Feedback Algorithms Enabling Privacy-Preserving Location Proofs for Mobile Users

Rajashekar M.B[1], Vishwesh J[2], Yathiraj G.R[3], Akshatha M[4]
[1, 2, 3, 4] Department of Computer Science and Engineering
[1, 2, 3, 4] GSSSIETW, Mysuru.

Abstract- *In this paper, we propose a class of RF algorithms inspired by quantum detection to re-weight the query terms and to re-rank the document retrieved by an IR system. These algorithms project the query vector on a subspace spanned by the eigenvector which maximizes the distance between the distribution of quantum probability of relevance and the distribution of quantum probability of non-relevance. The experiments showed that the RF algorithms inspired by quantum detection can outperform the state-of-the-art algorithms. We present the Spatial-Temporal provenance Assurance with Mutual Proofs (STAMP) scheme. STAMP is designed for ad-hoc mobile users generating location proofs for each other in a distributed setting. However, it can easily accommodate trusted mobile users and wireless access points. STAMP ensures the integrity and non-transferability of the location proofs and protects users' privacy. A semi-trusted Certification Authority is used to distribute cryptographic keys as well as guard users against collusion by a light-weight entropy-based trust evaluation approach. Our prototype implementation on the Android platform shows that STAMP is low-cost in terms of computational and storage resources. Extensive simulation experiments show that our entropy-based trust model is able to achieve high collusion detection accuracy.*

I. INTRODUCTION

IR is concerned with indexing and retrieving documents including information relevant to a user's information need. Although the end user can express his information need using a variety of means, queries written in natural language are the most common means. However, a query can be very problematic because of the richness of natural language. Indeed, a query is usually ambiguous; a query may express two or more distinct information needs or one information need may be expressed by two or more distinct queries. Consider topic 329 which is provided with the Text Retrieval Conference (TREC) test collection1 from which the query Mexico City has the worst air pollution in the world. Pertinent documents would contain the specific steps Mexican authorities have taken to combat this deplorable situation. is submitted to an IR system based on the Vector Space Model (VSM). This system would return both relevant documents and irrelevant documents as shown in Fig. 1a.Although the number of relevant documents in the top ten document list is quite high, there are some irrelevant documents– for example, LA062790-0048 is irrelevant because itis about a very specific case of river pollution at the Mexican border – and the Mean Average Precision (MAP) is only15.2 percent. An IR system addresses the problems caused by query ambiguity by gathering additional evidence that can bused to automatically modify the query [3]. Usually a query is expanded because the queries are short and it cannot exhaustively describe every aspect of the user's information need; however, some irrelevant documents may be retrieved or relevant documents may also be missed when a query is not short as shown in the previous example. RF can be positive, negative or both. Positive RF only brings relevant documents into play and negative RF makes only use of irrelevant documents; any effective RF algorithms include a "positive" component. Although positive feedback is a well established technique by now, negative feedback is still problematic and requires further investigation, yet some proposals have already been made such as grouping irrelevant documents before using them for reducing the query [38].As Location-Enabled mobile devices proliferate, location-based services are rapidly becoming immensely popular. Most of the current location-based services for mobile devices are based on users' current location. Users discover their locations and share them with a server. In turn, the server performs computation based on the location information and returns data/services to the users. In addition to users' current locations, there is an increased trend and incentive to prove/validate mobile users' past geographical locations. This opens wide variety of new location-proof based mobile applications. Saroiu et al. described several such potential applications in [1].Let us consider three examples: (1) A store wants to offer discounts to frequent customers. Customers must be able to show evidence of their repeated visits in the past to the store. (2) Accompany which promotes green commuting and wellness may reward their employees who walk or bike to work. The company may encourage daily walking goals of some fixed number of miles. Employees need to prove their past commuting paths to the company along with time history. This helps the company in reducing the healthcare insurance rates and move towards sustainable lifestyle. (3) On the battlefield, when a scout group is sent out to execute a mission, the commanding center may want every

soldier to keep a copy of their location traces for investigation purpose after the mission.

II. BACKGROUND

In this section, we illustrate the main technical background of the framework proposed in this paper

2.1 Vector Space Model

The VSM for IR represents both documents and queries as vectors of the k-dimensional real space Rk [34]. This vector space is defined by k basis vectors corresponding to the terms extracted from a document collection; for example, if the document collection stores three documents “orange juice”, “apple juice” and “apple”, the vector space is defined by three canonical basis vectors corresponding to “apple”, “juice” and “orange”, and the three documents are represented, respectively, by the following vectors. Each document vector results from the weighted linear combination of the basis vectors which represents the terms extracted from the document collection. In the example above, the weights are binary, that is,1 if the term occurs in a document, 0 otherwise. Other weighting schemes that assign vector coordinates are reported for example in [32] and [34]. The state-of-the-artist given by the pivoted normalization [36].

2.2 Relevance Feedback

The RF algorithm is also known as Rocchio’s algorithm [25]and it is designed to compute the new query vector using alinear combination of the original vectors, the relevant document vectors and the non-relevant document vectors, where the labels of relevance are collected in a training set. Suppose y is the query vector, x1; . . . ; xR are R relevant document vectors in Rk, and xRþ1; . . . ; xN are N _ R no relevant document vectors in Rk. The RF computes the following new query victory _ ¼ yz}|{ original query yþz}|{positive RF_ y_z}|{ negative RFmodified query; (3)whereby ¼1RXRi¼1xi (4)involves relevant document vectors and y_ ¼PNi¼Rþ1 xi=ðN _ Rfnvolves non-relevant document vectors.

III. RELATED WORK

The notion of unforgivable location proofs was discussed by Waters et al. [10]. They proposed a secure scheme which advice can use to get a location proof from a location manager. However, it requires users to know the verifiers as a prior.Saroiu et al. [1] proposed a secure location proof mechanism, where users and wireless APs exchange their signed public keys to create time stamped location proofs. These schemes are susceptible to collusion attacks where users and wireless APs may collude to create fake proofs. VeriPlace [2] is a location proof architecture which is designed with privacy protection and collusion resilience. However, it requires three different trusted entities to provide security and privacy protection: a TTPL (Trusted Third Party for managing Location in formation), a TTPU (Trusted Third Party for managing User information) and a CDA (Cheating Detection Authority). Each trusted entity knows either a user's identity or his/her location, but not both. VeriPlace's collusion detection works only if users request their location proofs very frequently so that the long distance between two location proofs that are chronologically close can be considered as anomalies. This is not a realistic assumption because users should have the control over the frequency of their requests. Hasan et al. [5] proposed a scheme which relies on both location proofs from wireless APs and witness endorsements from Bluetooth-enabled mobile peers, so that no users can forge Proofs without colluding with both wireless APs and other mobile peers at the same time. It eliminates the necessity of multiple trusted parties. Two privacy preserving schemes based on hash chains and Bloom filters respectively are described for protecting the integrity of the chronological order of location proofs. All the above systems are centralized, that is, they all require central infrastructures (wireless APs) to act as the location authorities and generate location proofs. However, we want to design framework that can also work for distributed scenario where users are far from any trusted Aping Davis et al.'s alibi system [6], their private corroborator scheme relies on mobile users within proximity to create alibi's(i.e., location proofs) for each other. The security and privacy of the system is achieved based on a cryptographic commitment scheme. However, they do not deal with any collusion attacks

1) Location Granularity Levels: We assume there are granularity levels for each location, which can be denoted by , where represents the finest location granularity (e.g., an exact Geo coordinate), and represents the most coarse location granularity (e.g., a city). Hereafter, we refer to location granularity level as location level for short. When a location level is known, we assume it is easy to obtain a corresponding higher location level where .The semantic representation of location levels is assumed to be standardized throughout the system.

2) Cryptographic Building Blocks: STAMP uses the concept of commitments to ensure the privacy of proves. A commitment scheme allows one to commit to a message while keeping it hidden to others, with the ability to reveal the committed value later. The original message cannot be changed after it is committed to. A commitment to a message

can be denoted as where is a nonce used to randomize the commitment so that the receiver cannot reconstruct , and the commitment can later be verified when the sender reveals both and .A number of commitment schemes [14]–[16] have been proposed and commonly used. Our system does not require a specific commitment scheme. Any scheme which is perfect binding and computational hiding can be used. In our implementation, we used [14], which is based on one-way hashing.

A. Protocol

1) Overview: Our protocol consists of two primary phases: STP proof generation and STP claim and verification.

Fig. 2 gives an overview of the two phases and the major communication steps involved. When a prove collects STP proofs from his/her co-located mobile devices, we say an STP proof collection event istartby the prove. An STP proof generation phase is the process of the prover getting an STP proof from one witness. Therefore, an STP proof collection event may consist of multiple STP proof generations. The prover finally stores the STP proofs he/she collected in the mobile device. When a prover encounters a verifier (the frequency of such encounters is specific to the application scenarios) and he/she intends to make a claim about his/her past STP to the verifier, the STP claim and verification phase takes place between the prove and the verifier. A part of the verification job has to be done by CA. Therefore, communication between the verifier and Cheapens in the middle of the STP claim and verification phase.

IV. PROPOSED SYSTEM

We propose a class of RF algorithms inspired by quantum detection to re-weight the query terms and to re-rank the document retrieved by an IR system. These algorithms project the query vector on a subspace spanned by the eigenvector which maximizes the distance between the distribution of quantum probability of relevance and the distribution of quantum probability of non-relevance In this paper, we define the past locations of a mobile user at a sequence of time points as the spatial-temporal provenance (STP) of the user, and a digital proof of user's presence at a location at a particular time as an STP proof. In this paper, we propose an STP proof scheme named Spatial-Temporal provenance Assurance with Mutual Proofs (STAMP). STAMP aims at ensuring the integrity and non-transferability of the STP proofs, with the capability of protecting users' privacy. We propose an entropy-based trust model to detect the collusion scenario. A distributed STP proof generation and verification protocol (STAMP) is introduced to achieve integrity and non-transferability of STP proofs. No additional trusted third parties are required except for a semi-trusted CA.STAMP is designed to maximize users' anonymity and location privacy. Users are given the control over the location granularity of their STP proofs. STAMP is collusion-resistant. The Bussard-Bagga distance bounding protocol is integrated into STAMP to prevent a user from collecting proofs on behalf of another user. An entropy-based trust model is proposed to detect users mutually generating fake proofs for each other. STAMP uses a entropy-based trust model to guard users from prove-witness collusion. This model also encourages witnesses against selfish behavior.

V. CONCLUSION

In this paper we have presented STAMP, which aims at providing security and privacy assurance to mobile users' proofs for their past location visits. STAMP relies on mobile devices in vicinity to mutually generate location proofs or uses wirelesses to generate location proofs. Integrity and non-transferability of location proofs and location privacy of users are the main design goals of STAMP. In this paper, a class of RF algorithms inspired by quantum detection has been proposed to re-weight query terms by projecting the query vector on the subspace represented byte eigenvector which is the optimal solution to the problem of finding the maximal distance between two quantum probability distributions. RF is then viewed as a signal detection technique – relevance is the document state to be detected and the queries are the detectors. First, the documents retrieved by an IR system to answer the original query are used to extract a feature matrix. Second, some relevance assessments are obtained according to whether RF is explicit or pseudo. The quantum probability distributions can be estimated and the optimal solution of a distance between two quantum probability distributions cane calculated. The eigenvector that results from this optimization problem can be utilized to project the query vector. Third, the retrieved documents can be re-ranked to answer the modified query.

REFERENCES

[1] A. Aji, Y. Wang, E. Agichtein, and E. Gabrilovich, “Using the past to score the present: Extending term weighting models through revision history analysis,” in Proc. 19th ACM Conf. Inf. Knowl. Manage., 2010, pp. 629–638.

[2] R. Blanco and P. Boldi, “Extending BM25 with multiple query operators,” in Proc. 35th Int. ACM SIGIR Conf. Res. Develop. Inf. Retrieval, 2012, pp. 921–930.

[3] C. Carpineto and G. Romano, "A survey of automatic query expansion in information retrieval," ACM Comput. Surv., vol. 44,no. 1, pp. 1–50, Jan. 2012.

[4] K. Collins-Thompson, P. N. Bennett, R. W. White, S. de laChica, and D. Sontag, "Personalizing web search results by reading level," in Proc. 20th ACM Int. Conf. Inf. Knowl. Manage.,2011, pp. 403–412.

[5] W. Croft, D. Metzler, and T. Strohman, Search Engines: Information Retrieval in Practice. Reading, MA, USA: Addison-Wesley, 2009.

[6] I. Frommholz, B. Larsen, B. Piwowarski, M. Lalmas, P. Ingwersen, and K. van Rijsbergen, "Supporting polyrepresentation in a quantum- inspired geometrical retrieval framework," in Proc. 3rd Symp. Inf. Interaction Context ins, 2010, pp. 115–124.

[7] I. Frommholz, B. Piwowarski, M. Lalmas, and K. van Rijsbergen, "Processing queries in session in a quantum-inspired IR framework," in Proc. Eur. Conf. Inf. Retrieval, 2011, pp. 751–754.

[8] R. B. Griffiths, Consistent Quantum Theory. Cambridge, U.K.: Cambridge Univ. Press, 2002.

[9] M. Gupta and M. Bendersky, "Information retrieval with verbose queries," Found. Trends Inf. Retrieval, vol. 9, nos. 3/4, pp. 91–208,2015.

[10] B. Waters and E. Felten, "Secure, private proofs of location," Departmentof Computer Science, Princeton University, Princeton, NJ, USA,Tech. Rep., 2003.

[11] X. Wang et al., "STAMP: Ad hoc spatial-temporal provenance assurance for mobile users," in Proc. IEEE ICNP, 2013, pp. 1–10.

[12] A. Pfitzmann and M. Köhntopp, "Anonymity, unobservability, and pseudonymity-a proposal for terminology," in Designing Privacy Enhancing Technologies. New York, NY, USA: Springer, 2001.

[13] Y.-C. Hu, A. Perrig, and D. B. Johnson, "Wormhole attacks in wireless networks," IEEE J. Sel. Areas Commun., vol. 24, no. 2, pp. 370–380,Feb. 2006.

[14] S. Halevi and S. Micali, "Practical and provably-secure commitment schemes from collision-free hashing," in Proc. CRYPTO, 1996, pp. 201–215.

[15] I. Damgård, "Commitment schemes and zero-knowledge protocols," in Proc. Lectures Data Security, 1999, pp. 63–86.

[16] I. Haitner and O. Reingold, "Statistically-hiding commitment from any one-way function," in Proc. ACM Symp. Theory Comput., 2007, pp. 1–10.

[17] D. Singelee and B. Preneel, "Location verification using secure distance bounding protocols," in Proc. IEEE MASS, 2005.

[18] J. Reid, J. Nieto, T. Tang, and B. Senadji, "Detecting relay attacks with timing-based protocols," in Proc. ACM ASIACCS, 2007, pp. 204–213.

[19] C. Kim, G. Avoine, F. Koeune, F. Standaert, and O. Pereira, "The Swiss-knife RFID distance bounding protocol," in Proc. ICISC, 2009,pp. 98–115.