DECISION SCIENCES INSTITUTE

A Trustworthy P2P Retrieval for Dense Mobile Networks

Yung-Ting Chuang
Department of Information Management, National Chung Cheng University
Email: ytchuang@mis.ccu.edu.tw

Qian-Wei Wu
Department of Information Management, National Chung Cheng University
Email: william132004@yahoo.com.tw

ABSTRACT

Nowadays, the networked devices have become essential for daily use, and have turned the world into a large-scale wireless networks. The problem behind such wireless networks is the difficulty of information retrieval in dense mobile networks. In this paper, we present a Trustworthy P2P Retrieval (TPR), where it divides the entire network into a number of regions, applies Locality Sensitive Hash (LSH) functions to map metadata to a geographical region, allows nodes to maintain partial membership views, and employs relocation method for mobility resilience. Experimental results shows that TPR could achieve high membership accuracy and mobility resilience.

KEYWORDS: Peer-to-Peer, Mobile Ad-Hoc Network, Wireless Sensor Networks

INTRODUCTION

Nowadays, the networked devices have become essential for daily use, and people can now use any mobile device to access or share information at anytime from across the world. People can now use any mobile device to access or share any information at anytime and from anywhere. The Internet is so powerful that it helps us to share information and perspectives from across the world.

Our trust in the accessibility of information over the Internet currently depends on benign and unbiased administration of centralized search engines and indexes. However, the experience of history shows that we cannot depend on such administrators to remain benign and unbiased forever. Therefore, some decentralized search and retrieval systems such as (Ferreira et al., August 2005; Gnutella; Wong and Guha, February 2008) have previously proposed to ensure the free flow of information over the Internet, preventing anyone to censor or filter information accessed over the Internet. These decentralized search systems give better assurance to the users of the Internet since a small number of administrators cannot prevent them from exchanging their information with others.

Because of the recent technical advancements, the network have slowly shifted to a large-scale wireless network, such as wireless sensor network (WSN) and mobile ad-hoc network (MANET), where there are enormous number of mobile nodes existed over a wide area. Some search systems (Xu et al., 2008; Shen et al., 2014) have therefore proposed in both WSNs and MANET environment. Nevertheless, these systems still generate too much overhead for data distribution and database maintenance, nor guarantee high retrieval rate when in a dense mobile networks.

Therefore, in this paper, we present a Trustworthy P2P Retrieval (TPR) for dense mobile networks. Our system divides the entire network into a number of regions, applies Locality Sensitive Hash (LSH) (Rajaraman and Ullman) functions to map metadata to a geographical region, allows nodes to maintain partial membership views, and employs relocation method for mobility resilience. Experimental results shows even with enormous mobile nodes in the network, our TPR could achieve high membership accuracy and mobility resilience.

RELATED WORK

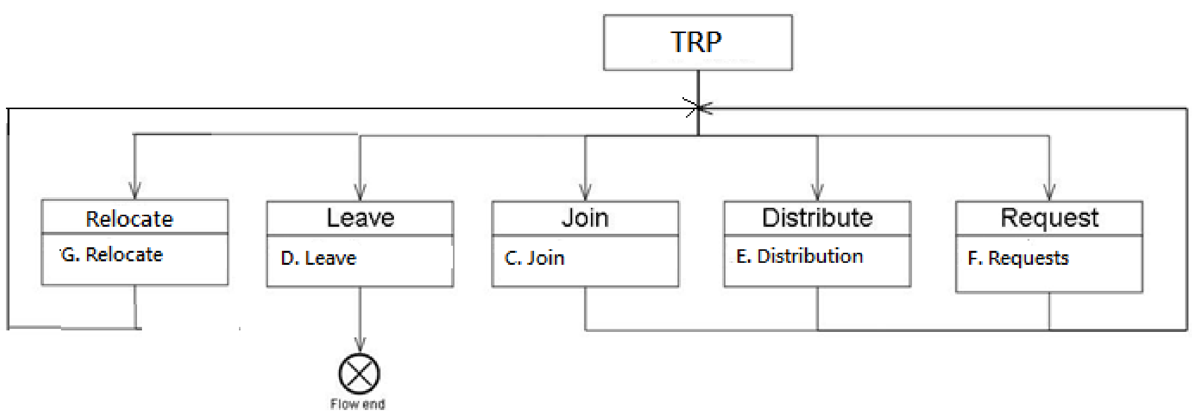
Peer-to-Peer Network

Mischke and Stiller (Mischke and Stiller, 2004) provide comparisons of distributed retrieval services for peer-to-peer networks. The structured approach (Bianchi et al., August 2007; Gupta et al., 2004) requires the nodes to be organized in an overlay network based on distributed hash tables (DHTs), trees, rings, which is efficient but vulnerable to manipulation by untrustworthy administrators. The unstructured approach (Ferreira et al., August 2005; Gnutella; Terpstra et al., 2007) is typically based on gossiping and randomization, where it requires the nodes to find each other by exchanging messages over existing links.

Wireless Sensor Networks (WSN) and Mobile Ad-Hoc Network (MANET)

Wireless Sensor Networks (WSNs) are widely applied to several applications, where these applications use sensors to constantly cooperate with others to exchange or collect data. Similarly, in Mobile Ad Hoc Network (MANET), nodes are allowed to distribute, to retrieve, or to move freely in any direction. Some geographic routing based data search systems (Xu et al., 2008; Shen et al., 2014) use distributed hash table (DHT) data mapping policy to map the file to a geographic location, and apply geographic routing methods (Frey and Stojmenovic, 2010; Shen et al., 2014) to store the file to a node closes to this geographic location for better retrieval rates. Unfortunately, these proposed methods are not energy-efficient since they create too much message costs on data maintenance, nor guarantee high retrieval rates when in a dense mobile network.

Figure 1 TRP System Flow Chart



TPR DESIGN

Our TPR allows nodes to join the network, to distribute metadata, to make requests, to move to other region, to leave the network, or to update its membership. The TPR system flow chart is shown in Fig. 1, and the detailed explanation for each action is described in the following subsections.

A. Types of Nodes

There are seven types of the nodes in our TPR:

1) Source Node: The node that has the actual file.
2) Joining Node: The node that wishes to join to a new region.
3) Leaving Node: The node that wishes to leave the current region.
4) Relay Node: The node that forwards the metadata or requests to other nodes in its region.
5) Requesting Node: The node seeks for the file.
6) Bootstrapping/AP Node: The node that knows all the nodes in its region.
7) Metadata node: The node that has metadata

B. Dividing Network

In TPR, we first consider the scenario where there are some wireless access points (APs) in the network located at the center of each region. Next, we divide the whole network into multiple regions, where each region has an AP located in the center in the region.

C. Joining a new region

The steps of joining a new region are given below:

1) A node joining to a new region first contacts the bootstrapping nodes.
2) A bootstrapping node returns all of its members, along with the boundary information (e.g., source node's URL) of the current region back to the joining node.
3) Joining node saves the information to its database.
4) Joining node distributes a join message to some nodes selected at random in its view.

D. Leaving the current region

The steps of leaving the current region are given below:

1) A node that wishes to leave the current region first determines if it has metadata in its database. If so, the leaving node notifies bootstrapping node about this information, and transfers its metadata to another random node.
2) Leaving node distributes a leaving message to all the nodes in its view, deletes its metadata, and leaves the network.

E. Distribution of Metadata

The steps of distributing metadata are given below:

1) A source node first produces metadata. Then, the source node uses Locality Sensitive Hash (LSH) Function (Rajaraman and Ullman) to calculate the designated location, and sends the metadata to the calculated location.
2) The first node (we later call this node as "relay node") in a mapped region receives this metadata first calculates 2√ nodes to which it needs to distribute this received metadata

based on its current view, and then distribute this metadata to some randomly chosen nodes in its view.
3) The relay node notifies the source node all the nodes that has its metadata.

F. Distribution of Requests

The steps involved when a requesting node distributes its query are given below:
1) A requesting node first produces query. Then, the requesting node similarly uses Locality Sensitive Hash (LSH) Function (Rajaraman and Ullman) to calculate the designated location, and sends the requests to the calculated location.
2) The first node (we later call this node as ”relay node”), that receives such query calculates the 2√view number of nodes to which it needs to distribute this query based on its current view, and then distribute the query to randomly chosen nodes in its view.
3) The nodes that receive such query compares the keywords in the request with the metadata it holds. If it finds a match, the node responds to the relay node with the URLs of the source nodes.
4) The relay node sends the response back to the requesting node.

G. Relocating to other region

The steps involved in relocating to other region are given below:
1) When a requesting node moves out of its current region before it receives the requesting response, it sends a relocation message to the relay node from region where it just left.
2) If a source node moves out of its current region, it first sends a relocation message to a relay node from its source's mapped region, and have this relay node to forward the relocation message to all the nodes that has its metadata.
3) If a node with metadata moves to other region, it first transfers its resources/metadata to a random node, and then informs the source node about this transfer.

CALCULATE RETRIEVAL RATES

The match probability for our TPR is based on the hypergeometric distribution (Feller, 1968), which describes the number of successes in a sequence of random draws from a finite population without replacement. The parameters for the hypergeometric distribution are as follows:

- n: number of nodes in a node's view.
- m: number of nodes that has metadata.
- r: number of nodes that has query.

Thus, the analytical probability of k matches is given as follows:

$$\mathrm{P(k)} = \frac{\binom{m}{k}\binom{n-m}{r-k}}{\binom{n}{r}}, \text{for } m + r \le n \text{ and } k \le \{m, r\} \qquad (1)$$

From (1), the probability P(k≥1) of one or more matches is then derived as:

$$P(k \ge 1) = 1 - P(k = 0) = 1 - \frac{n-m}{n} \cdots \frac{n-m+1-r}{n-r+1} \qquad (2)$$

We have previously shown that when the membership size n distribution of the metadata to m = 2√n nodes and distribution of the requests to r = 2√n nodes results in a probability P(k ≥1) that exceeds 1 $-e^{4} \approx 0.9817$, independent of n. Therefore, we have determined to distribute metadata and requests to $2\sqrt{n}$ in our subsequent experiments.

PERFORMANCE EVALUATION

Experimental Setting

In the simulation program, we consider the network with 900m by 900m area, and divide this area into 9 region, where the area of each region is 100m by 100m. In order to simulate a large wireless networks, we set total of 9000 nodes in the network, and have initially placed these 1000 nodes to the random places in each region. Next, we randomly divided these nodes into three groups: 1) fast speed, 2) medium speed, and 3) slow speed group. The nodes are free to move between these four directions: 1) North; 2) South; 3) West; and 4) East. The ratio of these nodes and its range of randomly chosen speeds are: 1) 30% of fast group with speed randomly chosen between 0.5 and 2.5 m/s; 2) 30% of medium group with speed randomly chosen between 1 and 5 m/s; 3) 40% of fast group with speed randomly chosen between 20 and 30 m/s. Then at each time step, all of the nodes in the network can move freely to other regions, distribute metadata, or distribute requests. Whenever a node hits a boundary of the network during its movement, it will bounce to the other direction and continue move to that direction. We initially chooses 100 source nodes and distribute 100 files to these nodes.

Analytical vs. Simulation

Figure 2 Analytical model versus Simulation results for n=4500, n=9000, n=13500, and n=18000

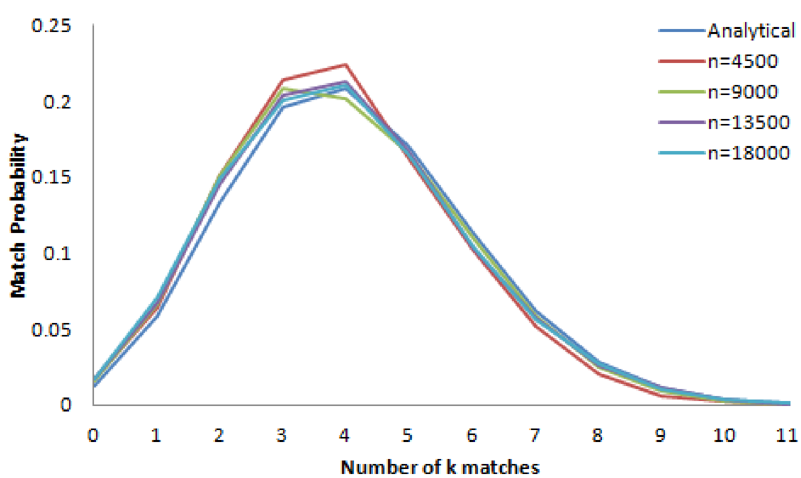


Figure 2 shows the analytical and emulation results for probability p of k matches, where the probability are obtained from Equation 1. This figure shows the analytical curve against emulation curves of n = 4500, n = 9000, n = 13500, and n = 18000 nodes, where each region contains r = m = $2\sqrt{n}$ metadata and requests. As we see from these graphs, the simulation results are all very close to the analytical results for each k matches for all n = 4500, n = 9000, n = 13500, and n = 18000 cases. Hence, we have demonstrated that our TRP retains significant utility in a range of circumstances, even when the number of nodes changes, which might be the circumstances in which the information is most needed

Membership Accuracy (MA)

Figure 3 MA in fast speed, medium speed, and slow speed

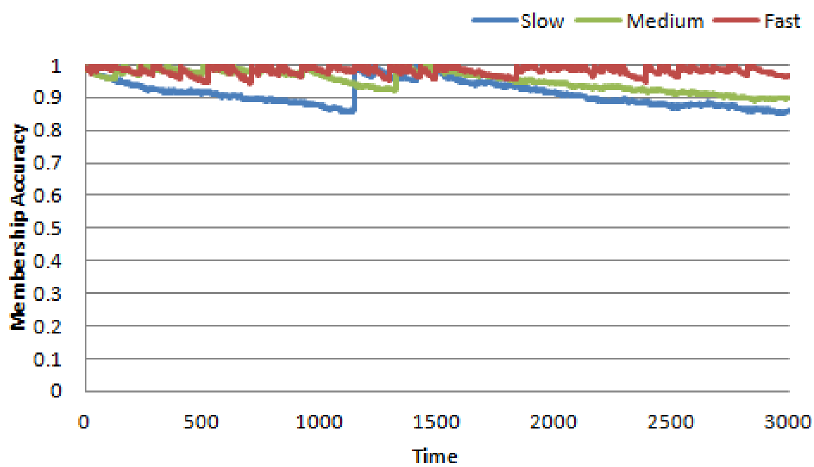


Figure 3 shows the membership accuracy (MA) where the network contains n = 18000 nodes, and is simulated in three different speed, namely, fast speed, medium speed, and slow speed. When the node travels with fast pace, it tends to jump to multiple regions quickly, and therefore it has greater chances to contact with the APs from these regions and get the most up-to-date membership from them. Next, when the node travels with medium speed, it does not move to other regions that frequent comparing to the node moving with fast speed, therefore it does not have that much chances to contact to the APs to get the most up-to-date membership, resulting its membership accuracy is slightly lower than the curve with fast speed. However, the MA produced from the medium speed are still very high, whereas MA are all above MA = 0.9. Lastly, when the node travels with slow speed, it tends not to move to other regions from the beginning to the end. As a result, the membership accuracy curve of slow speed are lower than both the curve of medium and fast speed. Nevertheless, the MA produced from the slow speed are still very high, whereas all the MA are all above MA = 0.9. In summary, from Figure 3, we can confirms that our TCSC is robust and effective since it could achieve high membership accuracy, and is thus mobility resilience.

CONCLUSION

This paper presents a Trustworthy P2P Retrieval, where its goal is to provide robust and effective search in a dense mobile networks. Extensive experiment results show that even with enormous mobile nodes in the network, our TPR could achieve high membership accuracy and mobility resilience.

ACKNOWLEDGMENT

This research is supported by MOST 104-2410-H-194-090-MY2 of Ministry of Science and Technology, Taiwan.

REFERENCES

Andoni, A., Indyk, P., (2006). Near-optimal hashing algorithms for near neighbor problem in high dimensions, in: Proceedings of the Symposium on the Foundations of Computer Science, pp. 459–468.

Bianchi, S., Felber, P., Gradinariu, M., August (2007). Content-based publish/subscribe using distributed r-trees, in: Proceedings of Euro-Par, Rennes, France. pp. 537–548.

Feller, W., (1968). An Introduction to Probability Theory and Its Applications. volume I. John Wiley & Sons.

Ferreira, R.A., Ramanathan, M.K., Awan, A., Grama, A., Jagannathan, S., August (2005). Search with probabilistic guarantees in unstructured peer-to-peer networks, in: Proceedings of 5th IEEE International Conference on Peer-to-Peer Computing, Konstanz, Germany. pp. 165–172.

Frey, H., Stojmenovic, I., (2010). On delivery guarantees and worst-case forwarding bounds of elementary face routing components in ad hoc and sensor networks. Computers, IEEE Transactions on 59, 1224–1238.

Gnutella, . http://en.wikipedia.org/wiki/Gnutella.

Gupta, A., Sahin, O., Agrawal, D., El Abbadi, A., (2004). Meghdoot: Content-based publish/subscribe over P2P networks, in: Proceedings of the 5th ACM/IFIP/USENIX International Conference on Middleware, Toronto, Canada. pp. 254–273.

Mischke, J., Stiller, B., (2004). A methodology for the design of distributed search in P2P middleware. IEEE Network 18, 30–37.

Rajaraman, A., Ullman, J.D., Mining of massive datasets. volume 77.

Shen, H., Li, Z., Chen, K., (2014). A scalable and mobility-resilient data search system for large-scale mobile wireless networks. Parallel and Distributed Systems, IEEE Transactions on 25, 1124–1134.

Terpstra, W.W., Kangasharju, J., Leng, C., Buchmann, A.P., (2007). Bubblestorm: Resilient, probabilistic, and exhaustive peer-to-peer search, in: Proceedings of the ACM Conference on Applications, Technologies, Architectures and Protocols for Computer Communications, Kyoto, Japan. pp. 49–60.

Wong, B., Guha, S., February (2008). Quasar: A probabilistic publish-subscribe system for social networks, in: Proceedings of the 7th International Workshop on Peer-to-Peer Systems, Tampa Bay, FL.

Xu, J., Tang, X., Lee, W.C., (2008). A new storage scheme for approximate location queries in object-tracking sensor networks. Parallel and Distributed Systems, IEEE Transactions on 19, 262–275.