Computer Networks 57 (2013) 1853–1862

Contents lists available at SciVerse ScienceDirect

Computer Networks

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

Solution for the broadcasting in the Kademlia peer-to-peer overlay

Zoltán Czirkos *, Gábor Hosszú

Department of Electron Devices, Budapest University of Technology and Economics, Magyar tudósok körútja 2, Building Q, Section B, 3rd Floor, Budapest H-1117, Hungary

ARTICLE INFO

Article history:
Received 8 February 2012
Received in revised form 29 December 2012
Accepted 28 February 2013
Available online 25 March 2013

Keywords:
Peer-to-peer
Distributed hash table
Kademlia
Broadcast

ABSTRACT

Structured peer-to-peer networks are capable of fast and efficient lookup operations as a distributed hash table. The topology of these networks makes it possible to send broadcast messages among nodes, either for the purpose of providing a complex query service for participants, or to disseminate information valuable for all nodes. In this article a broadcast algorithm for the Kademlia XOR topology is presented. The algorithm, which was developed specifically for Kademlia, uses replication mechanisms similar to that of the storage and retrieval service of this overlay topology. This allows for increased reliability and speed of the broadcast, and also efficient operation, as the routing table for lookups are already available and can be used. An analytical model is presented, which can be used to calculate the required level of replication for any desired reliability at runtime, and is validated with simulation as well.

1. Introduction

The purpose of developing structured P2P overlays was to reduce the traffic generated by data lookup requests [1], which were based on flooding mechanisms in earlier overlays like Gnutella [2]. However, some P2P applications require a broadcast messaging service. Broadcast messages are used to disseminate information valuable for every peer. Types of such information include system-wide configuration parameters or global messages in instant messaging services. They can also be used to implement partial keyword searches or so-called blind searches [3]. These are not much different from a broadcast: messages are forwarded like in a broadcast, but forwarding is stopped once the request can be fulfilled.

The topologies of structured P2P networks are organized in such a way, that the diameter of the overlay, i.e. the number of hops between any two nodes in the topology, is small. By using the already present connections, broadcast messages can be delivered to nodes quickly, usually in $O(\log N)$ time, where N is the number of nodes in the overlay [4,5]. Problems arise when nodes or network links fail, or malicious nodes join the network [6]. These hinder the correct functioning of communication, and even result in a significant number of nodes not receiving the message at all [7,8].

Most of the broadcast algorithms presented in the literature are developed for Chord [9] or CAN [10]. In this paper, we present a broadcast algorithm for the Kademlia network, which only uses entries in the routing tables of Kademlia nodes, and therefore it requires no preliminary steps or additional address space partitioning for the broadcast to take place. We deduce formulae to estimate the coverage correctness (ratio of nodes receiving the message) of the algorithm in the function of network packet loss ratio and node reliability. The equations presented in this article allow nodes to set the parameters of the algorithm during run-time, using data available through measurements of their usual traffic and investigation of their routing tables.

* Corresponding author. Tel.: +36 1 463 2797; fax: +36 1 463 2973.
E-mail addresses: czirkos@eet.bme.hu (Z. Czirkos), hosszu@eet.bme.hu (G. Hosszú).

1389-1286/$ - see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.comnet.2013.02.021

In order to enhance reliability and speed up the broadcast, we employ a replication method, that is suited especially for the Kademlia network. The effect of replication on reliability is also examined and an equation is derived, which can be used to calculate the required level of replication for any desired level of coverage by nodes at runtime. All analytical results are validated with simulation as well, and a small scale experiment is used to measure the efficiency of different methods aimed at shortening the time interval needed to carry out the full broadcast in an overlay network.

The rest of the paper is organized as follows. In Section 2, related works of similar methods are overviewed. Section 3.1 presents a transformation of the Kademlia overlay, which is used in the proof of correctness of the broadcast algorithm. Section 3 discusses the novel broadcast algorithm with replication. Analytical and simulation results are presented and compared in Section 4. Research is concluded and future research goals are set in Section 5.

2. Related work

Structured P2P networks are application level overlay networks utilizing a predefined topology. The overlay network is generally presented to the application as a distributed hash table service (DHT), storing $\langle key; value \rangle$ pairs and allowing the reliable retention and quick retrieval of any *value*, for which the associated *key* is precisely known. The mapping of keys to nodes makes it easy to lookup any $\langle key; value \rangle$ pair in the overlay. In turn, it makes it practically impossible to retrieve data if the key is not precisely known. In order to implement such kind of queries, usually broadcast messaging is used.

The broadcast algorithms of P2P networks generally use a deterministic method: partitioning the address space or probabilistic method: epidemic style communication [3]. Each have advantages, and by the combination of the two approaches, the number of messages does not increase significantly.

2.1. Deterministic broadcast algorithms

Deterministic broadcast algorithms are based on selecting *responsible nodes* for specific intervals of NodeID address space. By properly partitioning the address space, the broadcast can be sent using $N-1$ messages and in $O(\log N)$ time, where N is the size of the overlay. If the tree formed by the path of the messages is higher than that, the broadcast will be slower [4,11]. Churn (nodes frequently quitting, failing or maybe joining [12]) is handled by these algorithms with various levels of reliability and extra cost [9].

The *Efficient Broadcast* algorithm partitions the address space of Chord into equally sized intervals [5]. For example, in an overlay, which uses the address space $[0,8)$, the node with the address 0 initiating the broadcast will divide the address space into the intervals $[0,4)$ and $[4,8)$. In the latter one, node 4 is responsible for sending the message, while in the former, node 0 is responsible. So it refines the interval to $[0,2)$ and $[2,4)$ again, and node 4 does the same for its own selected interval. When experiencing packet losses or incorrect routing tables due to churn, the reliability of this algorithm drops rapidly [3]. Also, nonlinear network topologies like CAN [13] or Kademlia do not allow arbitrary selection of address space partitions.

The algorithm named *Self-correcting Broadcast* aims to extend the *Efficient Broadcast* to handle outdated entries in the routing table, by using a correction on use technique [14]. Address space partitions, or so-called intervals are selected, and the responsibility of a node is delegated to the node of the next interval upon failure.

Reliability can be improved by instructing selected nodes to provide an acknowledgment upon the successful completion of the broadcast [7]. This is a recursive mechanism: a responsible node will send the acknowledgment back to the one it is selected by, if it has received *all* the acknowledgments from the nodes it selected for its smaller intervals. After a predefined timeout, new responsible nodes are selected. While this solution is applicable for routing table deficiencies, it fails to provide adequate correctness when malicious nodes join the overlay, which reply to broadcast requests with false positive acknowledgments.

Partitioning of the address space can be carried out by address prefixes as well [15]. An algorithm presented in the cited article is implemented for Scribe on Pastry [16]. However, the arbitrary selection of responsible nodes requires additional lookup requests for the broadcast.

2.2. Probabilistic broadcast algorithms

Probabilistic algorithms use *epidemic-style* communication to cover the address space of overlays. In the algorithm *Efficient Broadcast in P2P grids*, nodes randomly select more neighbors for sending them the message as a duplicate [8]. This ensures reliability in case of nodes failing or quitting the overlay. The random selection, however, does not take overlay topology of Chord into consideration.

2.3. Other uses of broadcast-like communication

Algorithms based on selecting responsible nodes are generally faster than those working with random selection. However, the latter are more reliable when network errors are present. Efficiency can be improved if the aim of the broadcast is to provide *blind search*. For these, forwarding the message to neighbors is stopped if the search request is fulfilled [17], or the flooding can also be stopped by assigning a TTL (time to live) field to messages and decrementing it with every hop [18].

Several methods presented in literature implement multicast rather than broadcast [10]. The multicast tree in the communication can be the spanning tree of already present connections in the P2P network. For an overview of application level multicast algorithms, the reader is referred to [19].

2.4. The Kademlia structured peer-to-peer network

Kademlia is a P2P DHT routing protocol, which organizes its nodes to a binary tree. (For an in-depth discussion of its internal semantics, please refer to [20].) It is