24. Peer-to-Peer in Mobile Environments

Wolfgang Kellerer (DoCoMo Euro-Labs)
Rüdiger Schollmeier (Munich University of Technology)
Klaus Wehrle (University of Tübingen)

As we have seen in previous chapters, Peer-to-Peer-based applications are not limited to the well-known file sharing applications. Also, the Peer-to-Peer infrastructure is not limited to the hard-wired Internet infrastructure, but is starting to penetrate wireless networks of different characteristics. This chapter discusses the application of Peer-to-Peer based concepts in mobile infrastructure environments – including cellular systems and ad-hoc style networks. Starting with a motivation, application scenarios, and an overview of mobile system characteristics, the main part of this chapter describes challenges and possible solutions for optimizing Peer-to-Peer systems to meet the requirements of mobile scenarios. Both unstructured and structured Peer-to-Peer concepts for mobile scenarios are analyzed and discussed.

24.1 Why Is Peer-to-Peer Technology Interesting for Mobile Users and Mobile Services ?

Peer-to-Peer systems have previously been defined as self-organizing systems supporting to find and use distributed resources, i.e., services (cf. Chapter 2). With the increasing availability of mobile data communications – including Internet access in mobile networks – wired network originated Peer-to-Peer applications become available also to mobile users. They should enjoy the same level of service as they know from fixed line access while being mobile.

Furthermore, the Peer-to-Peer paradigm provides the unique opportunity for service offerings by individual users. For mobile users it is advantageous to be able, e.g., to offer instant services directly instead of uploading them to a centralized server. Moreover, services based on the local proximity of the users can benefit from a Peer-to-Peer style of provisioning without any infrastructure backing. However, for commercial applications there is still a need for some support by special, trusted parties to control service negotiation and to observe the fulfillment of service agreements [569].

However, wireless resources are usually limited and vary in performance and availability, which restricts the use of Peer-to-Peer applications in mobile environments. There are some general features in Peer-to-Peer that already address changes in the conditions of the users communication environment. For example, most Peer-to-Peer solutions can accommodate changes of a

R. Steinmetz and K. Wehrle (Eds.): P2P Systems and Applications, LNCS 3485, pp. 401-417, 2005.

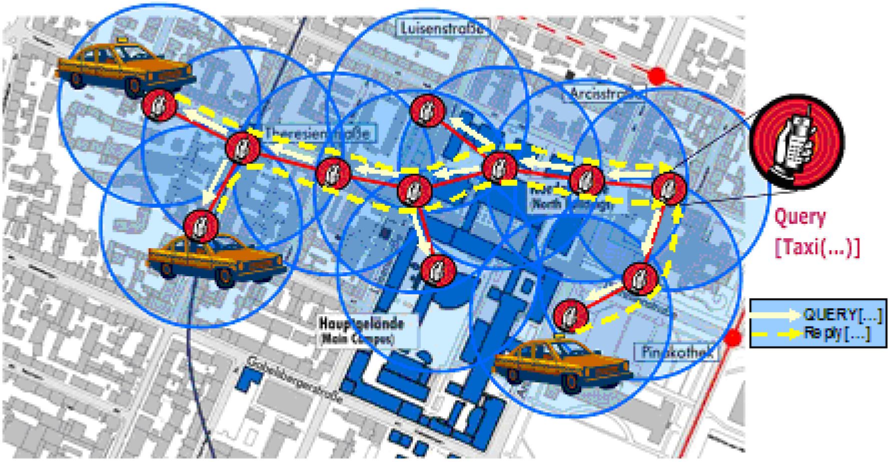

Fig. 24.1: Possible application scenario for location-based services: Locating a taxi.

node's availability caused, e.g., due to failures or joins and leaves of users. Too frequent changes as typical in wireless environments appear as a threat to conventional Peer-to-Peer-systems.

In general, the application of Peer-to-Peer to mobile environments provides a number of opportunities and challenges since, originally, Peer-to-Peer was not designed for mobile environments, which we will outline in what follows. As we will see in the next sections, two scenarios are providing the impetus for the application of Peer-to-Peer concepts in mobile environments.

Besides the well-known file sharing applications based on Peer-to-Peer networks, new wireless applications are also feasible in mobile networks, especially when we consider multi-hop links such as in mobile ad-hoc networks (MANETs). Therefore, we motivate two basic examples in the next two sections, that may be realized with Peer-to-Peer technology on top of an ad-hoc wireless network.

24.1.1 Scenario 1: Taxi Locator

Imagine, for example, a user standing at the side of a road, requires a taxi, but can not see any nearby. The user could now call the central taxi agency and order a taxi, having to state his current position. The agency, which has to track the current location of its taxis, could then direct the nearest one to the user.

If context-based routing was supported by the available MANET, the user could simply send out a request which would be broadcasted in a multihop manner, via a pre-configured number of hops, in its proximity. All participating nodes would forward the request until a taxi receives it, as illustrated in Figure 24.1. The taxi could then reply with an appropriate response message to the requesting node, and finally, pick up the user.

Thus our context-based routing scheme allows the utilization of Location-based Services (LBS) without the need for centralized elements. The underlying MANET limits flooding of the search request to the geographical proximity. Additionally, the creation of all kinds of search requests can be imagined. Possible request categories could thus also include bars, restaurants or closest bus stops.

24.1.2 Scenario 2: University Campus

The second scenario is not that highly dynamic as the first one, but also bases on ad-hoc network technology. On a university campus, today, students and teaching staff are often equipped with wireless technology, like laptops, PDAs, and Smartphones. During courses, seminars, reading groups, or in spare time, they may form groups of collaborating 'peers'. But often, collaboration with networked systems can not be deployed because of missing infrastructure support (e.g., network plugs) or due to restrictive network policies.

A solution to such problems is the formation of spontaneous wireless ad-hoc networks, e.g., based on bluetooth, IEEE 802.11x or both. Students are then able to collaborate, share teaching materials, and many more. After the courses are finished, students then move to the next activity. Thus, the spontaneous groups separate and – after a short and highly dynamic period – form new collaborative groups. These new groups may have a very different purpose.

The formation of spontaneous ad-hoc groups can easily be supported by existing MANET technology. They may also be interconnected by some Internet-connected peers, if two MANETs are within wireless reachability. But the sharing of information, the support for collaboration, and the realization of other services is not directly supported by MANET technology.

We think, that Peer-to-Peer technology has greatly proven to be appropriate for such tasks in the wired Internet. The interesting questions are now:

- Can these Peer-to-Peer-based approaches be used in MANET scenarios, e.g., like the two described here?
- What modifications and adaptations are necessary for such a purpose?

24.2 Introduction to Mobile Communication Systems

Before we take a deeper look into the challenges of Peer-to-Peer communication in mobile environments, we will briefly describe the variety of mobile communication systems and their features [348]. By "mobile communication" we mean systems with at least one hop of the communication path effected by a wireless link. They usually support interactive bi-directional communication. Broadcast systems such as Digital Video Broadcast (DVB) and Digital Audio Broadcast (DAB), or satelite-based systems are not considered in this chapter.

In general, mobile access networks can be divided into two subcategories: cellular networks, and mobile ad hoc networks (MANETs). The latter are realized either with wireless LAN or other short range wireless transmission technology. Table 24.1 summarizes the most important characteristics such as data rate and coverage.

Second generation cellular networks such as the Global System for Mobile Communication (GSM) have been enhanced to offer higher data rates and packet-based communication. For example, by providing up to 50 kbps the General Packet Radio Service (GPRS) in GSM offers much more efficient access to web resources than the 9.6 kbps, data service of GSM itself. With the recent introduction of third generation systems such as UMTS, offering up to 384 kbps, more demanding data services such as audio or video streaming, and the downloading of large media files, are also now possible. The described systems are only a small snapshot of the variety of cellular standards. Integration is expected with the introduction of the fourth generation. Cellular terminals are usually constrained in terms of processing power, storage space and battery power. Though the terminals offer only limited display capabilities, most data formats such as audio, video, image are processed. Connecting a laptop through a mobile phone to a wireless network overcomes end systems limitations, but still leaves data rate restrictions, the general problem of today's commonly available cellular networks such as GSM and GPRS.

System	Data Rate	Coverage	Frequency Range
Cellular Networks			
GSM	9.6 kbps	country-wide	900, 1800,
GSM/GPRS	$<$ 50 kbps	"	1900 MHz
UMTS	$<$ 384 kbps	partially country-wide	2 GHz
Wireless LAN			
IEEE 802.11b	max. 11 Mbps	50-300 m	2.4 GHz
IEEE 802.11a	6-24 Mbps	"	5 GHz
Short Range			
Bluetooth	max. 1 Mbit/s	10 or 100 m	2.4 GHz
IrDA	4 Mbps	line of sight	infrared

Table 24.1: Overview of mobile communication systems

Wireless data networks such as wireless LAN are of increasing importance for wireless access providing data rates of up to tens of Mbps. In contrast to cellular networks, which offer country-wide coverage and a feeling of always being connected, wireless LAN-based access is still restricted to certain isolated locations (hot spots).

Wireless LAN technology, in concert with short range communication, such as Bluetooth, also provides the infrastructure for mobile ad hoc networks. MANETs are self-configuring, multi-hop wireless networks that are not based on any fixed infrastructure such as base stations, hot spots or any central database. A node participating in a MANET acts at the same time as a data source, sink, and router on the network layer. This is very similar to Peer-to-Peer systems in that Peer-to-Peer nodes have the same roles on the application layer (cf. Chapter 2). In multi-hop MANETs, such as illustrated in both example scenarios, nodes in physical proximity are used as relay stations for routing. The main challenge for Peer-to-Peer techniques in MANETs is overcoming the instability of the physical network as connections are changing due to the movement of the nodes. This may lead to long unstable zig-zag routes on the Peer-to-Peer layer.

24.3 Challenges for Peer-to-Peer Techniques in Mobile Networks

Summarizing the observations about mobile communication systems described above, we can set down the following main challenges for Peer-to-Peer techniques in mobile networks: Mobile networks are usually resource-constrained systems limited by low data rates, low processing power, and low storage capacity of mobile terminals. Furthermore, they are characterized by frequent joins and leaves of nodes. This, what we call high churn rate, is resulting from node failures, e.g., when moving out of coverage, exhausted batteries, or from short session times, e.g., due to high online cost. To adapt to the scarce wireless resources, Peer-to-Peer solutions for mobile networks have to employ a more efficient search strategy than current flooding-based concepts. Traffic can also be minimized if Peer-to-Peer networks are aware of the underlying physical topology and long zig-zag routes are thus avoided. This is especially important for multi-hop networks in order to reduce the number of hops for the Peer-to-Peer path and to find the nearest shared resource.

As have been said, we mainly distinguish mobile ad-hoc and cellular networks in mobile environments. Cellular wireless networks are only one-hop wireless networks. Mobile Peer-to-Peer nodes are thus connected to the fixed Internet by a single wireless link. Consequently, we can state that even if the node moves, the physical path to this node does not alter very much. We therefore can regard Peer-to-Peer nodes connected to the Internet via a cellular wireless link as low performance nodes because we have to consider their bandwidth restrictions. Significant changes to the Peer-to-Peer protocol, which take into particular account the mobility of the nodes in cellular wireless systems are therefore not necessary from our point of view.

However, if we want to operate a Peer-to-Peer network on top of a MANET, we certainly do have to take into account the mobility of the nodes. The reason of course is that MANETs are self-configuring, wireless multi-hop networks, within which we have to assume the possibility of the physical path between two nodes changing frequently, due to the movement of the nodes [263]. If the overlay structure was to be established completely independently from the underlying MANET topology, it would result in a significant number of long and unstable zig-zag routes in the MANET layer, as already said. This would lead to high traffic volumes, which might not be sustainable by the MANET [356]. In the following, therefore, we will focus more narrowly on MANETs.

24.3.1 Peer-to-Peer Systems in Mobile Ad-Hoc Networks

Peer-to-Peer and MANET network architectures are very promising concepts. Although they use different layers for operation, we think that the idea to combine both architectures shows a high potential. As described in [547], the architectures have several similarities, but also many differences. To achieve a workable integration of the two architectures, a new inter-layer communication is necessary. Peer-to-Peer protocols are not aware of the underlying MANET and assume a fixed network infrastructure, causing additional and unnecessary network traffic. Thus a simple utilization of ad hoc networks for common Peer-to-Peer applications is not feasible, as we shall now see.

Common Peer-to-Peer applications have optimized algorithms to find information within their overlay network, but for information exchange they generally rely on TCP and assume stable connections. Whenever connections break, Peer-to-Peer nodes assume that their distant communication partners have left the network and switch to other, randomly chosen nodes, which can also provide the requested content, or which offer further connectivity to the network to initiate new search requests.

In MANETs, link breaks are common, as all nodes are in motion. Whenever two adjacent nodes move out of each others' radio range, the link between them breaks. A MANET protocol unaware of the Peer-to-Peer application, tries to re-establish a new route to the same destination, independently from the necessary effort. Instead of trying to create a new route to the same source of information after the network topology changed, other sources could provide the information at less cost. Therefore the MANET nodes have to report route breaks to the upper Peer-to-Peer node which then decides whether the old source is still utilizable, or wether another connection partner is more appropriate.

As common Peer-to-Peer networks operate on top of fixed network infrastructures, Peer-to-Peer nodes mostly do not distinguish between communication partners in close proximity and those more distant. Peer-to-Peer networks use multi-hop connections only for search requests. Subsequent information downloads use direct TCP connections between source and destination. The distance between nodes generally does not affect the stability and error-free operation of connections. Therefore Peer-to-Peer nodes might create connections to distant nodes, even though the information searched for might also be available more close by.

The distances between communication partners in MANETs are quantified by the number of intermediate hops. It is the most important parameter for route lifetimes [263]. Numerous unnecessarily lengthened routes induce additional routing overhead, delay data packet transmissions and as a result, greatly reduce overall network performance. Therefore, the client application must classify incoming search replies according to the hop distances to their respective destinations. As a consequence, a combined Peer-to-Peer-MANET approach must be well aware, on the one hand, of the underlying network infrastructure, and on the other hand, of the overlaying application.

This situation introduces specific requirements on flexibility and resilience of Peer-to-Peer systems in ad-hoc networks. Another difference between wired and wireless networks is that global connectivity cannot be presumed in mobile ad-hoc networks. Even if the network splits up in several parts the Peer-to-Peer system still have to be operational to enable communication between nodes.

Limited resources of mobile devices – such as computational power, memory, energy, and bandwidth – complicate the situation even more. Consequently, mobile Peer-to-Peer-systems have to use these scarce resources efficiently. Avoiding unnecessary transmissions may be thought of as an example. In regions with a high density of nodes low bandwidth must be expected because of the shared medium. Because of the long distances between nodes a low signal quality has to be expected in regions with a low density of nodes. Therefore, ad-hoc networks have an obviously lower bandwidth while the self-organization of the network causes a noticeably higher overhead.

24.4 Solutions for Peer-to-Peer in Mobile and Wireless Networks

In the following, we will present and discuss approaches to enable the use of Peer-to-Peer applications in a mobile environment [286]. As seen in the previous parts of this book, Peer-to-Peer technology can roughly be divided into two basic approaches: structured and unstructured Peer-to-Peer systems (see Chapter 2). Therefore, we will discuss the use of both paradigms separately in Sections 24.4.1 and 24.4.2 though there may be some overlapping similarities between both.

24.4.1 Solutions Based on Unstructured Peer-to-Peer Networks

If we want to employ Peer-to-Peer networks in MANETs as mentioned in Section 24.3, we have to take into account, in contrast to cellular wireless networks, that the topology of the underlying physical network frequently changes. Due to the assumed node movements (sensor networks are not regarded in this context), the links between the nodes frequently change leading to frequent path breaks [263]. Therefore it is necessary to align the virtual to the physical topology to avoid zig-zag routes as described in [550].

Summarizing, we can state that although pure Peer-to-Peer networks and MANETs employ similar networking approaches, their combination by simply establishing a Peer-to-Peer network upon a MANET, without providing any interaction between them will most probably show low performance [356]. Thus, only recently have some approaches been developed to provide content-based routing, caching, or Peer-to-Peer networking in mobile ad hoc networks, which take into consideration the physical topology.

7DS [387], for example, employs local broadcast messages to provide an infrastructure for web browsing without a direct connection to the Internet. Therefore, every node acts as a mobile cache, which is continually renewed if a direct connection is available.

Besides providing Internet content in MANETs, ORION [356] is, next to the Mobile Peer-to-Peer protocol (MPP) [548, 264], a system, which aims to provide Peer-to-Peer services in a MANET indexMPP. ORION provides a general purpose distributed lookup service and an enhanced file transmission scheme to enable file sharing in MANETs. In contrast to MPP, ORION, to a certain extent, separates the Peer-to-Peer network from the physical network. In order to describe the methods used and the properties exploited for Peer-to-Peer in MANETs, we will first describe some of the details of the Mobile Peer-to-Peer protocol.

The Mobile Peer-to-Peer protocol adapts the overlay structure to the physical MANET structure via a cross-layer communication channel between the physical network layer and the virtual Peer-to-Peer layer [547]. This integrated approach significantly reduces the messaging overhead and increases the search success rate compared with approaches using a separated treatment of the overlay and the physical networks. Further, MPP allows the introduction of a variety of new services, since it provides the possibility of context based-routing and location-based services, instead of simple address routing as is provided by current MANET routing algorithms. Thus new services like the search for a taxi or a cash dispenser in your proximity can be

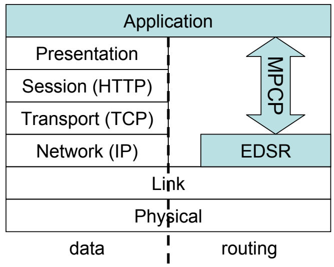

Fig. 24.2: Layered structure of MPP

realized without the necessity for further location-sensitive sensors and central instances. Further applications of such a combined approach have been described previously.

To minimize the effort to create a new protocol and to benefit from former developments, the MPP protocol stack reuses existing network protocols as much as possible. For node-to-node communication, the protocol utilizes an enhanced version of the Dynamic Source Routing (DSR) protocol [326] . For the transportation of user data it uses HTTP over TCP, as illustrated by Fig. 24.2. Thus the Enhanced Dynamic Source Routing (EDSR) requires only a new application layer protocol and minor changes within the DSR protocol. To connect the application layer protocol (MPP) with the physical network layer protocol (EDSR), the Mobile Peer Control protocol (MPCP) is used.

Since MANETs already provide routing algorithms which enable the localization of network participants by their IP addresses, an additional Peer-to-Peer implementation of this functionality is unnecessary and even degrades the performance. Consequently, EDSR is designed to perform the necessary routing tasks on the network layer and supplements the application layer protocol (MPP). This approach provides valuable advantages compared with a separate treatment of both networks:

– The MANET controls the organization of the network. Thus changes in the topology of the mobile network are taken into account automatically by the Peer-to-Peer network.
– The network layer is responsible for routing and the application controls the data exchange.
– The integration of both networks avoids redundant information requests.
– The inter-layer communication of the protocol optimizes performance, since the overlay network can be optimally adjusted to the physical network.
– The application layer protocol MPP simplifies the implementation of new services.

The separation of data exchange and routing tasks allows the reuse of existing protocols like TCP and HTTP. Only for routing tasks must MPP directly interact with EDSR residing in the network layer (cf. Fig. 24.2).

MPP allows distant peers to transparently exchange data. Therefore MPP is responsible for file transfers within the Peer-to-Peer network and resides in the Peer-to-Peer client application. MPP utilizes HTTP for data exchange since it is simple to implement and well tested. The HTTP content range header is able to resume file transfers in case of network errors due to link breaks. EDSR is mostly based on the DSR protocol, but additionally specifies new request and reply types to provide the means for finding peers by criteria other than the IP address. EDSR thus extends DSR and therefore EDSR nodes can be an integral part of DSR networks.

MPCP is the inter-layer communication channel between the application and the network layer. Thus MPCP links the EDSR Protocol in the network layer with the Peer-to-Peer application in the application layer. Using MPCP, the application can register itself in the EDSR layer to initialize search requests and to process incoming search requests from other nodes. It communicates to the corresponding protocol all incoming and outgoing requests and responses, exept the file exchange itself.

On startup, the Peer-to-Peer application on the mobile device announces itself to the EDSR layer via MPCP. If a user initializes a data search, MPCP forwards the request to EDSR which transforms it into a search request (SREQ). Similar to DSR route requests (RREQ), EDSR floods SREQs through the MANET. EDSR nodes receiving the request, forward it to the registered Peer-to-Peer application via MPCP. Thus the Peer-to-Peer application can determine whether locally shared data satisfies the request's criteria. If the request matches the description of a file shared by the node, the application initializes an EDSR file reply. This reply is sent back to the source node and contains all necessary information for the file transfer. Similar to DSR route replies (RREP), a file reply (FREP) includes the complete path between source and destination.

To compare the performance of a protocol adapted to the underlying physical network, like MPP, with a protocol which establishes its overlay still completely independently of the physical layer, we use an analytical approach. First, we have to evaluate the number of reachable nodes in a MANET environment.

If we assume x_0 neighbors on a per node average and a radio reach of R_0, we can compute the node density as:

$$nodedensity = \frac{x_0}{R_0^2 \Pi} \tag{24.1}$$

If we further assume a uniform distribution of the nodes in the plane, the cumulative distribution of the distance between two nodes is given by:

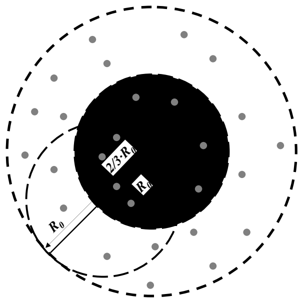

Fig. 24.3: Multihop reach of an average ad hoc node

$$F(r) = \frac{r^2 \Pi}{R_0^2 \Pi} = \left(\frac{r}{R_0}\right)^2 \tag{24.2}$$

which results in an increasing probability of occurrence for an increasing distance between two ad hoc nodes. The pdf of this function can now be computed by taking the derivate of $F(r)$:

$$f(r) = \frac{2r}{R_0^2} \tag{24.3}$$

This simply reflects the fact that the differential surface increases linearly with an increasing radius r. Accordingly, this also means that the probability of occurrence of nodes within the distance r also increases linearly. Thus the average distance between two nodes can be computed by:

$$\bar{d} = \int_0^{R_0} \frac{2r}{R_0^2} \cdot r dr = \frac{2}{3} R_0 \tag{24.4}$$

This means, as illustrated by Figure 24.3, that the multihop reach of an average node only increases by $\frac{2}{3}R_0$ instead of R_0. Thus the number of reachable nodes via h physical links can be computed by:

$$\Delta N_{phys} = \begin{cases} R_0^2 \Pi \cdot \frac{x_0}{R_0^2 \Pi} = x_0, \ h = 1 \\ \left(\left(1 + (h-1) \cdot \frac{2}{3}\right)^2 - \left(1 + (h-2) \cdot \frac{2}{3}\right)^2\right) R_0^2 \Pi \cdot \frac{x_0}{R_0^2 \Pi} \\ = \frac{8}{9} h x_0, \ h > 1 \end{cases} \tag{24.5}$$

If we now assume that a node in a Peer-to-Peer network does not adapt its overlay network to the underlying physical network, then it must establish its connections randomly. From equation 24.5 we can already observe that the further away a node is, in terms of physical hops, the higher the probability is that it connects to it. If we assume random selection, we can compute the probability in detail by:

$$p_{con}(h) = \frac{\Delta N_{phys}}{\left(R_0 + (h_{\max}-1)\cdot\frac{2}{3}R_0\right)^2 \Pi \cdot \frac{x}{R_0 \Pi}} = \begin{cases} \frac{1}{\left(1+\frac{2}{3}(h_{\max}-1)\right)^2}, & h = 1 \\ \frac{8h}{9\left(1+\frac{2}{3}(h_{\max}-1)\right)^2}, & h > 1 \end{cases} \tag{24.6}$$

where h_{max} defines the maximum possible number of physical hops, which is commonly limited to six [263]. The average path length of a not-adapted overlay network in a physical network can thus be computed by:

$$\begin{aligned} l &= 1\cdot\left(1+\frac{2}{3}(h_{\max}-1)\right)^{-2} + \sum_{h=2}^{h_{\max}} \frac{8}{9}h^2\left(1+\frac{2}{3}(h_{\max}-1)\right)^{-2} \\ &= \left(\frac{8}{9}\left(\frac{h_{\max}(h_{\max}+1)(2h_{\max}+1)}{6}-1\right)+1\right)\left(1+\frac{2}{3}(h_{\max}-1)\right)^{-2} \end{aligned} \tag{24.7}$$

Thus if we assume, for example, a maximum of six hops in the overlay network, we can compute the average path length in a not-adapted overlay network to 4.31 physical hops. This means that every message is transmitted in the network 4.31 times more often, than in an adapted overlay network like in MPP. In MPP, every overlay message is transmitted only once via one physical link, as a result of the cross-layer communication channel between the physical and the application layers. Further reduction of signaling traffic can thus be achieved because no additional keep-alive messages are necessary.

24.4.2 Solutions Based on Structured Peer-to-Peer Networks

Due to characteristics like flexibility, scalability, and resilience, structured Peer-to-Peer systems became a common approach for building decentralized self-organizing networks of any size. In particular, Distributed Hash Tables (DHT) have been proven as the primary design choice, as already discussed in Chapter 2 and Part III[1]. On the other hand, ad-hoc networks gain greater importance due to the increasing occurrence of scenarios without a centralized infrastructure. Whenever there is need for a scalable data management without any infrastructure, the combination of ad-hoc network- and DHT-technology seems to be an obvious solution [286]. The question, whether this is a fruitful combination, will be discussed in the following.

In general, DHTs have been developed for the infrastructural Internet, and therefore, some basic assumptions can not be directly applied for the use in ad-hoc networks. In the following, we will show the main differences

[1] Therefore, in the following, we will focus on DHTs as the state-of-the-art technique in structured Peer-to-Peer systems and will use both terms interchangeably.

between these network technologies, discuss resulting problems, give solutions for dedicated problems, and will show scenarios in which ad-hoc networks and DHT-protocols may be combined.

Challenges

As shown in Part III, numerous DHT approaches with versatile characteristics determined by the subjacent topological structure (routing geometry) exist. This structure induces different characteristics regarding flexibility and resilience. A high flexibility in the choice of neighboring nodes enables optimizations with respect to the underlaying network topology. A resilient DHT structure can still operate without the need for evoking expensive recovery-algorithms – even if many nodes fail at the same time. These issues are essential for the DHT to react flexibly on topological changes of the underlying network.

Flexibility in the Choice of Neighboring Nodes

According to [266], the flexibility of a DHT is the "*algorithmic freedom left after the basic routing geometry has been chosen*". For a Peer-to-Peer system this freedom is essential to build an efficient overlay-network. DHT structures with a high degree of flexibility are capable of adapting the routing to the frequent changes in an ad-hoc network. Two respective approaches have been proposed in [266].

Proximity node selection (PNS) describes the selection of nodes as optimal routing-table entries. A node can build its routing table with respect to several criteria, e.g., hop-count, delay, bandwidth, battery life, remaining capacity, or transmission power. As a result, short paths and stable underlay-connections improve overlay routing. This pre-selection of neighboring peers can be integrated into existing DHT protocols if the DHT structure is flexible enough. As shown in [266], PNS can easily be achieved in ring-based DHTs such as Chord. The strict rules for building the routing table make an efficient PNS impossible for tree-, butterfly-, or hypercube- based DHTs. With node movement in mind, the PNS procedure has to be repeated from time to time to achieve continuous improvements.

Proximity route selection (PRS) can be used to find optimal routes when using a pre-existing routing table. During the routing process a node can decide to which known node the request will be routed to. This decision is made upon the same criteria which are used for PNS. Less latency and a higher network stability can be achieved when routing to weak nodes, which are likely to fail, is omitted. Depending on the geometric structure used by a DHT, the routing protocol is able to adapt its routing behavior. According to [266], hypercube based DHTs are very well suited to use PRS because of varying, equidistant paths between two distant nodes (all are shortest). On the other hand, tree and butterfly based DHT protocols are not capable of using PRS because there is only one path existing in the DHT structure that allows a decrease in the distance between the two nodes. For this reason the routing algorithm does not allow any variations. Ring-based DHT-protocols have to make a tradeoff between an increased hop-count in the overlay and an eventually shorter or more stable path in the underlay.

In wireless networks the inherent broadcasting of packets to neighbors can be used to improve the peers overlay routing, e.g., like "shortcuts" in the overlay network. Routing to these surrounding peers causes just one hop in the underlying network which makes it very efficient. Keeping the connections to these nodes causes only small amounts of locally restricted network traffic.

If the ad-hoc network protocol is able to analyze packets, a message can be intercepted by a node which takes part at the routing process in the underlay network. This node may redirect the request to a node which is closer to the destination based on stored information on surrounding peers. The decision whether intercepting a message or not must be taken in respect to the progress that would be achieved in the overlay structure compared to the distance on the alternative route. If connection speed is not an issue, even nodes not directly involved in the routing may intercept a message. The route can be changed, and a route change message has to be sent to the node responsible for processing the routing request. If no route change message is received within a specified period of time, the routing progress will be continued. However, this procedure will decrease network traffic at the cost of increased latency. Route interception and active routing of non-involved peers can be used to achieve more redundancy, leading to a more stable network in case of node failures.

Resilient Networks

DHTs are considered to be very resistant against node failures. Backup and recovery mechanisms, that use distributed redundant information, ensure that no information is lost if a node suddenly fails. Depending on the subjacent DHT topology, the DHT experiences a reduced routing performance until the recovery has finished. [266] shows that especially tree and butterfly based topologies are vulnerable to node failures. Due to the high flexibility of ring based topologies, these DHTs are still operable in case of massive node failures.

When DHT protocols are used in an ad-hoc environment, resilience has to be considered as a very important issue. The resilience of a DHT determines how much time may pass before expensive recovery mechanisms have to be evoked. As the quality of connections in ad-hoc networks is highly dependent of the environment of the nodes, some nodes may be temporarily inaccessible or poorly accessible because of node movement. If the recovery process is started too early, an avoidable overhead is caused if the node becomes accessible again. If the topological structure allows the DHT protocol to delay recovery mechanisms without losing routing capability these costly recovery measures can be avoided. This approach has a positive effect on the maintenance costs of a DHT. In a worst case scenario, a node which is partly available and unavailable over a longer period of time can stress the whole network because of numerous join and leave procedures. This scenario can easily be provoked by node movement along the network perimeter. Resilience is therefore an important factor when DHTs are used in combination with ad-hoc networks. Resilient DHT structures are capable of compensating node failures and are able to use recovery mechanisms more accurately.

Member Selection Based upon Node Characteristics

To avoid problems with heterogenous nodes in ad-hoc networks it is important to judge the suitability of nodes by means of measurable criteria. The duration of the connection to the network, energy, link quality, or node mobility are indicators for the reliability of a node. If the decision whether a node may join the network or not, is based upon these parameters, a higher stability can be achieved due to a network consisting of reliable nodes. If free-riding is not a problem or if measures can be taken against free-riders, nodes rejected because of their low reliability could still use the DHT without being involved in the routing process. That way all nodes can benefit from the DHT without imposing weakness.

The indicators described above can also be used to predict node failures due to low energy or weak connections. Nodes which are likely to fail soon can inform their neighboring nodes in the DHT about their likely failure. Costly recovery mechanisms can be avoided, which leads to a lower overhead caused by sudden node failures.

Splitting and Merging DHTs

A separation of the network into multiple parts may be caused by the failure of one single node. For the nodes connected to a smaller part of the separated network the separation is equal to the failure of the majority of all nodes. This can hardly happen in infrastructural networks. Thus, a higher degree of redundancy is necessary to keep a DHT operational. By distributing a nodes' information to n neighbors the probability of a fatal DHT failure can be kept below $1/2^n$ if the network splits up in two equal parts.

Common DHT approaches are built upon the fact that only one instance of a DHT – accessible for every node in a global network – exists. When DHTs are used in an ad-hoc network this assumption leads to a serious problem: If two independent DHTs (parts of a separated DHT or completely independent DHTs) encounter, suited methods – to either merge those structure or to enable communication between them – have to be found.

Merging two DHTs imposes a vast amount of network traffic. Many key-value pairs have to be redistributed and new neighborhood connections must be established. This stress may often be unacceptable, especially if the connection between the DHTs is weak or short-lived. A method to merge large DHT structures has to be designed in respect to the limitations (low bandwidth, high latency, etc.) of an ad-hoc network, to avoid network overload. Criteria like the stability of inter-DHT connections must be judged to avoid a merge of two DHTs which is likely to split up again. However, if the merging of DHTs is omitted – and structured communication between DHTs is chosen as an alternative – continuously separating a DHT will create many small DHTs and causes an enormous communication overhead. In both cases the simultaneous coexistence of DHTs requires an unambiguous DHT-identification.

Improving the Teamwork Between Ad-Hoc Network and DHT

The characteristics of the underlying ad-hoc network protocol has great effect on the performance of the overlay as the DHT induces a constant flow of control and query messages. An optimized interaction between ad-hoc network and DHT is essential to create an efficient combination. Two aspects of the teamwork between over and underlay will be discussed exemplarily in the following section.

Reactive routing algorithms as AODV maintain their routing tables as they process routing requests. Old routes (to other nodes) become invalid after a certain period of time. If the time between keep-alive-messages, which are sent by the DHT-nodes is shorter than the time that may pass before an AODV-route becomes invalid, constant route finding requests can be avoided. The steadily used routes never become invalid and efficient underlay routing on overlay connections is possible.

On the other hand, proactive routing algorithms [483] are suited better for the use of PNS. PNS opens many connections to different nodes to examine the connection characteristics of many potentially optimal neighbors. The best nodes will be chosen as routing table entries. High initial delay and the additional overhead caused by the route requests of an reactive ad-hoc protocol interdict the use of PNS in reactive ad-hoc networks. These examples show that an arbitrary combination of ad-hoc network protocols and DHT protocols can result in inefficient solutions.

24.5 Summary

In this chapter, we have investigated the opportunities and challenges of the application of Peer-to-Peer concepts to mobile environments. A discussion of mobile network characteristics shows that mobile ad-hoc networks impose most severe problems in terms of low data rate, limited terminal capabilities, and high churn rate to conventional Peer-to-Peer systems. Therefore, they are in the focus of this chapter.

Although there is an inherent similarity of ad-hoc networks and Peer-to-Peer systems a powerful combination of both technologies cannot be created without a careful investigation of the resulting problems. Common Peer-to-Peer systems must be modified in many ways to enable their use in ad-hoc networks. Due to the nature of ad-hoc networks the choice of a specific Peer-to-Peer approach has to be considered carefully.

Structured and unstructured Peer-to-Peer-concepts have been discussed for their suitability to run on ad-hoc networks. As of their robustness and reactive behavior, unstructured Peer-to-Peer concepts seem the natural choice for ad-hoc networks. The application of cross-layer interworking of application layer and routing layer protocol, as described here, improves the performance of unstructured Peer-to-Peer systems over ad-hoc networks significantly. This is shown with an example protocol and in an analytical investigation.

Regarding the practical use of structured Peer-to-Peer concepts (DHT structures) in ad-hoc networks further investigation on split and merge algorithms for multiple DHTs are necessary. Furthermore, the impact of node movements on the DHT structure will be interesting topics of future research. The question what degree of redundancy is needed to provide a stable service – while operating in a highly dynamic network – has to be answered if real-life applicability is the final goal.

As applications operating on DHTs gain greater importance in the infrastructural Internet, efficient porting to wireless scenarios might also be needed soon. DHT structures suited for the use in ad-hoc networks could provide a generic interface which enables their usage without profound modifications. Applications like the ones presented in Part IV might be usable in ad-hoc scenarios if efficient Peer-to-Peer solutions can be provided for the challenges discussed in this chapter.