Decentralized Search And Retrieval For Mobile Networks Using SMS

Gaya Nair. P
M.Tech, CSE Department, Calicut University, Kerala, India

Abstract

Existing systems for mobile search are based on conventional centralized search engines on the Internet. The accessibility of the information over the Internet depends on the unbiased administration of centralized search engines and centralized search indexes. Centralized Internet search engines can be tampered with easily by their administrators to bias the results, concealing or censoring information. Because of these we cannot rely on centralized Internet search to remain unbiased forever. A trustworthy distributed search and retrieval system for the Internet is developed. This paper presents a theory-based literature review of the extant approaches used within Decentralized Search and Retrieval of information in mobile networks. The review also shows that the performance evaluation based on the number of nodes that are participated in the network.

1. Introduction

Mobile phones' computational power has been improving approaching the capabilities of general purpose computers. Mobile phones possess an extra set of concerns such as Personalization, Interactivity, Location and Context dependence and Dynamicity; that are not present in normal web servers. Today the human communications and interactions are based on mobile devices [7]. Anyone with access to a mobile phone, laptop, tablet, or other networked device can communicate with a friend or acquaintance simply and almost instantaneously. Recently, there has been a growing interest in how to explore the mobile phone capabilities in the web search context and how to merge them with existing phone functionalities. However the research has tended to focus on centralized approaches or Peer-to-Peer web search, rather than on the Peer-to-Peer web search in the social network context. In future the information can distribute, search for, and retrieve in a decentralized peer- to-peer fashion from one mobile device to another. To that end, a trustworthy distributed system called iTrust was developed. iTrust over SMS is a decentralized search and retrieval system that enables any two mobile devices to share information using the Short Message Service (SMS) that is available on many mobile phones.

2. Methodology

The purpose of this paper is to contribute to a shared understanding of the concepts of the iTrust decentralized network search systems and its associated benefits than the centralized search systems. To this end, we conduct a systematic survey of the extant literature and systematize this literature regarding the core concepts and the effects of the decentralized search systems.

The Centralized search systems use a limited set of pre-defined topics, and either special keywords within the search query a specialized parser to determine the intended topic. Those such systems tampered with easily by their administrators to bias the results, concealing or censoring of information. [1]. In this survey we concentrate on the search systems used for mobile search.

The rest of this paper is organized as follows. Section 3 describes the features of the mobile search and the techniques that are used before. Performance evaluation results, based on the analysis operational nodes in an iTrust network are presented in Section 3. Section 4 presents conclusions and future work.

2.1. Mobile Search Characteristics

Mobile search is viewed as an extension of the desktop search model. Yet the studies by Kamvar et al. [9,10] shown in recent that fundamentally it is different in several ways. The first study [9] found that the mobile search click-through rate and the search page views per query were both significantly lower in comparison to desktop search. It means that most mobile search users tend to use the search service for short time-periods and are either satisfied with the search engine snippet responses or do not find what they were looking for. The study also found that the persistence of mobile users was very low indicating that the vast majority of mobile searchers approach queries with a specific topic in mind and their search often does not lead to exploration. The second study [10] showed that the diversity of search topics for low-end phone users was much less than that of desktop or iPhone-based search. This result suggests that the information needs are broad, but are not satisfied by the information services available on low-end phones. As a whole, these studies indicate a pressing need for rethinking the current mobile search model for low-end mobile devices.

2.2. SMS-based Search Services

SMS-based search is very different from conventional mobile search via XHTML/WAP. An attractive aspect of SMS-based search is the lower barrier to entry of SMS due to the use of low-end phones and widespread availability of SMS. In developing countries, SMS is the most ubiquitous protocol for information exchange next to voice. In addition, economically SMS is cheap, it is reliable, it is universal, and it has unrivaled utility as a bearer for communications, information and services.

An SMS search system is not automatic and it does not provides accurate query responses. One reason for this problem is that search queries are inherently ambiguous, yet returning a disambiguated result is especially vital to SMS search queries for various reasons.

2.3. SMS-Based Web Search for Low-end Mobile Devices

SMSFind, is an SMS-based search system that enables users to obtain extremely concise appropriate search responses for queries across arbitrary topics in one round of interaction [1]. SMSFind is designed to complement existing SMS-based search services that are either limited in the topics they recognize or involve a human in the loop.

The SMS search system consists of a query server that handles the actual search query and results, and an SMS gateway that is responsible for communication between the phone clients and the query server. The client is a user with a mobile phone who sends an SMS message to the short code for our service, which arrives at the SMS gateway and is then dispatched to our server for processing. At our query server the query is then sent to a general search engine and result pages are downloaded. The query server extracts the results from the downloaded pages.

SMSFind search algorithm is used to extracts snippets from the downloads. To deploy SMSFind as a search service we implemented a front-end to send and receive SMS messages. This is the first effort that addresses the problem of SMS-based search for long tail mobile queries.

In terms of the performance of this system, it is observed that the queries that are not answered properly regardless of format are often ambiguous, explanations, enumerations, problems that require analysis, or time-sensitive queries. This is similar to statiistical techniques that have similar limitations. In addition this simple algorithm is able to answer over half of the dataset in complex queries.

2.4. Gnutella File Sharing Network with mobile agents

Gnetulla is a decentralized P2P file sharing protocol by sending “heartbeat” messages to its peers. This can result in devices transmitting and receiving large amounts of network traffic, but its use is bandwidth consuming due to the broadcast nature of some peer to peer protocols [3]. This is undesirable for mobile devices due to their bandwidth and power constraints. To address these issues, we introduced architecture that uses mobile agents to support mobile devices in a P2P network. A mobile agent attaches itself to the P2P network and acts on behalf of the mobile device, communicating with the mobile device using a lightweight communication protocol.

The architecture consists of a number of hosts on the network that contain execution environments for mobile agents. The execution environment not only provides a place for mobile agents to execute their program code, but also serve as a resource and security control for the host. The mobile agent that represents its mobile device executes the Gnutella file sharing protocol, joins the existing P2P network and communicates with other P2P hosts in one of these execution environments. Finally, the mobile device and the mobile agent communicate via a lightweight communication protocol, The mobile agent provides the advantages of reducing unnecessary traffic to the mobile device, greater support for mobility, and enabling the participation on the P2P network with fewer interruptions. Based on this concept a number of issues have to be address. These include, the lightweight communication protocol is text based, and it is of a request/response nature. This will affect the optimization of our lightweight communication protocol. Rather it generates a significant overhead from the migration of the mobile agent, and the effects of frequent migration on the architecture.

2.5. A Distributed Search Service for Peer-to-Peer File Sharing in Mobile Applications

The Passive Distributed Indexing (PDI), is a general-purpose distributed search service for document exchange in mobile applications, which is based on peer-to-peer technology. PDI defines a set of simple messages for transmission of queries and responses [4]. All messages are exchanged using local broadcast transmission. PDI supports forwarding of messages over several hops similar to techniques known form routing in mobile ad-hoc networks . Beside, PDI eliminates the need for flooding the entire network with query messages by maintaining an index cache at every device.

To implement PDI, each mobile device maintains a local repository, consisting of a set of files stored in the local file system. PDI provides search services for all documents in the repository. Each document can uniquely be identified by the path in the local file system together with a unique device identifier known as document identifier.

Besides the repository each mobile device implementing PDI maintains a local index cache. this index cache is used to store pairs of keywords and document identifiers investigated in recently received reports. The index cache can be used to answer popular queries without forwarding them to a device that actually stores a matching document.

The concept of Passive Distributed Indexing, a general-purpose distributed document search service for mobile file sharing applications. PDI is based on peer-to-peer technology, i.e., PDI does not require any centralized infrastructure for providing searching capabilities.

The issues related to this concept are : First, the impact of document modifications on the performance of PDI and design appropriate mechanisms for providing index caches consistency. Second, we have to evaluate the performance of PDI considering more sophisticated workload models, e.g. workloads consisting of location depended queries. Third, we have to develop a prototype implementation of PDI and test it in a mobile e-learning environment

2.6. Mobile Search –Social network search using mobile devices

The Mobile Search system is based on pure Peer-to-Peer architecture and it offers scalability, efficiency, resilience to failures and privacy at a higher degree than current centralized solutions. The advantage is that we can navigate through the data in a social network.

Social network's connections are determined from an address book of a mobile device. Users search one graph level of their social network at a time usually starting from their neighbours. However, users may also start a query anywhere in the social network. Every time a user issues a search query the mobile device forwards it to all its neighbours. The neighbours answer back by returning a result set and a list of their neighbours. If the user who issued the query is not satisfied by the results he can always ask new results from the next level neighbours as long as there are non-visited nodes in the network. This concept was named manual multihopping. In manual multi-hopping the user needs to select which of the non-visited nodes will be used for querying the next level.

Another way of navigating is by searching neighbour content tags and getting the result set composed by the content links with the tags and the list of next level neighbors. Tags work as links between content categorized similarly. At each hop the user gets the list of contents tagged in a similar way by nodes in its neighbourhood.

The Mobile Search system can be divided to two logical parts: local web search engine and meta crawling. Local web search engine is a search service, which manages the search index of a mobile device. Meta- crawling term refers a search service, which uses other local web search engines for getting the results and then combines different result sets into one. The part responsible for the meta crawler gets it's results from direct neighbors. The way the results are presented can always be changed thus the mobile device bears the load of processing the returned references.

Mobile Search complements traditional web search engines. It gives the user means to explore the neighbours' contents by traveling to the friends network topology. It covers a multitude of environments not covered by the centralized solutions.

One of the main advantages in relation to current centralized social network sites is the possibility to manage the site without interference from an external entity.

2.7. Trustworthy Distributed Search and Retrieval over the Internet

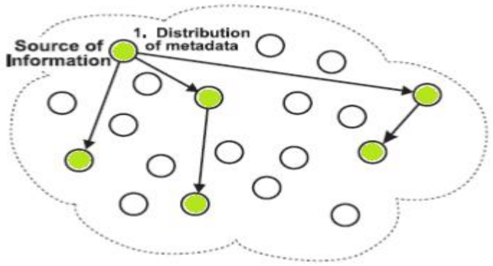
Existing systems for mobile search, are based on conventional centralized search engines on the Internet. Those systems use a limited set of pre-defined topics, and either special keywords within the search query or a specialized parser to determine the intended topic [2]. The centralized search engines are subject to censorship, filtering, and subversion of information. Because of these the centralized Internet search to remain unbiased forever. It is important to ensure that a trustworthy distributed search and retrieval system iTrust is developed.

The iTrust system involves distribution of metadata and requests, matching of requests and metadata, and retrieval of information corresponding to metadata. iTrust has no centralized mechanisms that can be tampered with easily by a small group of administrators. iTrust is inevitably more costly in bandwidth, processing and storage than a centralized search engine.

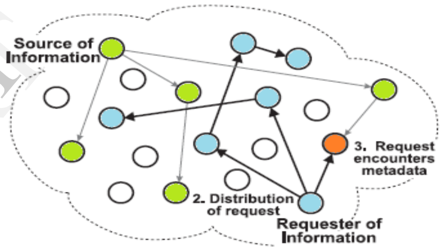
The iTrust system is deployed on a set of participating nodes in the Internet which is referred to as the membership. iTrust distributes both metadata that describes information, and requests for information, to a random subset of the participating nodes in the Internet. Because the metadata and the requests are distributed to nodes that are chosen at random from among all of the participating nodes, no one node or small group of nodes can suppress or censor information.

In the iTrust system, source nodes produce information and publish that information to make it available to other participating nodes. The source nodes create metadata keywords for their information, and communicate that metadata, together with a URL, to a subset of the participating nodes that are chosen at random, as shown in Figure 1. Requesting (querying) nodes generate requests (queries), containing metadata keywords for information that they seek to retrieve. The requesting nodes distribute their requests to a subset of the participating nodes that are chosen at random, as shown in Figure 2.

If a participating node receives a request, it compares the metadata in the request with the metadata that it holds. If the metadata match, which we call an encounter or a match, the matching node returns to the requesting node the URL that the source node included with the metadata, as shown in Figure 3. The requesting node then uses the URL to retrieve the information from the source node.



“Figure.1: A source node distributes metadata, describing its information, to randomly selected nodes in the membership”



“Figure. 2: A requesting node distributes its request to randomly selected nodes in the membership. One of the nodes has both the metadata and the request and, thus, an encounter occurs.”

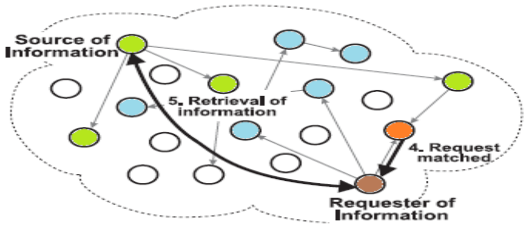


Fig. 3: A participating node matches the metadata and the request and reports the match to the requester, which then retrieves the information from the source node.

iTrust implementation on a node consists of the Web server foundation, the application infrastructure, and the public interface. These three components interact with each other to distribute metadata and requests, and to retrieve resources from the nodes.

2.8. iTrust Systems over HTTP and the Internet

iTrust information distribution and retrieval system, can be used to access information over HTTP and the Internet.The iTrust system operates over HTTP and, thus, TCP/IP.As such, it establishes a direct connection between any two nodes that need to communicate; the iTrust implementation uses neither flooding nor random walks.

iTrust uses to distribute metadata and requests [5]. This is implemented using PHP (PHP: Hypertext Preprocessor) on an Apache Web server, thereby allowing any user with a Web browser on any platform to interact with the node. Node information is stored locally in an SQLite database. Multiple nodes can be installed on a single Web server by creating multiple virtual Web sites; A single Web server have separate SQLite databases for multiple nodes.

2.9. iTrust: Trustworthy Information Publication, Search and Retrieval

For iTrust, a different messaging protocol is used for information publication, distribution and retrieval [6]. The design of the iTrust messaging protocol can be explained as: The source nodes produce and distribute their metadata randomly to a set of participating nodes in the network. Some of those nodes might forward the metadata they receive to other nodes in the network. The requesting nodes distribute their requests randomly to a set of participating nodes in the network. Again, some of those nodes might forward the requests they receive to other nodes in the network. If a node receives both the metadata and a request, the node determines whether the metadata and the keywords in the request match. If a node finds that its metadata matches the keywords in the request, the matching node provides, to the requesting node, the URL where the requesting node can retrieve the information. If a node finds that its metadata does not match the keywords in the request, it does nothing. The requesting node then retrieves the information from the source node using the URL provided by the matching node.For appropriately chosen parameters, it is probable that at least one node receives both the metadata and a request with corresponding keywords.

10. Detecting and Defending against Malicious Attacks in the iTrust

The iTrust system is a decentralized and distributed information retrieval system, that is designed to defend against censorship of information on the Internet. Even though the communication cost of iTrust is greater than that of a centralized search engine, the users are concerned about censorship or suppression of information, they should be willing to incur that extra cost. The decentralized and distributed nature of iTrust makes it very robust against malicious attacks that aim to prevent information retrieval.

However, a specific type of malicious attack that affect the iTrust systems is, to insert a large number of nodes in to the network that behave normally except that they do not match requests and metadata on certain topics [20]. The appropriate response to such an attack is to increase the number of nodes to which the metadata and the requests are distributed, to restore the probability of a match to the desired level.

The algorithm for detecting malicious attacks collects statistical data on the number of responses that a requesting node has received for a number of requests. Then, it computes the analytical probabilities of the exact number of matches for n, m and r and for different values of x. Where, n : The number of participating nodes, m: The number of participating nodes , r: The number of requesting nodes, and x: The number of operational nodes. Finally, it compares the empirical probabilities against the analytical probabilities of exact number of matches to estimate the proportion of operational nodes in the iTrust network.

The algorithm for defending against malicious attacks increases the number m of nodes to which the metadata are distributed and the number r of nodes to which the requests are distributed. It increases m and r to achieve the same probability of a match.

In iTrust, a different technique to be maintained for the desired degree of replication of the metadata and the requests. It is also the ongoing procedure to investigate other kinds of malicious attacks on iTrust, and detection and defensive adaptation algorithms for those kinds of malicious attacks.

2.11. iTrust over SMS System

The current development of iTrust uses Short Message Services (SMS) for information sharing and retrieval.

iTrust over SMS brings information sharing to any mobile device with instant text messaging capability. The iTrust over SMS system retains features of the iTrust over HTTP system that protect information against censorship, filtering, and subversion of information.

The iTrust over SMS system is completely independent from the mobile network service Providers. The System consists of a Short Message Service Center (SMSC), which store-and-forward message center for the sending and receiving of SMS text messages.

SMS message sent to the SMSC eventually reaches a mobile phone node; likewise, an SMS message received by the SMSC was originated by some mobile phone node. This conceptualization allows mobile phones using the iTrust over SMS API to be directly connected to each other in a peer-to-peer fashion.

The iTrust over SMS system is implemented as on the Android platform. The components are designed to be used in conjunction with any suitable graphical user interface or application, and are described without reference to a particular graphical user interface or application.

iTrust over SMS system can be used to public users and it also allows any type of data to be transmitted; The System can be extended to Wi-Fi Direct and/or Bluetooth to support search and retrieval over mobile ad-hoc networks.

3. Conclusion

This paper presents a theory-based literature review of the mobile search systems in the network . The majority of this research builds straight on the concepts of decentralized search systems in mobile network. This review points to research opportunities in decentralized search systems in mobile network. First, the search methods that are used in mobile networks are refined. However, such systems does not support security systems in network. So a trustworthy distributed search and retrieval system called iTrust is developed.A second avenue for future research is to provide this iTrust over SMS to public users and test the feasibility of average-size social networks in real-life scenarios using hysical mobile devices.

3. References

[1] J. Chen, L. Subramanian and E. Brewer, "SMS-based Websearch for low-end mobile devices," *Proceedings of the 16thACM International Conference on Mobile Computing and Networking*, Chicago, IL, September 2010, pp. 125–136.

[2] Y. T. Chuang, I. Michel Lombera, L. E. Moser and P. M. Melliar-Smith, "Trustworthy distributed search and retrieval over the Internet," *Proceedings of the 2011 Interntational Conference on Internet Computing*, Las Vegas, NV, July 2011, pp. 169–175.

[3] H. Hu, B. Thai and A. Seneviratne, "Supporting Mobile devices in Gnutella file sharing network with mobile agents," *Proceedings of the 8th IEEE Symposium on Computers and Communications*, Kemer-Antalya, Turkey, July 2003.

[4] C. Lindemann and O. P. Waldhorst, "A distributed search service for peer-to-peer file sharing in mobile applications," *Proceedings of the Second International Conference on Peerto-Peer Computing*, Linkoping, Sweden, September 2002, pp. 73–80.

[5] I. Michel Lombera, Y. T. Chuang, P. M. Melliar- Smith and L. E. Moser, "Trustworthy distribution and retrieval of information over HTTP and the Internet," *Proceedings of the International Conference on the Evolving Internet*, LuxembourgCity, Luxembourg, June 2011, pp. 7–13.

[6] P. M. Melliar-Smith, L. E. Moser, I. Michel Lombera and Y. T. Chuang, "iTrust: Trustworthy information publication, search and retrieval," *Proceedings of the 13th International Conference on Distributed Computing and Networking*, Hong Kong, China, January 2012, Lecture Notes in Computer Science 7129, Springer, pp. 351–366.

[7] P. Tiago, N. Kotiainen, M. Vapa, H.Kokkinen and J. K. Nurminen, "Mobile search – Social network search using mobile devices," *Proceedings of the 5th IEEE Consumer Communications and Networking Conference*, Las Vegas, NV, January 2008, pp. 1201–1205.

[8] The Gnutella Development Forum, http: // groups. yahoo. com/ group/ the_gdf/

[9] M. Kamvar and S. Baluja. A large scale study of wireless search behavior: Google mobile search. In*Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 701–709, 2006.

[10] M. Kamvar, M. Kellar, R. Patel, and Y. Xu. Computers iphones and mobile phones, oh my!: a logs-based Comparison of search users on different devices. In *Proceedings of the 18th international conference on World wide web*, pages 801–810, 2009.

[11] J. Wikman, Ferenc Dosa, and Mikko Tarkiainen. Personal Website on a mobile phone. Technical report, Nokia Research Center, 2006.

[12] T. Isdal, M. Piatek, A. Krishnamurthy and T. Anderson,

"Privacy preserving P2P data sharing with OneSwarm," Technical Report UWCSE, Department of Computer Science, University of Washington, 2009.

[13] J. Risson and T. Moors, "Survey of research towards robust peerto-peer networks: Search methods," Technical Report UNSW-EE-P2P- 1-1, University of New South Wales, September 2007, RFC 4981, http://tools.ietf.org/html/rfc4981

[14] B. Yang and H. Garcia-Molina, "Improving search in peer-to-peer networks," *Proceedings of the 22nd IEEE International Conference on Distributed Computing Systems*, Vienna, Austria, July 2002, pp. 5–14

[15] I. Clarke, O. Sandberg, B. Wiley and T. Hong, "Freenet: A distributed anonymous information storage and retrieval system," *Proceedings of the Workshop on Design Issues in Anonymity and Unobservability*, Lecture Notes in Computer Science, Berkeley, CA, July 2000, pp. 46–66.

[16] G. P. Jesi, D. Hales and M. van Steen, "Identifying malicious peers before it's too late: A decentralized secure peer sampling service," *Proc. 1st Intl. Conf. Self-Adaptive and Self- Organizing Systems*, July 2007, pp. 237–246.

[17] J. Webster, R.T. Watson, "Analyzing the Past to Prepare for the Future: Writing a Literature Review", *MIS Quarterly*, Vol. 26, pp. xiii-xxiii, 2002.