Future Generation Computer Systems 62 (2016) 161–174

Contents lists available at ScienceDirect

Future Generation Computer Systems

journal homepage: www.elsevier.com/locate/fgcs

Efficient and privacy-preserving skyline computation framework across domains☆

Ximeng Liu [a], Rongxing Lu [a,*], Jianfeng Ma [b], Le Chen [a], Haiyong Bao [a]

[a] School of Electrical and Electronics Engineering, Nanyang Technological University, 50 Nanyang Avenue, 639798, Singapore

[b] School of Computer Science and Technology, Xidian University, Xi'an, Shaanxi, 710071, China

HIGHLIGHTS

- We propose an efficient and privacy-preserving skyline computation framework for multi-domains.
- Lightweight Additive Homomorphic Public Key Encryption Scheme are proposed for big data processing.
- Fast Secure Permutation Protocol and Fast Secure Integer Comparison Protocol are designed.
- Skyline set can be computed in an efficient and a privacy-preserving way by extensive simulation.

ARTICLE INFO

Article history:
Received 19 May 2015
Received in revised form
7 October 2015
Accepted 10 October 2015
Available online 30 October 2015

Keywords:
Skyline
Secure computation framework
Multi-domain
Semi-honest
Lightweight additive homomorphic encryption

ABSTRACT

Skyline computation, which returns a set of interesting points from a potentially huge data space, has attracted considerable interest in big data era. However, the flourish of skyline computation still faces many challenges including information security and privacy-preserving concerns. In this paper, we propose a new efficient and privacy-preserving skyline computation framework across multiple domains, called EPSC. Within EPSC framework, a skyline result from multiple service providers will be securely computed to provide better services for the client. Meanwhile, minimum privacy disclosure will be elicited from one service provider to another during skyline computation. Specifically, to leverage the service provider's privacy disclosure and achieve almost real-time skyline processing and transmission, we introduce an efficient secure vector comparison protocol (*ESVC*) to construct EPSC, which is exclusively based on two novel techniques: fast secure permutation protocol (*FSPP*) and fast secure integer comparison protocol (*FSIC*). Both protocols allow multiple service providers to calculate skyline result interactively in a privacy-preserving way. Detailed security analysis shows that the proposed EPSC framework can achieve multi-domain skyline computation without leaking sensitive information to each other. In addition, performance evaluations via extensive simulations also demonstrate the EPSC's efficiency in terms of providing skyline computation and transmission while minimizing the privacy disclosure across different domains.

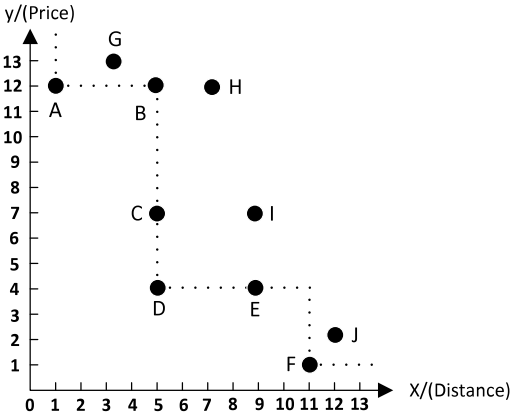
1. Introduction

As surveyed by IDC and EMC [1], the digital universe (all the digital data created, replicated and consumed in that year) was 130 exabytes in 2005 and forecasted to grow to 40,000 exabytes by 2020—almost doubled every two years. Such a large amount of data cannot be stored in a centralized way and are thus suggested to be stored in a distributed fashion. Although these data stored distributively seems irrelevant, the tremendous potential value contained in big data can be mined by using some data analytical methods and can contribute to many emerging areas such as e-healthcare, electronic commerce, government surveillance, etc. Due to this reason, big data analytic techniques have been drawn much attention not only by the government, but also by the industry and academia. For example, the US government announced "Big Data Research and Development Initiative" [2] to provide more than $200 million to extract knowledge and insights from large and complex collections of digital data. IBM released a unique big data analytic platform [3] allowing enterprises to address the full spectrum of big data business challenges. Some research labs [4,5] have also been built to focus on big data research. With such huge data generated everyday, it is undoubtedly challenging for us to design

☆ This research is supported by the Key Program of NSFC-Guangdong Union Foundation under Grant No. U1135002; The NSFC Program under Grant No. 61100214, No. 61402109, No. 61502248, and No. 61370078; The Nanyang Technological University under Grant MOE Tier 1 (M4011450).

* Corresponding author.
E-mail addresses: rxlu@ntu.edu.sg (R. Lu), jfma@mail.xidian.edu.cn (J. Ma).

http://dx.doi.org/10.1016/j.future.2015.10.005
0167-739X/



Point	Coordinates
A	(1,12)
B	(5,12)
C	(5,7)
D	(5,4)
E	(9,4)
F	(11,1)
G	(3,13)
H	(7,12)
I	(9,7)
J	(12,2)

Fig. 1. An example of traditional skyline.

a new processing tool for capturing, storing, managing and analyzing data.

Skyline computation is an important operation in many applications to return a set of interesting points which are not dominated by any other point from a huge multidimensional data space [6,7]. Due to its power of retrieving data from multidimensional data set, skyline computation plays an important role in multi-criteria decision making and client preference applications. For example, consider a database which contains information about hotels with numerous dimensions. A tourist can issue a skyline query to retrieve the hotels with the cheapest prices and the shortest distance from bus stop, each point in the figure represents the price of the hotel per night and distance to the bus stop. As can be seen from Fig. 1, points *A*, *D*, *F* are not dominated by any other point in the space, they are called *skyline points* in the database.

Most existing works on skyline computation focused on providing efficient skyline computation over centralized data storage [8], however, it is impossible to store such large amount of data in a sole database in practical scenarios. These data are more commonly stored distributively in different service providers and inter-connected via Internet. The distributed skyline computation certainly can benefit clients by providing high-quality services, however, its flourish still hinges on understanding and managing the information security and privacy challenges, especially during the period of skyline computation over different domains. To clearly illustrate the challenge in multi-domains skyline computation, we consider the following scenario: a client wants to get a skyline set from a service provider. Because one service provider, containing only a small set of data, cannot provide enough services to the client, more service providers should be involved into the system for skyline processing. Although the idea is promising, these service providers cannot offer the entire database directly to the others because these service providers are always commercial companies and their databases are usually regarded as the trade secrets. In addition, these private information cannot be leaked to unauthorized parties during the data transmission and skyline computation.

In this paper, to address the above-mentioned privacy issues in the skyline computation, we propose an **E**fficient and **P**rivacy-preserving **S**kyline **C**omputation framework across different domains, called EPSC. Specifically, the main contributions of this paper are Fourfold.

- Firstly, we propose EPSC, an efficient and privacy-preserving skyline computation framework for multi-domains. With EPSC, the databases in different domains can be gathered together to offer a better skyline service to the requesting clients. During the skyline set processing, each domain will not directly reveal its own data to the other domains.
- Secondly, as the basis of realizing EPSC, we present a new cryptosystem called Lightweight Additive Homomorphic Public Key Encryption Scheme (LAHE) in order to fit for big data processing. Our LAHE can achieve message additive homomorphism with extremely low computational and communication cost.
- Thirdly, to achieve the EPSC, an Efficient Secure Vector Comparison Protocol (*ESVC*) is introduced to compare relationships between two vectors. In order to implement *ESVC*, two basic techniques called Fast Secure Permutation Protocol (*FSPP*) and Fast Secure Integer Comparison Protocol (*FSIC*) are designed as the components of *ESVC*. Notice that secure permutation protocol and secure integer comparison protocol have been studied in privacy-preserving data mining [9], however, most of them rely on time-consuming homomorphic encryption techniques. To the best of our knowledge, our *FSIC* and *ESVC* are the most efficient one in both computational cost and communication overhead.
- Finally, to validate the efficiency of EPSC, we also develop a custom simulator built in Java. The extensive simulation results show that our EPSC can help the client to get a skyline set efficiently and minimize domain privacy disclosure without overburdening the whole framework.

The remainder of this paper is organized as follows: In Section 2, we formalize the system model, the security requirements and identify our design goals. In Section 3, we describe some preliminaries which serve as the basis of our proposed framework. Then, we present the proposed EPSC in Section 4, followed by security analysis and performance evaluation in Section 5 and Section 6, respectively. In Section 7, we also discuss some related works. Finally, we draw our conclusions in Section 8.

2. Models and design goal

In this section, we formalize the system model, security model, and identify our design goals.

2.1. System model

In our system model, we mainly focus on how to respond client's skyline query in a privacy-preserving way. Specifically, we consider the system model by dividing EPSC into three parts: clients who request the skyline service, Local Skyline Service Provider (LSP) and Collaborative Skyline Service Provider (CSP), as shown in Fig. 2.

(1) *Clients*: Clients can launch either local skyline query or global skyline query across different service providers. When skyline set is needed, a client just sends his request to LSP. Once a