1. Introduction

The amount of hosts on the Internet had and still has a tremendous growth rate. It is today's dominating communication platform and it became an important tool for a huge amount of people and also a strong business platform. After the initial rise of the Internet another communication technology underwent a similar popularity development. The cellular phone rapidly found its way into the every day lives of people changing their way of communication drastically towards the "anytime, anywhere" paradigm. Nowadays, devices such as SmartPhones and Personal Digital Assistants (PDAs, refer to the list of abbreviations in the Appendix) for personal communication, but also cars, appliances and even clothes just to name a few are being equipped with communication technology, predominantly using a license-free spectrum. This allows devices to communicate with each other, forming a flexible, distributed, decentralized and spontaneous network. Those ad hoc networks allow for new applications and with the current technological development might soon be widely adopted and experience a similar increase in popularity as cell phones and the Internet.

The concept of ad hoc networking can be found in a vast amount of possible scenarios. Those range from disaster response, vehicle-to-vehicle communication, sensor networks, community networks, smart spaces and e-learning to battlefield networking as partly depicted in Fig. 1.1. All these scenarios have different characteristics and requirements which potentially vary enormously from application to application. Independent from applications and scenarios there are some key characteristics which can be found in all ad hoc networks although with varying impact on the networking performance. The key property is that ad hoc networks are self-organizing, infrastructure-less networks. This lack of infrastructure adds a high degree of complexity as centralized approaches might not be feasible to be realized. Many problems that can easily be solved using some central entity such as a server or a high performance communication backbone need to be solved differently from existing solutions that assume an infrastructure. Such problems which include for example unique IP address assignment, object and service location or synchronization have to be solved in a distributed fashion in ad hoc networks. In contrast, cellular networks for example can always rely on the high performance communication backbone and on the existence of centralized services. In addition, nodes in ad hoc networks can be mobile resulting in topological changes and disconnection. The shared wireless medium is error-prone with fast, time-varying channel characteristics and only provides a relatively low bandwidth. Multi-hop capability is assumed and nodes must therefore act as a router and forward packets on behalf of other nodes. The devices used in such environments are likely to be battery-driven and therefore energy management can become an important issue. Clearly, ad hoc networks can be highly dynamic and resource constrained, resulting in a multitude of problems which can arise in such environments. Current real-world implementations of ad hoc network protocols have been proven to only perform poorly partly due to these problems [1] [2]. On the other hand they are quickly deployable, do not rely on any kind of infrastructure such as a communication backbone or servers, are flexible, can be deployed spontaneously, ideally without the help of experts, and come at a low cost since they quite likely operate in one or multiple license-free spectra.

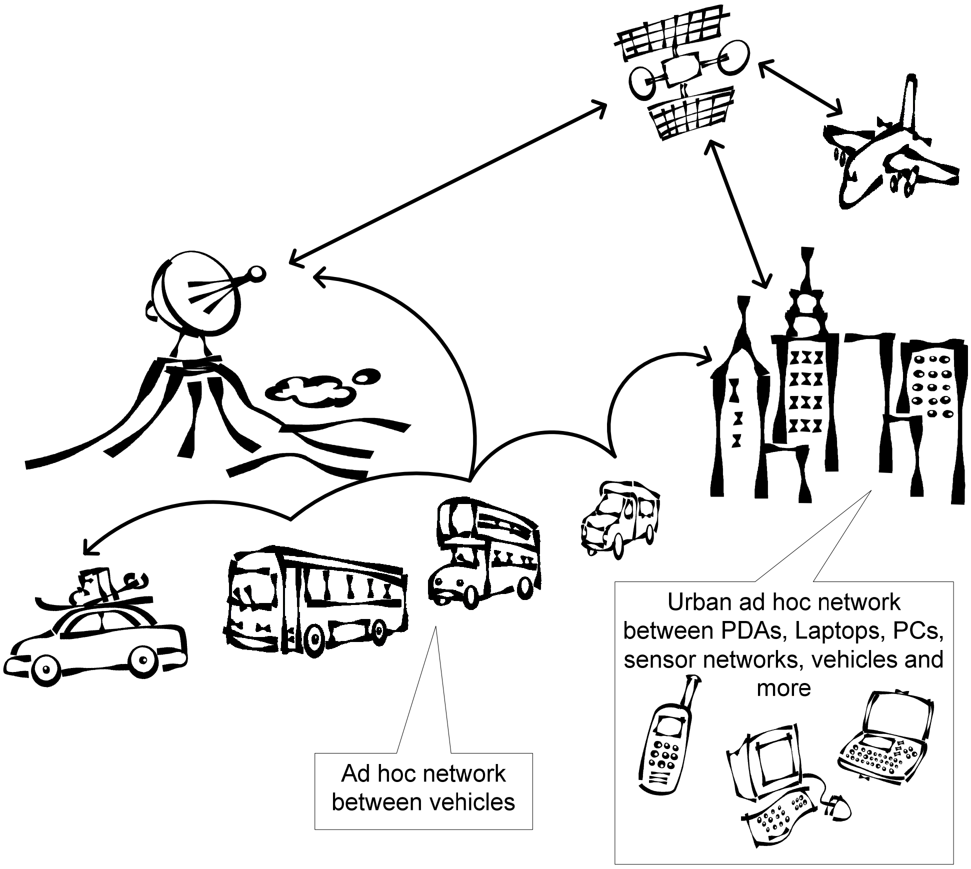

Fig. 1.1 Possible ad hoc networking scenarios and infrastructure integration

The flexibility and self-organizational properties of ad hoc networks are very compelling arguments for a future widespread use of this kind of technology. On the other hand the challenging environments and the limiting factors intrinsic to ad hoc networks [3] as described above seem deterrent. In vehicle-to-vehicle scenarios for example energy is of no concern since a car generates electrical energy but the mobility of a car might impose the biggest challenge to successful communication. But the impact of mobility varies over time, i.e. most likely with the speed limit, the street layout and the traffic density. In sensor network scenarios on the other hand nodes might not move at all but the nodes are expected to operate over a long period of time without any maintenance and external energy source. In this kind of environment energy management plays a key role. These two examples show that the deployment scenario for an ad hoc network determines the way protocols must behave. The scenarios vary drastically in their characteristics and even within a given scenario the network conditions can change over time in a non-periodical and non-deterministic way. All theses factors impose high demands on network protocols for ad hoc networks.

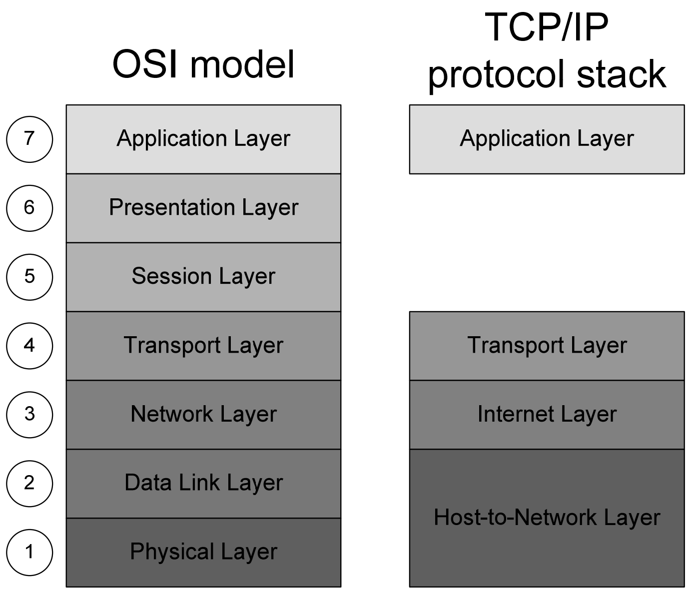

Fig. 1.2 Layered reference models (ISO/OSI and Internet)

Simply adopting existing protocols and mechanisms will either fail or suffer from a poor performance since typical protocols such as the ones the TCP/IP stack comprises of, meet requirements not applicable to mobile wireless ad hoc networks [1]. Ad hoc networks suffer from various unique challenges as the ones mentioned before. The mobility of participating nodes for example causes a highly dynamic network topology which frequently results in routes to fail and new ones to be established. This is simply not an issue in the Internet, for example, and therefore novel routing algorithms need to be investigated that are able to compensate the topological dynamics found in ad hoc networks. Going a step further the question arises whether the network architecture used in most modern communication systems, namely the layered protocol architecture conceptually based on the ISO/OSI model (see Fig. 1.2), is suitable in highly dynamic networked environments.

1.1. Problem Statement

A vast amount of protocols for ad hoc networks have been proposed. Especially routing protocols have been in the spotlight of attention for a long time and they are still an intensively researched topic in the field of ad hoc networking. They differ in various aspects such as the technological assumptions made for the hardware including Global Positioning System (GPS) or certain radio technologies. Another difference would be the performance focus of a protocol ranging from energy efficiency over Quality of Service (QoS) provisioning and overhead to throughput and delay minimization. To cope with the network dynamics, often periodic topology control messages are exchanged that contain path and link information. Another approach would be to let routing information expire once it is not actively used any more for a certain time period. The obvious problem with those and similar approaches is the static parameterization of a protocol. If a node sends out periodic route maintenance packets the time period might be chosen too large to compensate for rapid topological changes. When choosing a time interval which is far too small resources such as energy and bandwidth are wasted lowering the overall performance. Similar considerations apply to all network protocols and issues other than mobility. That means that such protocols only operate efficiently in a narrow spectrum of network dynamics. Clearly, for dynamic environments more adaptive mechanisms must be applied to achieve operational optimality over a broad range of possible networking conditions.

Static parameterization is one of the main reasons for inefficiencies in protocol behavior. Another important aspect that is closely connected are the various bottlenecks that can be found in ad hoc networking environments such as energy and bandwidth. These bottlenecks are not necessarily performance metrics in a classical sense. If energy consumption is an issue than the overall communication performance might be sacrificed to preserve the energy resources. Energy consumption as such must be available as a metric for protocols in order to be able to adapt and being efficient. An architectural framework would make sure that such metrics are calculated and interpreted identically on different nodes therefore guaranteeing interoperability and compatibility.

Since all protocols of the network stack should be based on adaptive algorithms the network architecture itself should provide a common framework for adaptation provisioning instead of having independent adaptation mechanisms on a per-protocol basis. An architectural solution to the problem will greatly simplify and unify the problem domain.

A network architecture for ad hoc networks or in general for highly dynamic networks must allow for adaptations on all protocol layers. Essentially, that means that relevant network information must be gathered and provided to all protocols of the network stack. According to that information a protocol can adapt to changing network conditions and can therefore be efficient in a broad range of different network scenarios and environments. That of course is only the special architectural consideration for ad hoc networks. Furthermore, such an architecture should of course also share the conceptual key characteristics of the ISO/OSI model as displayed in Fig. 1.2, which has proven its benefits through its widespread use und longevity.

A suitable general approach that conceptually satisfies the above mentioned requirements would be the cross-layer design paradigm. An architecture based on cross-layer mechanisms on the one hand preserves conceptual characteristics of a layered protocol stack but at the same time partly weakens the very strict functional independence of protocols by information interchange.

1.2. Goals

Considering the problem statement there are certain requirements a (cross-layer) architecture for ad hoc networks should fulfill to overcome the special problems that can be found in such highly dynamic and distributed systems. Additionally, there are various other requirements which have to be taken into account:

- Ad hoc networking issues

The main reason to change to an alternative architectural design model for ad hoc networks is the multitude of intrinsic network dynamics that can hinder a more “traditional” network architecture to support network protocols efficiently. A suitable architecture must be able to help alleviate the performance degradation due to these issues such as the mobility of nodes.

- Conventional network issues

Conventional or traditional network challenges including issues such as load balancing or QoS will also occur in ad hoc networking environments. Due to the nature of distributed wireless networks, QoS provisioning for example needs to be handled differently as compared to infrastructure based networks and the load distribution in a network will most likely be based on differing patterns and phenomena. A well designed architecture will support protocols to efficiently handle conventional network issues in novel ways that are tailored to ad hoc networking environments considering their peculiarities.

- Support for novel applications

Applications for ad hoc networks algorithmically will most likely go beyond their counterparts found in infrastructure-based, centralized networks. Problems such as replication or auto-configuration are far more difficult to solve and will therefore require more complex solutions. The architecture should provide some support to reduce this complexity on the protocol level and should provide data to support decision processes such as replication placement for example. The architecture should also be able to even further improve applications that are already tailored towards mobile ad hoc networks designed for a purely layered architecture and finally support novel applications only applicable in ad hoc networks.

- Optimization metric generation and provisioning

The architecture will provide the facility to exchange information in a well structured way. Part of the architectural design must be that the information exchanged is provided in a way that it is interpreted the same way on every node and that the data itself is generated uniformly. Ideally, some of the optimization metrics are generated within the architecture from low-level protocol or system information such as data from the hardware or operating system. Generic, low-level information would guarantee that this information is present independent from the actual protocol, operating system or hardware. An additional task of the architecture would be to derive more meaningful and expressive metrics from the basic information provided.

- Elimination of static parameterization to a large extend

Static parameterization is one of the main reasons for inefficiencies and the waste of resources in ad hoc networks. By gathering data from all protocol layers and using it for further metric refinement, parameter values can be set based on those metrics generated. This way protocols can operate more efficient and utilize resources closer to the optimum.

- Seamless system integration and interoperability

It cannot be assumed that a new communication architecture will replace existing technology fast enough to be not concerned with system integration and interoperability. Protocols should therefore be designed in a way that they can function efficiently in a network where there are also nodes running a purely layered network stack and none-cross layer optimized versions of the corresponding protocols. Furthermore, on a given node it might be possible that a mix of cross-layer optimized and none-cross layer optimized protocols are running. The overall functionality of the network stack must be guaranteed in such a case.

- Fallback capability and hardware independence

If for some reason an optimization metric is unavailable a cross-layer protocol must still be able to provide its function, although of course in a less optimized way. That means there must be a fallback mechanism to none-cross layer mode. Without optimization information the protocol should behave exactly as its un-optimized layered counterpart to guarantee interoperability as stated above. That also implies that protocols must be able to operate hardware independent to a certain extend. GPS for example might not work in indoor environments but networking should nevertheless be possible.

- Network-wide optimizations

Local optimizations based on locally available information might not have a strong impact on the overall network performance. Even worse, localized optimizations might as a whole influence each other negatively and worsen the performance. Some optimizations are in their nature network-wide problems. In ad hoc networks, routing for example is a task that has to be performed at every node. This makes load balancing a network wide issue as packets can potentially be routed through any node in the network and not only through dedicated backbone nodes.

In general an architecture should of course be efficient, scalable and robust. But these are generic requirements not particularly applicable to ad hoc networks or network architectures for dynamic systems. The solution, CrossTalk, developed and analyzed in this dissertation was designed considering these special goals. Its ability to actually fulfill these objectives is shown using exemplary cross-layer optimizations based on protocols which only act as demonstrators not limiting the generic nature of the CrossTalk architecture.

1.3. Summary

Ad hoc networks are highly dynamic network environments which makes protocols design very challenging. The vast amount of routing protocols proposed is a good indicator of the fact that solutions based on a layered protocol stack turn out to be far away from optimal operation depending on the scenario and the degree of the network dynamics such as mobility. Therefore, an architectural framework is needed that supports protocol adaptations and optimizations so that protocols cover much broader deployment scenarios efficiently. Cross-layer approaches are a solution to this problem as they allow protocols to exchange information which can be used as input for algorithms, for decision processes, for computations, and adaptations to protocol behavior. Such an architecture should expose certain characteristics such as interoperability with none-cross-layer optimized protocols and the ability to support network-wide optimizations.

1.4. Chapter Organization

The remainder of this document is organized as follows.

Chapter 2 discussed design alternatives to the cross-layer design including their advantages and disadvantages. Following the introduction of architectural network stack designs, DynaMO, a peer-to-peer substrate for ad hoc networks is described. DynaMO is part of the work that lead to CrossTalk, the cross-layer architecture described in later chapters and serves as a good example of in-layer adaptations that do not loose their importance by the introduction of cross-layer design.

Chapter 3 presents related work in the field of cross-layer design. First, individual exemplary cross-layer adaptations and optimizations are described followed by more generic cross-layer architectures and frameworks. This chapter also addresses some criticism that was expressed towards unbridled cross-layer design.

In chapter 4 the CrossTalk cross-layer architecture is presented and its mechanisms, components and functionality are discussed. That includes the Local View component that reflects the local nodal state and the Global View component that is used to estimate network-wide metrics.

Chapter 5 describes the results from exemplary applications of the CrossTalk architecture. The analysis presented in this chapter also comprises a novel application and investigates the scalability relevant questions concerning CrossTalk's mechanisms and components.

Chapter 6 concludes this document and the results concerning the CrossTalk architecture and chapter 7 gives an outlook on interesting future work.