

Pear2Pear (On Wifi): A Data Sharing Protocol Over Wifi through a Peer to Peer Network

Steve Tueno, Romeo Tabue, Forentin Jiechieu, Yacynth Ndonga, Billy Zafack, Audric Feuyan, Jonas Atibita, Alex Djouontse and Rodrigue Mbinkeu

National Advanced School of Engineering, Yaoundé, Cameroon

Abstract

A peer-to-peer system is a distributed system in which equal nodes (in terms of role and usage) exchange information and services directly. This paper describes a distributed peer-to-peer protocol that allows wifi-enabled smart devices (especially Android smart-phones) to exchange data using only wifi. The protocol is designed to allow the automatic establishment of a distributed peer-to-peer network of any size without geography constraint. It is applied to file sharing between devices, but can easily be adapted to support any data sharing. The protocol defines two layers: (1) a *kernel layer* responsible for creating, routing, establishing and maintaining links between nodes (peer), addressing a node and adding and removing nodes; and an (2) application layer to support data sharing (file sharing in this case). The structure of the peer-to-peer network is hybrid. Regarding the file sharing use case, the application allows a node to (i) search for a file in the catalog of its subnet that is held by the *root* node, and (ii) download a file: if the file is not in the subnet, the root node delegates a node to make vouchers in neighboring subnetworks to get the file and make it available. A proof of concept was made on Android.

Keywords: Distributed System, Peer-to-Peer, Network, File Sharing, Wifi, Routing

Introduction

Since the beginning of the Internet, the client-server model was the reference model for the provision of resources. In this model, the system relies on a dedicated server that centralizes and maintains all resources and services. Therefore, the increase in the number of users requires a greater investment of service providers. It is indeed required to guarantee the availability of resources and services, despite the large number of simultaneous requests. This requires significant resources and imposes software and

Email address: stuenofotso@gmail.com, steve.jeffrey.tueno.fotso@usherbrooke.ca
(Steve Tueno, Romeo Tabue, Forentin Jiechieu, Yacynth Ndonga, Billy Zafack, Audric Feuyan, Jonas Atibita, Alex Djouontse and Rodrigue Mbinkeu)

hardware constraints, which make such systems very expensive and difficult to secure and scale because of the single point of failure. However, the increase in the number of users also implies an increase in cumulative resources (computing and storage), hence the peer-to-peer paradigm.

The peer-to-peer paradigm consists of connecting users in order to pool resources and distribute tasks [1]. A peer-to-peer system is thus a distributed system of connected peers, consumers and service providers. Systems based on this paradigm are implemented as logical networks, connecting participants over physical networks. The work presented in this paper reports on the design of an hybrid peer-to-peer protocol in which peers are wifi-enabled mobiles not connected to Internet.

1. Background

1.1. The Peer-to-Peer Paradigm

The peer-to-peer paradigm (P2P) has been focused from the outset on the management of large amounts of data spread over a very large scale. Although it has been used so far mainly for file sharing purposes, many research projects are interested in other types of applications including collaboration, distributed computing, sharing and distribution of content.

1.1.1. Definition

A peer-to-peer system is a distributed system in which equal nodes (in terms of role and usage) directly exchange information and services [1]. Behind this academic definition is the reality of the context of use of these systems. Indeed, it is the role of peers to put and maintain each peer-to-peer system in place. There is no central entity that controls the system. This entails a number of advantages, but also constraints. For benefits, increasing the number of participants also increases the resources available while distributing the load among participants. It is therefore theoretically possible to scale to a relatively low cost. In addition, the absence of a central control entity gives the system greater robustness, since the availability of resources is not directly related to a particular peer [2]. For constraints, the absence of a central entity requires the establishment of a self-organizing mechanism. Peers must connect to or disconnect from the network without affecting the availability of resources. In addition, peer-to-peer systems must manage a network of highly heterogeneous nodes (hardware architecture: processor, memory and connection type: network interface, connection rate, ...).

1.1.2. Different centralization levels

P2P systems can be classified into three major families according to the level of centralization that results in a more or less important distribution of tasks to be accomplished. These are centralized, distributed and decentralized (hybrid) P2P models (Fig. 1).

The *centralized model* is at the border of the P2P model, because it is based on a dedicated server that centralizes and maintains the body of knowledge of peers, the resources always being hosted on peers. In this model, the server is a location entity and contains no resources. Only the retrieval of resources is decentralized. The advantage of

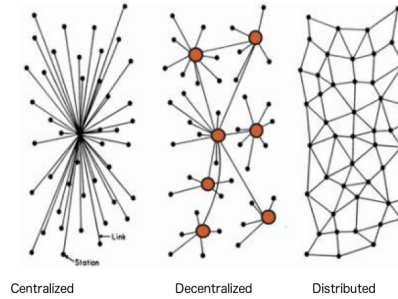


Figure 1: P2P systems classification

this model is that it allows a comprehensive search, fast resource localization and simple management thanks to the central entity. The major disadvantage is that the system operation is based solely on the central entity. The latter must be able to support a large number of queries and perform a lot of research. A well-known example falling into this category is that of the file-sharing system *MP3 Napster* [3].

Distributed P2P systems try to distribute all the system functions among the peers, namely: searching, routing and retrieving objects in the logical network. Peers need to organize to form an effective dynamic architecture connecting them to each other. There are two approaches to this: build a random graph and rely on random explorations/searches of the network, or build a structured graph in which an efficient routing is gradually supported [4]. A well-known example falling into this category is that of *Gnutella v0.4* [5]. Gnutella in its first version v0.4, proposes to connect the nodes in a random way and uses the flood for its searches in the logical network thus formed. For a better use of resources, the following version, *Gnutella v0.6*, proposes to organize the logical network according to a decentralized architecture on two levels.

The *decentralized model*, also called hybrid model, adds an hierarchy to the centralized model. This model relies on super node interconnections at the top of the hierarchy, according to the distributed model. Each leaf node is attached to one or more super nodes. A super node manages a set of leaf nodes. Objects shared by a leaf node are saved to the super node responsible for that node. When a leaf node looks for an object, it sends its request to its super node. This one then carries out the search among the objects of the nodes which are attached to it, or even the neighboring super nodes if necessary. A well-known example falling into this category is that of *FastTrack* [6]: 1 super node for 100 leaf nodes.

Existing peer-to-peer systems rely on Internet for data sharing. Our goal is to design a peer-to-peer system able to circumvent the need for an active Internet connexion by relying solely on device wifi network interfaces.

2. The Pear2Pear On Wifi Protocol

The protocol is designed according to the decentralized or hybrid model. A proof of concept was made on Android (Fig. 4).

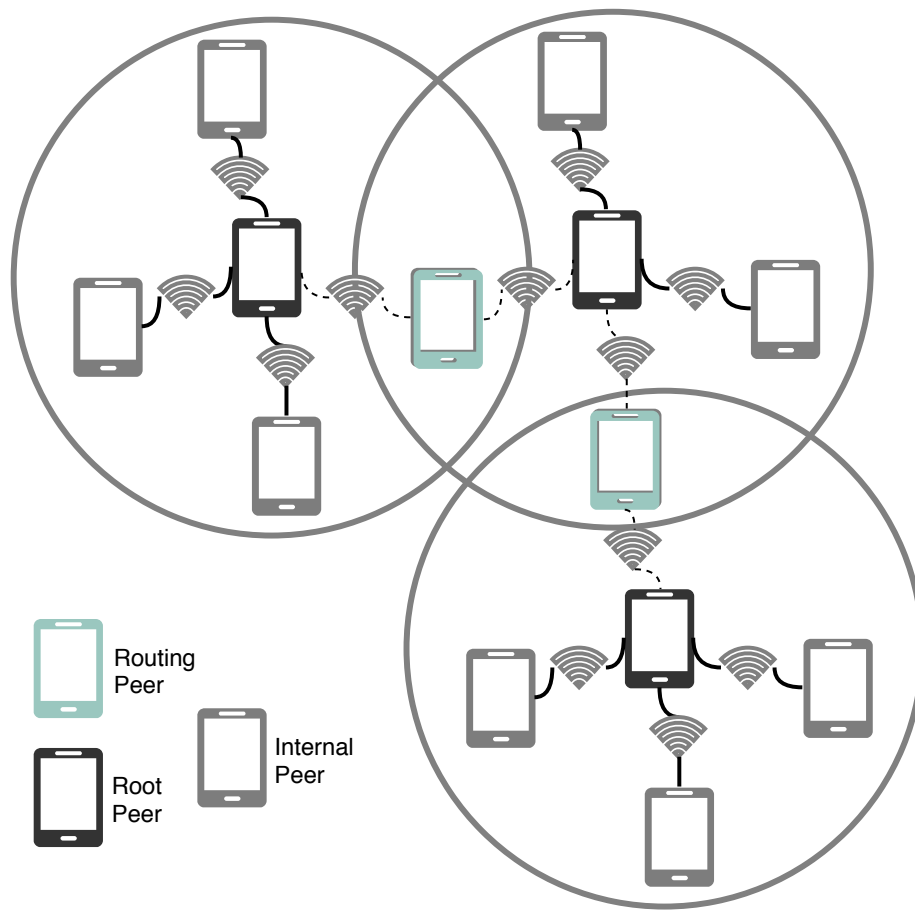


Figure 2: Architectural View of Pear2Pear On Wifi Subnetworks

2.1. Problems / Contributions

Several problems were identified as soon as this project started:

- Establishing and maintaining links between nodes
- Adding and removing a node from the network (does a node exit the network or is it removed ?)
- Routing (how to choose the best route for a given packet ?)
- Node addressing
- Management of available resources (battery, memory,...)
- File sharing protocol

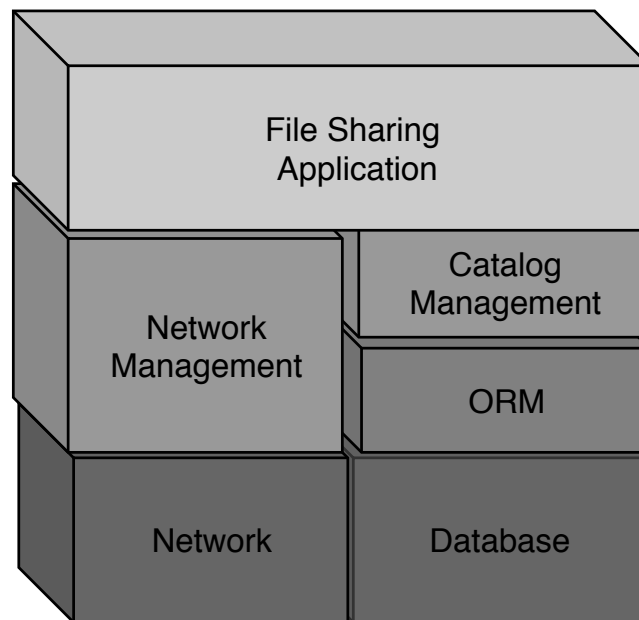


Figure 3: Building Blocks of a Pear2Pear On Wifi Subnetwork

- How to identify a file on the network ?
- Uniqueness, completeness, error handling for a given file
- Managing duplicate files (if a peer wants to download a file it already has)
- Storing files (file block, complete file)
- Parallelism management in file sharing
- Management of identical files but with different names

2.2. Algorithms

To solve the identified problems, several algorithms have been set up. This section presents some among the most relevant ones.

2.2.1. Subnet Creation

A wifi-enabled device wants to connect:

1. It scans surrounding wifi networks for available networks.
 - If it identifies a network, it connects to it; otherwise, it creates a network (hotspot) and becomes its *root*. The ssid of each network is defined so as to make it unique and quickly identifiable by other peers. Each network is protected. To access it, a complex computation function allows each device equipped with *Pear2Pear* to compute the passphrase from the ssid.



(a)



(b)

Figure 4: Overview of the Pear2Pear On Wifi Android Proof of Concept

- If it identifies a network and cannot connect to it after a while, it creates a network (hotspot) and becomes its *root*. Indeed, some operating systems, including Android, limit the number of devices likely to join the hotspot. This number can also be limited by the protocol in order to optimize the use of resources of the network root (battery, memory, ...).

Regarding the management of peer departures, if a peer notifies its root of its departure, it is removed from catalogs after the expiration of a certain countdown. Moreover, the root must make pings to scan for available peers and deals with silent leaves.

2.2.2. *The Network File Catalog*

Each subnet is associated with a file catalog. The catalog identifies for each file the proprietary device and its subnet. Each file is represented by its original name and an unique identifier, related to its content, obtained through the computation of the hash of this content. This makes it possible to manage the duplication of files and to optimize the acquisition of data.

1. The root device initializes the catalog of its subnet, when it is created, with the list of its shared files.
2. Each peer that connects to it, communicates the list of its available files.

Each non-root device in a subnet periodically scans other accessible subnets and informs its root to allow it to build a subnet catalog. The subnet catalog identifies other accessible subnets and for each, the list of peers that can access it.

2.2.3. *File Search and File Management*

Note: when a file search fails, a special frame must (or can) be emitted (once or periodically) in order to invite the peers to make this file available as far as possible.

To allow inter-network searches, each root must periodically designate, for each neighboring subnet, a peer that will switch to that subnet and retrieve the file catalog to bring it back to its original subnet and make it available to the root. This will enrich the network file catalog of the subnet which will support all file requests. A subnet that is no longer accessible is automatically removed from the catalog after the expiration of a certain countdown. Each time a file is added or deleted, a notification is sent to the root to update the catalog.

The choice of the peer which will retrieve the catalog of another subnet, among all peers having access to this subnet, must be done by an algorithm ensuring optimal use of peer resources. We choosed *round robin* for our proof of concept.

A catalog makes it possible to reference for each subnet whose file is referenced the number of jumps required to reach it and the nearest network that serves as gateway to it. This supports the computation of the shortest path to retrieve a file (also taking into account file duplicates that enable distribute file retrieval).

2.2.4. *File Retrieval*

A peer that searches for an available file gets the unique identifier of the file.

A Peer to a Peer of the Same Subnet. : A peer $P1$ of the subnet x wants to retrieve the file f present in the shared folder of peer $P2$ of the subnet x :

1. $P1$ sends a request to download the file f specifying the unique identifier of f in the network file catalog.
2. The root of subnet x identifies the peer $P2$ as the closest (same subnet) peer that has a copy of the requested file and sends to $P1$ the network address of $P2$.
3. $P1$ proceeds directly to download the file from $P2$.

If multiple peers of x have a copy of the file, the root makes all their addresses available to $P1$ so that it can proceed with a distributed download of the blocks of f .

A Peer to a Peer of a Different Subnet. : The peer $P1$ of the subnet $S1$ wants to retrieve the file f of a peer of the subnet S_n , where n is the number of hops separating the 2 subnets:

1. $P1$ sends a request to download the file f specifying the unique identifier of f in the network file catalog.
2. The root of $S1$ identifies the closest subnet that owns the file, based on available file copies and number of hops between networks, as S_n . It also identifies the peer $P1'$ of $S1$ as the most likely to go to $S2$, the gateway from $S1$ to S_n . Therefore, it sends $P1'$ to $S2$ with the file request.
3. $P1'$, once in $S2$, checks if $S2 == S_n$.
4. If $S2 == S_n$, it sends the file request to the root of $S2$ to receive the network address of the peer $P2$ of $S2$ hosting the file.
5. $P1'$ proceeds directly to download the file from $P2$.
6. If $S2 \neq S_n$, the protocol resumes in step 1 with $P1 = P1'$ and $S1 = S2$.
7. $P1'$ jumps back to $S1$ and directly delivers the file to $P1$ since it has its network address.

If multiple peers of $S2$ have a copy of the file, the root makes all their addresses available to $P1'$ so that it can proceed with a distributed download of the blocks of f . In addition, if the file is very large, several peers of $S1$ having access to $S2$ can jump to $S2$ to request specific blocks of f and return to distributively deliver it to $P1$.

References

- [1] B. Yang, H. Garcia-Molina, Comparing hybrid peer-to-peer systems, in: P. M. G. Apers, P. Atzeni, S. Ceri, S. Paraboschi, K. Ramamohanarao, R. T. Snodgrass (Eds.), VLDB 2001, Proceedings of 27th International Conference on Very Large Data Bases, September 11-14, 2001, Roma, Italy, Morgan Kaufmann, 2001, pp. 561–570. URL <http://www.vldb.org/conf/2001/P561.pdf>
- [2] O. Dalle, F. Giroire, J. Monteiro, S. Pérennes, Analysis of failure correlation in peer-to-peer storage systems, Ph.D. thesis, INRIA (2008).

- [3] S. Saroiu, P. K. Gummadi, S. D. Gribble, Measuring and analyzing the characteristics of napster and gnutella hosts, *Multimedia Syst.* 9 (2) (2003) 170–184.
doi : 10 . 1007 / s00530 - 003 - 0088 - 1.
URL <https://doi.org/10.1007/s00530-003-0088-1>
- [4] L. Viennot, Autour des graphes et du routage, Ph.D. thesis, Université Paris-Diderot-Paris VII (2005).
- [5] M. Ripeanu, Peer-to-peer architecture case study: Gnutella network, in: *Proceedings first international conference on peer-to-peer computing*, IEEE, 2001, pp. 99–100.
- [6] J. Liang, R. Kumar, K. W. Ross, The fasttrack overlay: A measurement study, *Computer Networks* 50 (6) (2006) 842–858.