Context-Aware Routing Method for P2P File Sharing Systems over MANET

Taoufik Yeferny
Dept. of Computer Science
Faculty of Sciences of Tunis
MOSIC Research Group
Tunis, Tunisia
Taoufik.Yeferny@it-sudparis.eu

Khedija Arour
Dept. of Computer Science
National Institute of Applied Sciences and Technology of Tunis
URPAH Research Group
Tunis, Tunisia
Khedija.arour@issatm.rnu.tn

Amel Bouzeghoub
Dept. of Computer Science
Telecom SudParis
SAMOVAR CNRS LAB
Paris, France
Amel.Bouzeghoub@it-sudparis.eu

ABSTRACT

Mobile devices have achieved great progress. They allow user to store more audio, video, text and image data. These devices are also equipped with low radio range technology, like Bluetooth and Wi-Fi, etc. By means of the low radio range technology, they can communicate with each other without using communication infrastructure (e.g. Internet network) and form a mobile ad hoc network (MANET). The peers in the MANET are typically powered by batteries which have limited energy reservoir also they are free to move from their locations at anytime. Recently, P2P file sharing systems are deployed over MANET. A challenging problem in these systems is (i) the selection of best peers that share pertinent resources for user's queries and (ii) guarantee that the pertinent peers can be reached in such dynamic and energy-limited environment. To tackle this problem, we propose a context-aware integrated routing method for P2P file sharing systems over MANET. Our method selects the best peers based on the query content and the user's profile. Furthermore, it considers the energy efficiency, peer mobility and peer load factors into the query forwarding process to guarantee that the pertinent peers can be reached.

1. INTRODUCTION

In the last few years, peer-to-peer file sharing systems have emerged as platforms for users to search and share information over the Internet network. There are different kinds of P2P systems architectures that can be roughly classified into structured, unstructured and hybrid architectures [7]. Nowadays, mobile and wireless technology has achieved great progress. Cell phones, PDAs and other handheld devices have larger memory, higher processing capability and richer functionalities. They allow user to store more audio, video, text and image data with handheld devices. These devices are also equipped with low radio range technology, like Bluetooth [1] and Wi-Fi [2], etc. By means of the low radio range technology, they can communicate with each other without using communication infrastructure (e.g. Internet network) and form a mobile ad hoc network (MANET). Mobile peers that are in the transmission range of each other can communicate with their peers directly. To communicate with peers outside the transmission range, messages are propagated across multiple hops in the network. Hence, P2P file sharing systems can be also deployed over MANET. Due the nature of MANET, these systems suffer from tow principles constraints. Firstly, wireless medium is much more dynamic due to peer mobility and the frequent variations in channel quality due to interference and fading [4]. Secondly, mobile devices are battery operated and energy-limited. If a peer is frequently asked to provide or relay files, its battery would be quickly exhausted.

A challenging problem in these systems is (i) the selection of best peers that share pertinent resources for user's queries and (ii) guarantee that the best peers can be reached in such dynamic and energy-limited environment (query routing problem).

In the literature, several works proposed different techniques of query routing in P2P systems on wired scenarios [16]. However, they are not applicable to MANET, since they don't consider the constraints of this network; thus they cannot grantee that the pertinent peers can be reached in such dynamic and energy-limited environment. Hence, energy efficiency and peer mobility are uncompromising factors in the design of query routing P2P file sharing systems over MANET. Several routing methods have been proposed for P2P file sharing systems over MANET. Each of them has its own advantages and limits.

In this paper, we propose a context-aware integrated routing method for P2P file sharing systems over MANET. The key contributions of our proposal are:

- The selection of best peers that share pertinent resources is based on the query content and the user's profile. Indeed, each peer builds a profile of its neighbors. The profile contains the list of the most recent past queries and neighbor that supplied answers for. We defined a similarity function that computes the aggregate similarity of a peer to a given query.

- Our routing method takes into account the constraints of MANET environment to guarantee that the best peers can be reached. Hence, we defined a Link_stablity function that combines the peer mobility and battery energy factors to compute the stability of link between two peers. In addition, we defined a function to guarantee a load balancing and palliate the congestion problem.

The rest of the paper is organized as follows. In Section 2, we present a critical overview of query routing methods in P2P systems over MANET. Section 3 discusses our approach. Section 4 concludes with some proposed direction for further works.

2. RELATED WORK

In the literature there are several points of view of the routing problem in unstructured P2P systems over MANET. Bin Tang et al [15] classify the existing approaches for unstructured P2P systems over MANET into layered or integrated design approaches.

2.1 Layered design approach

The layered design decouples functionalities of the application layer and the network layer, which enables independent development of protocols at the two layers. In this design, routing protocol at application layer (for example, Gnutella) are operated on top of an existing MANET routing protocol at network layer. This design is similar to the approach in the Internet, which layers a P2P protocol on top of the existing IP infrastructure. The routing protocol at the application layer selects the overlay neighbor to forward the search query then it uses an existing routing protocol at the network layer (i.e. DSR [8], AODV [11], DSDV [12], etc) to localize this neighbor. However, due to peer mobility, these overlay neighbors may not reflect the current physical topology of the ad hoc network, and thus may need a multi-hop route to be reached. As a result, each such overlay hop required by Gnutella at the application layer could result in a costly flooding-based route discovery by the multi-hop routing protocol.

2.2 Integrated design approach

MANETs are a limited resource environment where the performance can be more important than portability and separation of functionalities. Hence, integrated design approach is proposed as alternative to layered design approach. In integrated design, routing protocol at the application layer is integrated with a MANET routing protocol at the network layer. In the literature, there are several integrated approaches.

A first idea consists to build an efficient unstructured P2P overlay over MANET. In this overlay connections between mobile peers are closely match the physical topology of the underlying MANET. To find relevant resources for a given search query, flooding technique is used. Andrew et al [9] propose a decentralized and dynamic topology control protocol called $TCP2P$. This protocol allows each peer in MANET to select a set of neighbors according to preference defined function that take into account the energy efficiency, fairness and incentive. After building the network topology, each peer routes the query to its neighbors regardless the query content. Although this protocol virtually controls the macroscopic usage of energy and establishes a stable link, in term of energy efficiency, fairness and incentive, between a source and destination peers. However, it does not compromise the satisfaction of user because queries are flooded regardless their content. Moreover, user's mobility is not considered. $E-UnP2P$ method [14] builds an efficient overlay avoiding redundant links and redundant transmissions while ensuring connectivity among the peers, it introduces a root-peer in the P2P network connecting all other peers. Each peer maintains connection with other closest peers such that it can reach the root-peer. Using the information of its directly connected and 2-hop away (logically) neighbor peers, each peer builds up a minimum-spanning tree to identify far away peers and builds up the overlay closer to the physical network. Thereafter, when a peer wants to retrieve a file, it sends the query to all of its neighbor peers.

A second idea consists to define a progressive search mechanism that allows to route the search queries to the best neighbors. In order to find content, a peer sends a query to its best neighbors, which, in turn, forward the query to their best neighbors and so on, until the query time-to-live (TTL) expires. To select the best neighbors a peer is based on some factors (i.e. Battery energy, signal power, neighbor velocity, neighbor's content, etc.). In Data Dissemination in Mobile P2P Networks [13] each peer maintains a global description of other peers' content (content synopses), and utilizes that synopses in order to route queries more efficiently. A peer that receives a query searches in its local collection. If it is not possible to answer this query, it calculates a score of peers from the global index then propagates the query to the peers, which have the greatest score. If there is no match between the query and the content synopses, the query is forwarded to a set of random neighbors. Content synopses must be updated whenever an object is added, deleted or its contents have changed, which generates a lot of message traffic and load charge of peers. Furthermore, this method does not consider the mobility and the energy factors. In enhancing peer-to-peer content discovery techniques over mobile ad hoc networks [4], the authors propose to improve the unstructured P2P over MANET using Gossiping [5] approach of MANET routing protocol. This is achieved by computing the forwarding probability of a link based on the network load. Indeed, if a peer want to send a query it computes the forwarding probability for a given neighbor based on it computational load (the queue utilization of the neighbor) then forwards the query to neighbors with lower load. Significantly, this probability allows sending more messages to neighbors with lower load, while less messages are sent to saturated peers. This method grantees a load balancing between peers but it floods the query regardless its content. Furthermore, it does not consider the mobility and the energy factors.

3. PROPOSED APPROACH

In this paper, we consider the pure peer-to-peer systems (Gnutella system) over MANET and we propose new techniques that are more efficient than the Gnutella search. Flooding is a fundamental file search operation in pure peer-to-peer (P2P) file sharing systems, in which a peer starts the file search procedure by broadcasting a query to a random set of its neighbors, who continue to propagate it with the same manner to their neighbors. This procedure repeats until a time-to-live (TTL) counter is decremented to 0. If a contacted peer has pertinent resources for the search query, it sends a query hit message to the source peer. The query hit message is routed back to the source peer through the reverse path of the query message. This solution generates a very large number of messages and it cannot quickly locate the request resource. Furthermore, query hits may not be received by the source peer due to the peer mobility and energy limitation. Indeed, peers in the reverse path of the query message may turn off or move out of the network.

In our approach, to find relevant resources for a specific user query a peer sends the query to its best neighbors, which, in turn, forward the query to their best neighbors and so on, until the query time-to-live (TTL) expires. Neighboring peers refer to those peers which are within the transmission range of the forwarding peer.

Assume that a peer p_i which has a set N of neighboring peers. Now the question is "How we determine the best k neighbors?", k is a user defined threshold and $k \leq N$. In the following, we present the different context features considered to select the best k neighbors for a given query q. Thereafter, we present our neighbors selection algorithm.

3.1 Context features

3.1.1 User's profile and query content

We consider the query content to help the querying peer to find the most relevant answers to its query quickly and efficiency. To achieve this, a peer estimates, for each query, which of its neighbors are more likely to reply to this query, and propagates the query message to those peers. To determine the pertinent neighbors, we compute the similarity between the query and each neighbors. Hence, each peer maintains a profile for each of its peers. The profile contains the list of the most recent past queries, that the specific peer that provided the answer for. Although logically we consider each profile to be a distinct list of queries, we use a single Queries table with (Query-peer) entries that keeps the most recent queries the peer has recorded.

For each query it receives, the receiver peer uses the profiles of its peers to find which ones are more likely to have documents that are relevant to the query. To compute the similarity, the receiver peer compares the query to previously seen queries and finds the most similar ones in the repository. To find the similarity between the queries, it uses the cosine similarity [10]. Thereafter, we compute an aggregate similarity of a peer to a given query. The aggregate similarity of peer n_j to query q that peer p_i computes is:

$$Psim_{p_i}(n_j, q) = \sum_{q_k \ was \ answered \ by \ n_j} Cosine(q_k, q) \tag{1}$$

3.1.2 Link stability

We defined a Link_stablity function that combines the peer mobility and battery energy factors to compute the stability of link between two peers. Before describing our function, we present two principle metrics. The first one takes into account the peer mobility factor to predict lifetime of a link between tow peers. The second one predicts the remaining battery energy of a given peer.

Peer mobility

In MANET environment peers are free to move from their location at anytime. In our approach we consider this important factor, thus we predict the lifetime of a link between the forwarding peer and its neighbors. To predict the lifetime of a link $i - j$ between the peer p_i and its neighbor $n_j \in N$ we are based on the RABR protocol [3] functions. This protocol operates at network layer. It predicts the lifetime of a link $i-j$ using a metric called the "affinity" a_{ij} and it is a measure of the time taken by peer n_j to move out of the range of peer p_i. Peers exchange beacons periodically. Peer p_i periodically samples, for every $\triangle_t$ time units, the strength of the beacon signals received from peer n_j. The rate of change of signal strength is given as:

$$\Delta(S_{ij}) = \frac{S_{ij}(current) - S_{ij}(prev)}{\Delta_t} \tag{2}$$

The above quantity is then averaged over the last few samples to obtain $\Delta(S_{ij}(ave))$. Hence, based on this metric we define a link lifetime measure $Lifetime(i - j)$, which computes the time taken by peer n_j to move out of the range of peer p_i, as follows:

$$Lifetime(i-j) = \begin{cases} \Delta(S_{ij}(ave)) & if \quad \Delta(S_{ij}(ave)) \geq 0 \\ \frac{S_{thresh} - S_{ij}(current)}{\Delta(S_{ij}(ave))} & otherwise \end{cases} \tag{3}$$

Battery energy

The calculation of energy level is important to determine the battery level of every peer during active data transmission. We assume that the battery level of a wireless peer decreased when the peer initiated data transmission or when the peer forwards packets. A peer gets killed (disconnected) if the battery power finishes. To predict the remaining battery power we assume that the transmit power is fixed. As in [6], energy required for each operation like receive, transmit, broadcast, discard on a packet is given by:

$$E(packet) = b \times (packet\_size) + c \tag{4}$$

Coefficient b denotes the packet size dependent energy consumption whereas c is a fixed cost that accounts for acquiring the channel and for MAC layer control negotiation. Each peer has to maintain a table to record the remaining energy of its neighboring peer. This data is used by the peer to predict the remaining energy of the neighboring peer n_j. Assume the remaining energy, of a neighbor peer at time $t1$ and t2 are $rengy1(n_j)$ and $rengy2(n_j)$. The prediction of remaining energy of this peer at time t is given by

$$\begin{aligned} rengy(n_j) &= rengy2(n_j) \\ &+[(rengy2(n_j) - rengy1(n_j))/(t2 - t1)] \times (t - t2) \end{aligned} \tag{5}$$

Every peer has to calculate the $rengy$ by itself and sends it to its neighbors.

We combine the lifetime and the remaining energy metrics to define our function Link_stability. This metric calculates the time taken by peer n_j to move out of the range of peer p_i or the battery power of n_j finishes. The $Link\_stability(i-j)$ of a link $i-j$ between the peer p_i and its neighbor n_j is computed as follows:

$$Link\_stability(i-j) = Min(rengy(n_j), Lifetime(i-j)) \quad (6)$$

Where, $rengy(n_j)$ is the remaining energy of the neighbor n_j.

3.1.3 Peer load

A vital part of the optimal network is the load balancing. For instance, job completion becomes complex, if huge load is given to the peers with less processing capabilities. There is a possibility of load imbalance due to that the computing/processing power of the systems are non-uniform few peers may be idle and few will be overloaded. A peer which has high processing power finishes its own work quickly and is estimated to have less or no load at all most of the time. However, if we send queries only to peers that have hight processing capabilities data packets will take routes that could introduce more delay hence increasing latency. With proper ways to transferring traffic load onto routes that are relatively less congested can result in overall better throughput and reduced latency. An important parameter indicates the line congestion is the queue utilization of the neighbor (i.e. Number of packets waiting in queue), a high count indicates line congestion. We define a Peer_Load function based on the CPU capabilities and the queue utilization of the neighbor. The Peer_Load of a neighbor n_j is calculated as follows:

$$Peer\_Load(n_j) = cpu \times (1-u) \quad (7)$$

where cpu is the processing power and $0 \leq u \leq 1$ is the queue utilization of the neighbor n_i. This function allows to send more messages to neighbors with lower load, while less messages are sent to saturated peers.

3.2 Neighbors selection algorithm

To select its K best neighbors, the forwarding peer p_i ranks its neighbors according to a Preference function that we define. Thereafter, it selects the first k neighbors, which have the greatest score. Our Preference function computes the score of each neighbor n_j for a given query q, as a weighted arithmetic sum of Link_stability, Peer_Load and Psim metrics:

$$\begin{aligned} Pref(n_j) &= \alpha 1 \times Link\_stability(i-j) + \alpha 2 \times Peer\_Load(n_j) \\ &+ \alpha 3 \times Psim_{P_i}(n_j, q) \end{aligned} \quad (8)$$

where $\alpha 1$, $\alpha 2$ and $\alpha 3$ represent the relative importance of these three metrics.

4. CONCLUSION AND FUTURE WORKS

We have presented a novel context-aware integrated routing method for P2P file sharing systems over MANET. Our method selects the best peers based on the query content and the user's profile. Furthermore, it considers the energy efficiency, peer mobility and peer load factor into the query forwarding process to guarantee that the pertinent peers can be reached. As the future work, we plan to implement the proposed method and evaluate its retrieval effectiveness and routing efficiency.

5. REFERENCES

[1] Bluetooth. http://simple.wikipedia.org/wiki/Bluetooth, March, 2012.

[2] Wifi. http://simple.wikipedia.org/wiki/Wifi, March, 2012.

[3] S. Agarwal, A. Ahuja, J. P. Singh, and R. Shorey. Route-lifetime assessment based routing (rabr) protocol for mobile ad-hoc networks. In *ICC (3)*, pages 1697–1701, 2000.

[4] D. N. da Hora, D. F. Macedo, L. B. Oliveira, I. G. Siqueira, A. A. F. Loureiro, J. M. Nogueira, and G. Pujolle. Enhancing peer-to-peer content discovery techniques over mobile ad hoc networks. *Comput. Commun.*, 32(13-14):1445–1459, Aug. 2009.

[5] M. Dietzfelbinger. Gossiping and broadcasting versus computing functions in networks. In *STACS 97*, volume 1200, pages 189–200. Springer Berlin / Heidelberg, 1997.

[6] N. Gupta, K. Bafna, and N. Gupta. 2011 international conference on information and network technology. In *2011 Seventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, volume 4. IACSIT Press, 2011.

[7] H. Jin, X. Ning, H. Chen, and Z. Yin. Efficient query routing for information retrieval in semantic overlays. In *Proceedings of the 21st Annual ACM Symposium on Applied Computing*, pages 23–27, Dijon, France, April 23–27 2006. ACM Press.

[8] D. B. Johnson, D. A. Maltz, and J. Broch. Dsr: The dynamic source routing protocol for multi-hop wireless ad hoc networks. In *In Ad Hoc Networking, edited by Charles E. Perkins, Chapter 5*, pages 139–172. Addison-Wesley, 2001.

[9] A.-H. Leung and Y.-K. Kwok. On topology control of wireless peer-to-peer file sharing networks: energy efficiency, fairness and incentive. In *World of Wireless Mobile and Multimedia Networks, 2005. WoWMoM 2005. Sixth IEEE International Symposium on a*, pages 318 – 323, june 2005.

[10] J. Makhoul, F. Kubala, R. Schwartz, and R. Weischedel. Performance measures for information extraction. In *Proceedings of DARPA Broadcast News Workshop*, pages 249–252, Herndon, VA, February 1999.

[11] C. Perkins, E. Belding-Royer, and S. Das. Ad hoc on-demand distance vector (aodv) routing, 2003.

[12] C. E. Perkins and P. Bhagwat. Highly dynamic destination-sequenced distance-vector routing (dsdv) for mobile computers. In *Proceedings of the conference on Communications architectures, protocols and applications*, SIGCOMM '94, pages 234–244, New York, NY, USA, 1994. ACM.

[13] T. Repantis and V. Kalogeraki. Data dissemination in mobile peer-to-peer networks. In *Proceedings of the 6th international conference on Mobile data management*, MDM '05, pages 211–219, New York, NY, USA, 2005. ACM.

[14] N. Shah and D. Qian. An efficient unstructured p2p overlay over manet using underlying proactive routing. In *2011 Seventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN)*, pages 248 –255, dec 2011.

[15] B. Tang, Z. Zhou, A. Kashyap, and T. cker Chiueh. An integrated approach for p2p file sharing on multi-hop wireless networks. In *WiMob (3)*, pages 268–274, 2005.

[16] T. Yeferny and K. Arour. Learningpeerselection: A query routing approach for information retrieval in p2p systems. In *International Conference on Internet and Web Applications and Services*, pages 235–241, Barcelona, Spain, May 09-15 2010. IEEE Computer Society.