the same latent semantic meaning. For example, wearable smart healthy devices can monitor user physiological data, which can be used to inform medical diagnosis, and data collected by vision assistive devices can improve the quality of life for the visually impaired individual [1]–[4].

While a broad range and types of data collected by sensing devices can enrich the representation of things, the storage and computational capabilities of sensors are limited. The combination of IoT and cloud services can be utilized to process IoT data. However, data privacy is a potential concern. Therefore, in this article, we seek to determine how one can bridge two different modalities to facilitate searching for the same semantic information without affecting the privacy of the original data. Also, in this article, we focus on multimedia data.

A number of cross-modal retrieval methods have been developed in the literature, such as those designed for visual classification, searching of media, recognizing actions, and visual representations [5]–[8]. These methods are generally capable of filling the semantic gap among heterogeneous sources of data, but they are not designed to preserve user privacy. Thus, when sensitive, unencrypted data are uploaded to a search engine, privacy cannot be maintained, and an adversary may gain access to private information. However, after the original files are encrypted, the correlation between two similar files cannot be discerned. In other words, encryption complicates searching operations. Thus, we focus on achieving searchable, cross-modal encryption in this article.

The concept of searchable encryption (SE) was coined to facilitate searching and retrieval of encrypted contents that contained a concrete query keyword [9], and since then SE has been extended to support searching of multiple keywords, dynamic searching, ranked searching [10], [11], and other activities. For example, a number of SE methods have been proposed recently to provide secure image searching, such as label-based image searching and content-based image searching [12]–[14]. These methods were intended mainly

2327-4662 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See https://www.ieee.org/publications/rights/index.html for more information.

Authorized licensed use limited to: University of Texas at San Antonio. Downloaded on August 17,2020 at 17:06:34 UTC from IEEE Xplore. Restrictions apply.