

Ad-hoc Internet Connectivity

Eoin Curran and Jim Dowling

28th May 2003

Abstract

For a large class of wireless networks connectivity to fixed services such as the internet will be essential. However, since nodes in a wireless network are mobile, their point of connectivity to a fixed network may be continuously changing. In many cases the route to the fixed network may involve multiple hops across other mobile nodes. This document is an overview of current developments and research in the area of ad-hoc networking and mobility as they relate to the integration of ad-hoc networks with fixed networks.

1 Introduction

This document is an overview of current developments and research in the area of ad-hoc networking and mobility as they relate to the integration of ad-hoc networks with fixed networks. In particular, the following topics are examined:

- Physical Network Layer: 802.11
- Ad-hoc routing protocols
- Internet Connectivity for Ad-hoc Networks: *ad-hoc connectivity*
- Possible Scenarios for ad-hoc networks
- Other research which may be relevant to the topic:
 - Swarm Intelligence
 - Non-deterministic flooding
 - Service-based Routing

1.1 What is Ad-hoc?

There are a number of characteristics that make a network inherently '*ad-hoc*':

- Zero-configuration. A network may be made up of members from multiple administrative domains. Nodes should be able to join the network and access it's services easily.
- Peer-to-peer. A node both consumes and provides the services of the network.
- Dynamic Topology. Ad-hoc almost always means wireless. With widespread availability of 802.11 hardware, ad-hoc almost always means radio. Nodes may be mobile, may only be temporarily available.

The terminology *ad-hoc* is, unfortunately, used to mean different things in different contexts. In the context of IEEE 802.11, ad-hoc simply means the lack of infrastructure, but not a multi-hop network:

“a network composed solely of stations within mutual communication range of each other via the wireless media” ([IEE99, XS01]).

But in common usage, the term ad-hoc almost always means a multi-hop wireless network. For example, Perkins writes ([PB94]):

“Ad-hoc networks differ significantly from existing networks: The topology of interconnections may be quite dynamic. Users will not wish to perform any admin actions to set up such a network. We do not assume that every computer is within communication range of every other.”

Some more specific characteristics of ad-hoc networks have been identified. The IETF has a working group for mobile ad hoc networks, and there are commercial and community efforts in *mesh networking*.

1.1.1 MANET

In [CM99], the IETF identifies what they characterise as a Mobile Ad Hoc Network (MANET). It is a collection of nodes, each of which is equipped with one or more wireless network interfaces. The system may operate in isolation, or have gateways to and interface with a fixed network. When a MANET is connected to a fixed internetwork, it is envisioned that it will operate as a *stub network*, i.e. only carrying traffic originating at or destined for internal nodes.

A number of characteristics of MANETs are identified:

- Dynamic Topologies: Nodes are free to move arbitrarily. The network topology may change rapidly and randomly and contain both unidirectional and bidirectional links
- Bandwidth constrained, variable capacity links: After accounting for interference, noise, contention, the realized throughput may be much lower than a radio's maximum rate.
- Energy Constrained Operation.
- Limited physical security. Possibility of eavesdropping, spoofing, denial of service attacks.
- Networks may be large. This is described as tens or hundreds of nodes.

1.1.2 Mesh Networks

Mesh networks are multi-hop wireless networks at the *fringe* of the internet. They exist almost exclusively to provide internet access. The mesh network scenario is discussed in more detail in section 4.3.3.

1.1.3 Zeroconf

An ad-hoc network may contain nodes from multiple administrative domains. A successful ad-hoc network will require as little configuration as possible for operation. The IETF has a Zeroconf working group who work to enable Zero Configuration IP networking. [Wil03] defines requirements for zero-configuration of IP networks:

“A zeroconf protocol is able to operate correctly in the absence of configured information from either a user or infrastructure services.... ...benefits of zeroconf protocols over existing configured protocols are an increase in the ease-of-use for end-users and a simplification of the infrastructure necessary to operate protocols”.

In particular, the Zeroconf working group defines standards for address autoconfiguration, naming services, service location and multicast operation in IP networks without any pre-existing infrastructure or administrative effort.

1.2 Document Structure

In section 2 we explore the various routing and mobility protocols that may make up an ad-hoc internet connectivity solution. In particular, section 2.1 looks at the 802.11 standard for wireless networking and section 2.3 examines ad-hoc routing protocols.

Section 3 explores research into connectivity for ad-hoc networks.

Section 4 examines mobility models and usage scenarios for wireless networks that might require internet connectivity.

Section 5 examines other research which may be relevant to this area.

Section 6 identifies some interesting areas for future research in the area.

2 Background

2.1 IEEE 802.11 - Wireless Ethernet

IEEE 802.11 is the most widely available wireless networking system currently available. Since Intel's new chipset for mobile computing (*Centrino*) includes an integrated 802.11b interface, access to 802.11 networking will probably be widely available in the future. The IEEE 802.11 specification consists of both a physical-layer specification and a medium-access-control (MAC) sublayer.

2.1.1 Physical Layer

The 802.11 physical-layer specification provides for radio (unlicensed band) and infra-red transmission. 802.11 originally (1997) specified radio transmission at 1Mb/s or 2Mb/s. In 1999, the physical-layer standards 802.11a and 802.11b were released. 802.11b operates in the 2.4Ghz band at 5.5 or 11Mb/s, whereas 802.11a operates in the 5Ghz band at up to 54Mb/s. 802.11b is easier to implement, and so has become widely available earlier [Sta01].

2.1.2 Media Access Control Layer

802.11 provides two different modes of operation at the MAC layer. These are described in [CWKS97]. There is a *Distributed Coordination Function* (DCF), that allows for an infrastructureless network. No central control is required for media access while using DCF. It is essentially a carrier sense multiple access with collision avoidance (CSMA/CA). A radio interface cannot detect collisions itself while sending, so a positive acknowledgment scheme is used.

The collision avoidance scheme with positive acknowledgments may suffer from the *hidden node problem*: a hidden node is one that is close enough to the destination of a packet to interfere with it, but far enough from the sender that it does not hear it being sent (and hence does not know to avoid transmitting). This problem can be dealt with in 802.11 using a *virtual carrier-sense* mechanism: a node wishing to send a packet first sends a *Request-to-Send (RTS)* packet. The recipient node then replies with a *Clear-to-Send (CTS)* packet. Any node hearing either of these packets will then update their *network allocation vector (NAV)* and will not transmit for the duration specified in the RTS or CTS.

The virtual carrier sense mechanism can deal with the hidden node problem. But wireless packet networks also face the *exposed node problem* ([XS01]): nodes close enough to the sender to hear it's packets and RTS, but far away enough from the destination that they cannot interfere with the packet. Exposed nodes can lead to underutilization of the available bandwidth. This is discussed more in Section 2.2.

At the MAC layer, automatic retransmissions (up to seven) are used when a unicast packet is not acknowledged. However, broadcast packets in 802.11 are unacknowledged. They use neither positive acknowledgment nor virtual carrier-sense mechanisms. The error rate for broadcast packets that higher layers of the network stack experience may be much higher than those for unicast packets.

2.2 802.11 in multi-hop networks

2.2.1 Interaction of MAC layer with TCP

[XS01] examines the behavior of 802.11 in a multi-hop network. The interaction between the MAC protocol and the TCP (Reno) protocol is analysed. It is found that the 802.11 MAC protocol functions poorly in a multi-hop environment.

The multi-hop network analysed is a simple string topology. Each node can only communicate with its adjacent nodes. The DSR routing protocol is used for multi-hop route finding between nodes. The performance of file transfers between pairs of nodes is analysed.

It is found that the performance of TCP in this scenario is very unstable. The throughput of a single TCP socket will drop to almost zero very regularly. Reducing the maximum window size used by TCP can alleviate this problem, and keep the throughput stable. The authors' analysis suggests that this performance problem is due to the interaction of the 802.11 MAC protocol and exposed nodes. The interfering and sensing range in 802.11 may be more than twice the size of the communication range (this is the case in the ns-2 simulation of WaveLAN). If a node sends a number of packets sequentially, the re-transmission of the first packet is likely to be interfered with by the subsequent packets. Reducing the window size means that packets transmissions are spaced out in time, which avoids this problem.

It is also found that serious unfairness exists between multiple TCP streams within a multi-hop network. A TCP connection between adjacent nodes can be completely blocked by another TCP connection involving neighboring nodes. The binary exponential back-off scheme always favors the latest successful node.

2.2.2 Performance Measurements: Loss Rates

In [DACM02b] measurements from a real-world multi-hop wireless network are presented. It is found that multiple routes with the shortest hop-count may exist, but that the reliability of these routes can vary widely. Therefore, a choice of route based purely on hop-count is unlikely to choose the best route between two points. It is claimed that most existing ad-hoc protocols assume a bi-modal distribution of link quality (links are either very good or very bad), but that in reality the distribution is spread out.

In the network examined in [DACM02b], it is found that the best 40% of link pairs deliver at least 90% of their packets. It is also found that although the quality of the link in either direction are often highly correlated, at least 30% of the link pairs have a difference of more than 20% between the delivery rates in each direction. The quality of a link may also be quite variable with time (even though the network tested was stationary).

The strength of signal measured by a receiving node does not have a good correlation with the delivery rate. The authors suggest that signal to noise ratio might be useful as a fast predictor of delivery rates, but that it was not available from the radio hardware they tested.

In [DACM02a], it is found that the delivery rates on many routes are high enough that existing ad-hoc protocols will use them, but low enough that performance can be considerably sub-optimal. This reflects the assumption in these protocols that the presence of a link is a bimodal property, which is not supported by real wireless packet networks.

[DACM02a] goes on to propose a possible path metric that could be used as an alternative to hop-count when choosing multi-hop routes in a wireless network: the expected number of transmissions (including retransmissions) along a path. Similarly to hop-count, this is an additive metric. It is proposed to calculate¹ this from the delivery rate. The delivery rate can be provided by measurement or predicted based on signal strength and quality.

¹If the forward and reverse delivery rates are r_f and r_r respectively (possibly being dependent on packet size), then the expected transmission count is $1/(r_f \times r_r)$

2.2.3 Theoretical Capacity of Ad-Hoc Networks

[LBD⁺01] gives a simple analysis of the chain topology (as used in [XS01]). Labeling the nodes in a chain 1,2,3 etc., with nodes 200m apart, sending range of 250m and interference range of 550m. While 1 is transmitting to 2, neither 3 nor 4 can transmit without interfering at 2. This gives a maximum achievable utilisation of $\frac{1}{4}$. In practice, 802.11 achieves a throughput corresponding to a utilisation of $\frac{1}{7}$. 802.11 schedules it's packets using virtual carrier-sense and exponential back-off, which fails to achieve near the ideal scheduling of packet transmission.

[LBD⁺01] analyses the performance of 802.11 scheduling in regular chain and lattice topologies with regular data patterns, and in a random network with random traffic. These analysis are performed in simulation, but the results for the chain topology were verified with a real radio network. It is found that the exponential back-off of 802.11 can result in nodes wasting as much as 5.4% of it's time in long back-offs due to hidden nodes. Nodes at the edge of the network tend to have more capacity than interior nodes, and hence send more packets than the interior nodes can forward. This results in increased contention and packets being dropped.

The absolute limit on total one-hop capacity of an ad-hoc network is given by it's geographical area and the transmission and interference ranges of the network interfaces. Adding more nodes to a network while keeping the area constant only increases the contention for media access. The total capacity available to a node when multi-hop routing is in place decreases both with average path hop-count, and the number of nodes in the network.

In a random network, it is shown that the average path length varies with the square-root of the network area. If the density of nodes in the network is constant, this implies that the capacity available to each node is proportional to $1/\sqrt{n}$, where n is the node count. (A global scheduling achieving $1/\sqrt{n \log n}$ was demonstrated in [GK99]). If traffic patterns are taken into account, it is shown that pure-local traffic gives a capacity independent of network size, and will reduce to $1/\log n$ for a power-law distribution of correspondence with an exponent of -2 .

2.3 Ad-hoc Routing Protocols

As outlined in [RT99], routing protocols may be broadly classified as table-driven (pro-active) and source-initiated (on-demand). (It may be useful to add a third classification for routing protocols which utilise location - i.e. GPS co-ordinates). This section outlines the main ad-hoc routing protocols that are used in the research on connectivity for ad-hoc networks.

2.3.1 Pro-active protocols

Destination-Sequenced Distance-Vector Routing DSDV ([PB94]) is some of the earliest work on ad-hoc routing protocols. The authors note that existing routing protocols for fixed networks exhibit some of their worst-case performance in a highly dynamic interconnection topology: they have a heavy computational burden and poor convergence characteristics. They also note that there are significant differences in a wireless medium to a wired medium, i.e. mobile computers may have only one network interface but still be used to connect two separate networks.

The authors propose to extend a classical Bellman-Ford algorithm with destination-assigned sequence numbers to avoid formation of routing loops. A route table at each of the nodes lists all available destinations and the number of hops to each. The route table also contains the next hop for packets to that destination as well as the sequence number of the route advertisement. Each node periodically broadcasts it's route table to it's neighbors. This broadcast contains that nodes current sequence number.

A node receiving routing table information will merge it with it's own routing table. A route with a larger sequence number will replace an older route, as will a route with a shorter hop-length (the hop-length of a route is increased by one before storing in the route table). An exception to this is a hop-length of infinity, which signifies a broken route. An entry may only be made in the route table for a neighbor that shows that it can receive packets from the node. Hence, DSDV uses bidirectional links only.

Nodes broadcast routing table information periodically, and in response to changes. A node may either broadcast a *full dump* or an *incremental update*. Nodes keep track of the average settling time of routes (the time between the first route with a new sequence number and the shortest route), and routing update broadcasts can be delayed by the settling time to avoid sending unnecessary traffic.

DSDV allows for operation at either layer 2 or layer 3. It is proposed that if operating at layer 2, a node would advertise which layer 3 protocols it supports and some information (i.e. IP address). This information would only need to be propagated in routing updates when it changed, which would be infrequent. (Authors note that layer 3 operation violates the normal subnet model of operation, but is compatible with the model of operation offered by the IETF Mobile IP Working Group)

2.3.2 On-demand protocols

Dynamic Source Routing Dynamic Source Routing (DSR - [JMB01, JM96]) is an on-demand routing protocol designed for multi-hop wireless networks. Nodes discover *source routes*, a complete route from source to destination. The complete route is added to each packet being sent. Using source routes allows the loop-free property to be trivially implemented. Nodes forwarding or overhearing packets can easily cache the routing information they contain for future use.

Routes are discovered by flooding a route request. Replies to this request also require a route back to the originator, which may require a further flooding of route request. Although DSR is designed to work with uni-directional links, for MAC protocols which limit unicast packet transmission to bi-directional links (such as MACAW or 802.11), DSR can eliminate the second route discovery by reversing routes.

Using promiscuous receive, DSR supports quite aggressive caching of routes and automatic shortening of routes. There is also scope for caching negative information about intermittent links.

As described in section 3.1.2, DSR is designed for use with multiple network interfaces, and also supports advertisement of Internet Gateways and Mobile IP through the routing protocol.

A *route reply storm* is the situation where many neighbors of a node sending a route request have a cached route, and their simultaneous replies cause heavy media contention. To avoid this situation, nodes replying to a route request pause before replying. The length of the pause is related to the hop-length of the route being returned. While pausing, the node enters promiscuous receive and cancels it's reply if it hears another reply.

Ad Hoc On-Demand Distance Vector Routing AODV ([Per97]) is one of the best-studied ad hoc routing protocols in the literature. Like DSDV, it is a distance-vector routing protocol. However, AODV operates purely on-demand.

AODV's route discovery is a broadcast-based method. Destinations maintain a sequence number, similar to that used in DSDV. Broadcast packets also have a unique identity to avoid duplicates. AODV operates only on bidirectional links and the reverse path to a node is set up automatically during the propagation of route request (RREQ) by including the source's sequence number. If the route request reaches the destination, or an intermediate node with a route to the destination, a route reply (RREP) is unicast back to the source, and the appropriate routing table entries are made at the intermediate nodes. The route request may also contain a destination sequence number to specify how fresh a cached route must be to be accepted.

Each node maintains a routing table, which contains a subset of nodes in the network. Each entry contains the next hop on the route, a hop count to destination, and the destination sequence number. The route table also tracks how many of it's neighbors are using the route, so that it may expire route entries after a period of inactivity.

Broken links can be detected using information from the link-layer, and also using periodic *hello* messages. A broken link will result in a RREP with a hop count of ∞ . Nodes which have not sent any packets to all of it's active downstream neighbors for a *hello interval* will broadcast a

hello message with its current sequence number. A node missing a *hello* message from an active neighbor for a number of consecutive *hello intervals* will consider the link broken.

Further work on AODV has demonstrated how to implement multicast routing within a multi-hop wireless environment.

3 Ad-hoc Connectivity

3.1 Prior Work

This section presents an overview of research in the area of providing internet connectivity to ad-hoc networks.

3.1.1 Lei/Perkins - DSDV and Mobile IP

In [LP97] a scheme is proposed to integrate a Mobile IP implementation with a modified RIP routing protocol. This paper presents a mechanism that extends foreign agent coverage to a whole ad-hoc network instead of only being available to nodes in direct contact with the foreign agent. The modified RIP protocol being used is very similar to DSDV - [PB94].

The Foreign Agent in this scheme participates in the routing protocol of the ad-hoc network. This enables unicast routing between the Foreign Agent and every mobile node on the ad-hoc network. A mechanism is also required by which nodes may effect agent discovery. This may happen by agent advertisements or agent solicitations. Agent advertisements are piggy-backed onto the routing entry for the foreign agent, which causes them to be delivered to each node.

Most of the details of the paper involve the co-ordination of updates to the routing table by the separate Mobile-IP and RIP daemons. This is achieved by introducing a third, *route-manager daemon* to co-ordinate routing table updates.

3.1.2 Broch/Maltz/Johnson - DSR

In [JMB99, JMB01], the authors describe their efforts on integrating Dynamic Source Routing (DSR - [JM96]) with heterogeneous networks. The authors propose a number of mechanisms to achieve this.

Firstly, the situation where nodes may have multiple network interfaces, or where the network as a whole may have heterogeneity among its network interfaces is examined. The authors propose a *logical addressing model*. Under this model, each node has a unique identity, and its various interfaces are assigned an index which is locally unique. This scheme has been adopted by the IETF ([CM99]).

The interface index values are arbitrary except for two special cases. Special interface indices are reserved to act as logical identifiers for services which a node may provide. Two services are specified: a node that may act as a gateway to the internet, and a node may act as a Mobile IP home or foreign agent (termed a *mobility agent*). This allows these services to be effectively advertised to the whole network via DSR.

When a DSR node acts as a gateway, it will respond to route requests for addresses on the internet listing itself as the second-to-last hop. When it receives packets with these routes, it will then act as a proxy for the ad-hoc node.

The history of DSR is interesting. The design grew from an elaboration of the Address Resolution Protocol ([Plu82]) to a multi-hop environment. In [JMB01] the authors discuss the siting of the protocol within the ISO stack. They consider whether such a protocol should be placed at the link-layer (ISO layer 2) or the network layer (ISO layer 3). The authors had originally intended to do routing at the link layer for a number of reasons:

- Running at the link layer would allow IPv4, IPv6, IPX and other network protocols to take advantage of DSR, and maximise the potential number of nodes that can participate.
- DSR's design grew from that of ARP, which is based at link-layer level.

- DSR was designed to be implementable within the firmware of a wireless network interface, and hence operate below the operating system's network layer software.

Although the authors decided to implement DSR at layer 3 so that the routing protocol could support nodes with multiple network interfaces, the FreeBSD implementation makes the DSR available as a virtual interface so that higher layers may essentially treat it as if it were implemented at Layer 2.

3.1.3 MIPMANET

MIPMANET ([JAL⁺00, JA99]) presents an integration of AODV ([Per97]) with Mobile IP ([PM94]). Some assumptions are made:

- *“nodes within an ad-hoc network should not have to make any assumptions about their network ID's”*. This means that a node cannot decide whether a destination is within the ad-hoc network simply by looking at it's address. This point is made to support a zero-configuration setup.
- related to the previous point, the authors assume that nodes must use IP layer routing to reach a gateway to the fixed Internet.

Mobile IP must be adapted for use within an ad-hoc environment. The nature of the ad-hoc environment impacts on mobile IP:

- Mobile Nodes and Mobility Agents need to use multi-hop communication.
- Broadcasts are expensive and should be minimized
- Link-layer information about connectivity to mobility agent must be replaced with routing protocol information to enable movement detection and cell switching

In response to these requirements, agent advertisements are reduced from 1 second to 5 second intervals. The MIPMANET Cell Switching (MMCS) algorithm is proposed, whereby a node switches to a new Mobility Agent if it is at least 2 hops closer than it's current agent for two consecutive agent advertisements. MIPMANET uses reverse tunneling in it's Mobile IP setup, although this is not mandatory.

Integration of Mobile IP and the ad-hoc routing protocol is done using a MIPMANET Interworking Unit (MIWU). The MIWU looks to the Mobile IP Agent like *“a visiting node that is registering different IP addresses, but with the same link-layer address... All ad hoc routing functionality can be put in the MIWU”*.

MIPMANET is evaluated using ns2 ([Fal00, Mon98]). 15 mobile nodes are simulated plus two foreign agents, one on each side of the flat area (1000m x 500m). The random way-point model is used for mobility and the traffic pattern is constant bit rate (CBR) between a wired and a wireless node. The simulation results suggest that periodic agent advertisements are worth the overhead since they improve performance by causing shorter routes between mobile nodes and foreign agents (without advertisements, a node will stay with a foreign agent until it's link breaks).

The MIPMANET Paper ([JAL⁺00]) makes a number of criticisms of the work outlined in sections 3.1.1 and 3.1.2. Their main criticism of [LP97] is that it uses a pro-active routing protocol, and cannot be easily adapted to use an on-demand protocol. Some issues that are not dealt with by the work on DSR ([JMB99, JMB01]) are suggested:

- movement detection
- handoff
- choosing between FA's
- agent advertisements

A number of areas for future work are also identified:

- Dynamic address allocation
- co-operation between internet access points
- Should the mobility agent discovery be integrated with the routing protocol, rather than layered as in this work
- Use of multicast for agent advertisements
- Some pro-activeness may be beneficial (i.e. broadcast agent advertisements are pro-active)

3.1.4 Perkins/Sun/Belding-Royer

[PBR02] proposes a scheme similar to that of [JAL⁺00]. However, in this scheme, more changes are made to the routing protocol in order to efficiently support Mobile IP. A well-known multicast group address, the *All Mobility Agents* address ([Per96]) is used in an AODV Route Request when a mobile node wishes to use agent solicitation. Foreign Agents may also respond to Route Requests for addresses on the internet with a special *FA-RREP*. The scheme uses MIPMANET Cell Switching to decide when to switch Foreign Agents.

The evaluation is very similar to that of MIPMANET. Again, the number of nodes is modest (10, 20 and 50), a random way-point mobility model is used, and the traffic is constant bit rate between wireless and wired nodes. The parameters of agent advertisement interval, node mobility and the number of foreign agents are varied and some performance characteristics measured (packet delivery fraction, average latency, routing and Mobile IP overhead).

It is shown that adding extra foreign agents in this scheme improves packet delivery latency and shortens the average path length. The shorter path lengths cause a slight improvement in delivery fraction and reduces the AODV overhead.

3.1.5 MEWLANA

Mobile IP Enriched Wireless Local Area Network Architecture (MEWLANA - [EP02]) presents two different routing protocols that are designed around Mobile-IP based internet connectivity for ad-hoc networks. MEWLANA-TD is a table driven routing protocol, and MEWLANA-RD is a tree-based ad-hoc routing protocol.

It is proposed that the routing protocol for an ad-hoc network be chosen based on the size of the ad-hoc network and the fraction of traffic that is purely internal to the ad-hoc network. It is claimed that MEWLANA-RD is more suitable when a network is mostly used for access to fixed services. MEWLANA-TD and MIPMANET would be more suitable for networks with mainly internal traffic, with MIPMANET being better for larger networks.

The table driven approach uses DSDV, so every node maintains routing information for the whole network. This routing information is used to limit the spread of foreign agent beacons (beacon only needs to be sent when a new node joins the network).

The MEWLANA-RD scheme uses an ad hoc protocol called *Table Based Bidirectional Routing* (TBBR). Routing table formation is done only with MIP entities and no additional ad-hoc protocol is used. The protocol aims to serve mainly outside traffic. The routing table is formed from agent advertisements and registration messages, and is repeated after each registration renewal interval.

TBBR uses a *Depth Level Number* (DLN), acquired from the hop-count of agent advertisements. A node only processes advertisements with hop-count smaller than their DLN. Registration requests travel from the leaves of the tree to the root (foreign agent), and establish routes from the FA to each of the mobile nodes.

The two MEWLANA schemes are evaluated against the MIPMANET proposal. A Performance metric is defined as the sum of the reciprocals of Mobile IP Overhead, Ad Hoc Routing Overhead, and the Number of Hops to route inside traffic. The scenarios examined are for between 4 and 128 nodes with up to 10 nodes participating in inside traffic. The results with this metric support the

assertions above about the suitability of the three protocols to various situations. The evaluation seems a little strange: there does not seem to be any external traffic involved, and the metric chosen is not justified.

3.1.6 LUNAR

Lightweight Underlay Network Ad-Hoc Routing (LUNAR -[TG02]) presents a somewhat different approach to ad-hoc routing. With simplicity in mind, the protocol is designed for the “*small common case*”: 10-15 nodes and network diameters of not more than 3 hops. The code size should be small, without many subtleties. Operation should be simple, i.e. LUNAR should have a default profile which will work without special configuration.

The LUNAR approach is that of an *underlay network*. Ad-hoc path establishment is linked to the usual address resolution activities going on between the network layer and the link-layer. This is similar to the approach of DSR, which developed from multi-hop ARP. A virtual logical subnet is created on top of the underlying multi-hop, ad-hoc network. This allows IP to treat the ad-hoc network as if it were a physical subnet.

LUNAR opts for a decentralized approach for IP address configuration, as an alternative to DHCP. Nodes choose a random address and probe for conflicts in a manner similar to that proposed by the IETF’s zeroconf working group in [CAG02].

A node with a connection to the internet can share it’s connection with other nodes by responding to resolution requests for internet addresses, and employing Network Address Translation (NAT).

The routing protocol used by LUNAR is on-demand, with routing paths the responsibility of the sender. The current implementation relies on bidirectional links. Paths are established in response to ARP or broadcast-send requests, and are torn down after 3 seconds. This means that paths are completely rebuilt every 3 seconds. Any packet losses are left to the transport layer to deal with.

Due to lack of availability of stable linux implementations of AODV, DSR and TORA, LUNAR was only compared with Optimized Link State Routing (OLSR - [CJ03]). It was found to behave only slightly worse than OLSR even though it’s code is less than one-third the size of OLSR and includes address configuration and internet connectivity. The scheme was evaluate using the Ad Hoc Performance Evaluation Testbed ([LLN⁺02]), a real-world testbed.

3.1.7 Hybrid Proactive and Reactive Mobile IP

[RK03] proposes a hybrid scheme for agent advertisement and solicitation based on the work of [PBR02]. Agent advertisements (pro-active) are scoped with a small time-to-live (TTL) which allows them to be broadcast to nodes within a small distance of the foreign agent. Nodes outside this distance use an *expanding ring search* to try and locate a foreign agent. After a successful ring search, an agent advertisement is unicast to the mobile node. The scheme also employs caching and eavesdropping of agent advertisement and registration methods in order to reduce overhead.

The evaluation of this scheme shows that a hybrid approach to agent advertisements may reduce both Mobile-IP and AODV overhead overhead. The optimal TTL for advertisements varies with network size and density.

3.1.8 LocustWorld MeshAP

LocustWorld ([Loc03]) produce software and hardware for *mesh networking*. Their software allows people to cooperatively share internet access. Nodes in the mesh provide a service to their users.

A node is a dedicated computer with a wireless network interface. It may also have a connection to the internet, via DSL for example. A node is designed to provide internet connectivity to local users via a wireless or wired network. If a node has wired internet access, it will share this access with other nodes in the mesh over multiple hops.

Initially, a node allocates itself a random address in the class A 10.0.0.0 private address range. It attempts to find an internet gateway over it’s Ethernet gateway. If no gateway is found, the

node considers itself a repeater-cell. The node then starts an internal DNS and transparent web proxy. The node also picks a random class C private address range (192.168.128.0-192.168.254.0). The node allocates itself an address in this range and runs a DHCP server advertising itself as default gateway and DNS server.

AODV is used to find routes to gateway nodes within the mesh. A bogus address is used to indicate internet access. A node sends a route request for this bogus address, and gateway nodes will reply to this request. Once a repeater cell has a route to a gateway node, it sets up an encrypted IP-tunnel and uses this to forward IP traffic.

Once a repeater node has a tunnel set up, it will serve the gateway node's address as default gateway and DNS server. It will transparently proxy web and DNS requests to itself to the gateway node. Network Address Translation is used by the nodes to supply connectivity to clients. HTTP and FTP traffic is transparently proxied also.

4 Analysis of the Research

4.1 Mobility Model

In [MMPS00] the random way-point model commonly used with the Monarch extensions to ns-2 is examined. The authors analysis concludes that in this model you are more likely to travel to a distant point than to a nearby point. However, it has been observed ([Kle00]) that interpersonal associations are localised - nodes commonly travel short distances more frequently than they do large distances. This assumption is one that ad-hoc routing protocols should exploit.

[MMPS00] goes on to propose a *Kleinbergian model* for mobility: a node chooses an angle θ and a distance d to travel to it's next point. This provides increased locality. By choosing d not uniformly, but using some distribution that favours shorter distances, further locality can be provided.

These mobility models are used to evaluate DSDV, AODV and DSR. The original mobility model is compared to the Kleinbergian and the localised Kleinbergian. Interestingly, it is found that the performance of the protocols is almost identical across these mobility models. This suggests that the routing protocols do not exploit the locality of interpersonal associations (although the analysis here is quite short, and uses node speeds up to 200m/s!)

[CBD02] makes an analysis of the random way-point model, measuring the *average neighbor percentage* of all nodes in the network over time. It is found that if the nodes are initially randomly distributed and subsequently perform random way-point model that the average neighbor percentage varies a lot for the first 600 seconds of the simulation. After this initial period, the average neighbor percentage remains fairly constant (this is because nodes are more likely to be around the centre of the simulation to travel between two random points). [BMJ⁺98] for example uses a 900 second simulation, so this phenomenon may affect the results.

4.2 Network Scenario

The scenario in which a network will be used will determine which routing, mobility, and internet-connectivity solutions are most appropriate. We identify here some of the parameters of a network scenario that are most important when considering Internet connectivity for ad-hoc networks. Section 4.2.1 describes IETF's approach to network scenarios in mobile wireless networks. Section 4.2.2 describes some additional parameters of network scenarios where internet connectivity is required.

4.2.1 MANET *networking contexts*

In [CM99], the IETF present the idea of a *networking context*, and identify a number of parameters of the networking context which can be varied:

- Network size

- Network connectivity - the average degree of a node
- Topological rate of change
- Link capacity
- Fraction of unidirectional links
- Traffic patterns - non-uniform or bursty traffic to be considered
- Mobility - how does temporal and spatial correlation affect a routing protocol?
- Fraction, frequency of sleeping nodes.

4.2.2 Additional Parameters for Internet Connectivity

In the scenario of a mobile wireless network with internet connectivity, there are a some other parameters which may be relevant:

- Amount of internal traffic: what fraction of traffic is purely internal to the network, and what fraction is between the fixed network and the wireless network?
- Static/Semi-static nodes: At least some nodes (the internet gateways) will be static. Some fraction of other nodes in the network may be static or semi-static.

4.3 Example Network Scenarios

4.3.1 Random Scenario

Much of the literature evaluating ad-hoc routing protocols operates using a very simple scenario (because it is easily implementable in ns-2):

- 2-d, rectangular area
- No obstacles
- Bi-directional links
- Fixed number of nodes
- Nodes operational for whole simulation
- Random-waypoint mobility model. Nodes pause for a random amount of time before picking a point uniformly from the simulation area. Node then moves to new point at a random speed up to 20 m/s (72 km/h). Some evaluations use a fixed rather than a random speed.

This (or simple variations) are used for evaluation in [BMJ⁺98], part of the Monarch² project at CMU. [BMJ⁺98] is one of the main performance evaluations of ad-hoc routing protocols in the research. CMU implemented the Wireless and Mobility Extensions to ns-2 ([Mon98]), which are widely used for performance analysis.

²MOBILE Networking Architectures

4.3.2 Scenario-Based Performance Analysis

In [JLH⁺99], parameters are given for simulations of three example scenarios:

- Conference Auditorium: network access so speaker may share data with the audience
- Event Coverage: Reporters at a political or sports event, or stock-brokers at a stock exchange.
- Disaster area: Rescue operation at a natural disaster site.

They present parameters for each scenario: transmitter range, environment size, node mobility, traffic type and patterns. For instance, the conference auditorium is divided into three zones: the stage (with the speaker walking back and forth), the audience (who are fairly static), and the entrance (where people come and go). In the event coverage, the speed of node movement is fairly small (1 m/s), and clusters of around 10 nodes are formed spontaneously. In the disaster area, there are three groups of nodes that are connected only by vehicles (helicopters, cars) which are moving around quite quickly (20 m/s). There are multiple network partition events in the disaster area.

In each of the scenarios, there are a number of constant-bit-rate (CBR) sources and a number of receivers, giving a number of concurrent CBR flows. Each flow sends 512 byte packets 4 times a second.

The choice of some of the parameters are a bit strange (e.g. conference has transmitter range of 25m, somewhere between bluetooth and 802.11), but provide some results that are more meaningful than the random model.

4.3.3 (Sub)Urban Mesh Network

The sub-urban mesh network is one of the type envisaged by LocustWorld's MeshAP (see section 3.1.8). Many people - who don't necessarily know each other (multiple administrative domains) - co-operate to provide internet access to a suburban area. Nodes are typically very static. Some nodes will have internet connectivity that they are willing to share with others over the mesh network.

Mesh networking is a popular term ([Eco02]) for multi-hop wireless networks used to solve the *last-mile problem*: how to provide broadband connections to homes without running cables directly to each subscriber. At the moment, the main option is to reuse the telephone or cable-TV networks. Mesh networking is an approach being proposed by both commercial and 'open-source'/community movements.

In the commercial model, a *neighborhood access point* (NAP) is installed, which is a radio base station with a high-speed internet connection. Subscribers to the service install their own wireless node to gain access to the NAP. Once this node is installed, it can also act as a relay to extend the effective coverage of the NAP.

4.3.4 WAND

The Wireless Ad-hoc Network for Dublin (WAND), is "*a collaboration between the Dynamic Interactions Group at Media Lab Europe (MLE) and the Distributed Systems Group at Trinity College Dublin (TCD) and aims to realise a wireless network research testbed for new types of wireless applications running on ad hoc networks*".

The initial deployment of WAND will be a string of nodes, mounted on traffic lights between TCD and MLE. These nodes will be equipped with 802.11 network interfaces, and close enough to each other to communicate. In the future, the nodes may have dedicated high-speed internet access, or high-speed wireless communication among themselves.

The backbone of the WAND network is intended to act as a starting point for the network to grow. There will probably be two types of users of the network:

- Fixed nodes. Businesses and residents within range of WAND may choose to join the network. They will be able to access the services of the network, but also extend the coverage and improve the infrastructure of WAND.
- Mobile Users. People using PDAs, mobile phones or laptops equipped with 802.11 may temporarily join the network and access its services. These users may be mobile while they are in the network. These mobile nodes will also form part of the network infrastructure while they are present.

It is anticipated that much of the traffic will initially be the use of WAND for internet access. Nodes on the network will act as HTTP, FTP, email clients (POP/IMAP/SMTP), and possibly to access computing systems (ssh). There may also be usage of software such as instant messenger (which, interestingly provides an application-layer macro mobility protocol). These usages have in common that they can all work with client-initiated TCP connections (so nodes should not have a need to act as servers) and they all use WAND as a *stub network* (see section 1.1.1).

With such a network in place, however, other usages may evolve around the network. Obvious examples would be networked computer games between people on the network, or peer-to-peer file sharing. Less obvious might be facilities such as printing, scanning, or file backup provided over the network. These services might require nodes on the network to act as servers and use more intra-WAND traffic.

4.4 Performance Evaluation

Performance evaluation of a routing, mobility or ad-hoc connectivity solution can be broken down into two main approaches: simulation and imperical study. By far the most popular method of evaluation in the literature is simulation.

4.4.1 Network Simulation

As described in Section 4.3.1, ns-2 is usually used for simulation of wireless networks and evaluation of ad-hoc routing protocols. There are alternatives available, such as OPNET Modeler (commercial) and GloMoSim ([ZBG98]). However, as noted in ([CSS02]), the results of simulations in these three environments may vary widely (possibly due to the level of detail used in physical layer simulations) and should not be trusted to correspond with real-world results.

4.4.2 Real-world evaluation

APE³ ([LLN⁺02]) is a testbed for performing real-world protocol evaluations. It is essentially a small (~8Mb) linux distribution (which can be installed easily under windows). It provides for scripting and choreography for individual nodes, with data gathering performed at IP and physical layer. Using choreography can provide instructions to operators of the nodes on where to move physically. Support is also provided for simulating physical conditions by dropping packets at the MAC layer.

4.4.3 Performance criteria

[CM99] outlines some performance criteria for evaluation of ad-hoc routing protocols.

- End-to-end throughput and latency. Include statistical measures (means, variances, distributions)
- Route acquisition time.
- Percentage Out-of-Order Delivery.

³Ad-hoc Protocol Evaluation testbed

- Ratio of data bytes(packets⁴) transmitted / data bytes(packets) delivered
- Ratio of control bytes(packets) transmitted / data bytes(packets) delivered

4.5 Miscellaneous

4.5.1 Implementation

The Ad-hoc Support Library (ASL - [LLN⁺02, KZ02]) is a user-space library which provides an API to facilitate implementation of routing protocols for wireless ad-hoc networks in Linux. The authors argue that is preferable for on-demand route discovery to be performed outside the kernel, keeping computations that may be memory or CPU intensive out of kernel space.

The ASL library allows typical on-demand routing functions to be implemented on a standard Linux 2.4 kernel with no kernel modifications necessary. The Linux *Universal TUN/TAP* interface is used to pass packets requiring routes to a user-space daemon, and the packet filtering facility, *Netfilter*⁵ is used to monitor packet routing events.

Using ASL, the University of California at Santa Barbara implementation of AODV, AODV-UCSB was easily ported to run completely in user-space, with fewer packets needing to cross between kernel- and user-space.

5 Other Research

5.1 Artificial Intelligence / Mobile Agents for Routing

5.1.1 Swarm Intelligence for Routing

“Swarm Intelligence (SI) is the property of a system whereby the collective behaviours of (unsophisticated) agents interacting locally with their environment cause coherent functional global patterns to emerge. SI provides a basis with which it is possible to explore collective (or distributed) problem solving without centralized control or the provision of a global model.” [Ara02].

The term *Swarm Intelligence* is used to refer to systems whose design is inspired by models of social insect behaviour. Key characteristics of these models are:

- Large numbers of simple agents
- Agents may communicate with each other directly
- Agents may communicate indirectly by affecting their environment, a process known as *stigmergy*
- Intelligence contained in the networks and communications between agents
- Local behaviour of agents causes some *emergent global behavior*

Some research has focused on the use of swarm-intelligence type systems for routing within communications networks. Both [KESI⁺01] and [AGK⁺01] provide an overview of the main work in swarm intelligence as it applies to routing: AntNet and Ant-based Control.

⁴measuring packets (as opposed to bytes) gives some idea of channel access efficiency

⁵<http://www.netfilter.org/>

AntNet AntNet ([CD98]) is an adaptive, mobile-agents-based algorithm inspired by work on the ant colony metaphor. It has been found to out-perform the best-known routing algorithms on several packet-switched communications network.

In AntNet, each node keeps a routing table, which for each destination gives the probability of choosing each neighbouring node as the next hop. In actual network operation, the next hop with the highest probability is always chosen. Periodically each node will launch network exploration agents, called *forward ants* to every destination. At each node, the ants will choose their next hop probabilistically using that nodes routing table. As the ants visit a node, they record their arrival time and the node identity in a stack.

An ant reaching it's destination is converted to a *backward ant*. The backward ant pops the entries off it's stack and visits each of the nodes that the forward ant did. At each node along the return trip, the arrival time of the backward ant is compared to the arrival time of the forward ant. This gives a round-trip time to the destination over the route chosen by the forward ant. This round-trip time is compared to the average round-trip time to that destination. If the new round-trip time is smaller, the probability of choosing that route is increased. If the new time is larger, that route's probability is decreased.

Ant-based Control Ant-based Control ([SHBR96]) uses a very similar approach to AntNet, but designed specifically for telephone networks. Ants travel only in one direction and may be delayed at congested nodes. The effect of an ant's arrival at a node is decreased with the age of the ant. The use of a route for telephone calls may result in it's congestion, which will in turn cause the strength of routing entries involving congested nodes to decrease. The use of noise or jitter in the ant's movement decision is suggested to promote random exploration.

The ant-based control strategy was found to out-perform another (much more complex) mobile agents approach to load balancing from British Telecom's research labs. The ant-based strategy was better able to adapt to changing call patterns, both by reacting to them, and by choosing more robust routing strategies⁶.

5.1.2 Agent-based DVR

Agent-based Distance Vector Routing (ADVR - [AMM01]) alters a DVR routing protocol so that the routing messages exchanged become agents, which determine their own movement through the network: rather than being broadcast to all neighbors of a node, routing updates get transferred between nodes as a part of agent migration. A certain number of agents is active in the network at any particular time. The migration strategy of the agents in an ADVR approach needs to be chosen carefully to avoid looping and other negative side-effects.

A *Random Walk* is an agent migration strategy where agents simply select a node randomly from the neighbors of it's current node. It has been shown that, due to it's probabilistic nature, a Random Walk will visit all nodes and edges in a network (in infinite time).

A *Structured Walk* is a migration strategy that chooses an agent's destination based on some criteria (congestion, topological information, historical information). Three structured walk strategies are proposed: node-least-visited, edge-least-visited and *least-first walk*⁷.

Using a structured walk, ADVR can achieve convergence properties approaching that of a DVR protocol, but with significantly fewer messages. The messages that are sent in ADVR are much more efficient at distributing routing information.

5.2 Message Flooding and Broadcast Storms

Many of the approaches to ad-hoc routing and internet connectivity mentioned above utilise some form of flooding in order to disseminate messages throughout the network. However, the plain flooding algorithm may entail a large number of unnecessary packet rebroadcasts, which in turn

⁶One experiment allowed both approaches to run for a period of time. The routing tables were then frozen and their performance tested against changing call patterns for the remainder of the simulation.

⁷the agent chooses the destination node to minimise the combined visit-count of the edge and node

increase media contention and packet collisions and hence use up bandwidth. There has been some research into alternative flooding algorithms that may reduce unnecessary broadcasts.

5.2.1 The Flooding Problem

The flooding problem in the context of mobile wireless networks may be stated: a node wishes a message to be delivered to all nodes in the network. A node may *broadcast* a message, which will deliver it to all nodes within transmission range in the absence of message collisions at a receiving node. (In terms of 802.11, broadcasts are unacknowledged, and a RTS/CTS handshake does not apply).

The plain flooding algorithm is that the initial node broadcasts the message, which has a unique identifier. A node receiving a broadcast will in turn rebroadcast the message if it is the first time the node has received the message. Since receiving hosts are close to each other and the timing of rebroadcasts highly correlated there is a high probability of media contention. Also, the closer re-broadcasting hosts are to each other, the more redundancy exists in the rebroadcasts.

5.2.2 Broadcast Storms

[TNCS02] describes these problems with the plain flooding algorithm as the *broadcast storm* problem. Some simple analysis is made of coverage and contention for broadcasting and re-broadcasting in a 2-dimensional space with omni-directional antennas.

It is shown that a node receiving a broadcast can cover a maximum additional area 61% of the original area. The average additional coverage from rebroadcast is 41% of the original coverage area. The average additional coverage a rebroadcast may provide drops off sharply with each additional broadcast that a node receives. A node that has heard a broadcast twice will only provide 19% extra coverage on average. The expected additional coverage drops below 5% for a node that has received the broadcast 4 or more times.

A similar analysis may also be made for contention between rebroadcasts. For two nodes that receive a broadcast and decide to rebroadcast it, there is a 59% that their transmission areas overlap, and hence a 59% chance of contention between these two rebroadcasts. The probability of contention clearly rises with the number of receiving hosts.

[TNCS02] goes on to present a number of schemes whereby rebroadcast is inhibited in some nodes. These schemes aim to reduce broadcast redundancy, and hence contention and collision. Most of the schemes operate by attempting to maximise the utility of the rebroadcast, i.e. the additional coverage area or change of non-collision. All of these schemes use random delays between reception of a broadcast and rebroadcast to allow reception of duplicate messages and decide whether to rebroadcast.

Counter-based Scheme If a host attempts to rebroadcast a message, it may be temporarily delayed from doing so due to busy medium, back-off, or other queued messages. In this case, it is possible that the host will receive the broadcast one or more times before it gets the chance to start transmitting its own broadcast.

A counter-based scheme sets a threshold, C , for the number of times a message may be received and still re-broadcast. If a message is received more than C times, it is not rebroadcast. The initial rebroadcast is delayed by a random number of slots (the amount of time it takes to send a message). Messages received during this delay may cause the rebroadcast to be cancelled.

Distance-based scheme Related to the analysis of additional coverage of rebroadcasts given in 5.2.2 above, the greater the distance between the sender and receiver of the broadcast, the greater the additional coverage achievable by the receiver re-broadcasting. The signal strength of reception may be used to correspond to the distance. During the random delay before rebroadcast, if the node receives the message from another node which has its distance within a certain radius (corresponding to signal strength above some threshold), then the rebroadcast is cancelled.

Location-based scheme As an extension of the distance-based scheme, if nodes are equipped with GPS receivers they could attach their location to broadcasts. A receiving node could then calculate more accurately the additional coverage that it may provide by rebroadcast. (This is algorithmically hard, but acceptable approximations may be made).

Cluster-based scheme This scheme uses a clustered topology, such as one formed by the *cluster formation algorithm* of [JTT99]. Nodes which are not cluster heads or gateway nodes never re-broadcast messages. Cluster heads and gateway nodes may use some other scheme to decide whether to rebroadcast or not.

Evaluation The evaluation of the various flooding schemes show that a significant number of rebroadcasts may be eliminated without seriously affecting the reachability of the flooding. The location based scheme is the most effective, offering the best combination of reachability and rebroadcast savings. The balance between reachability and rebroadcast savings may be adjusted by choice of scheme and parameters.

5.2.3 Probabilistic Broadcast and Phase transitions

In [SCS02] a simple probabilistic broadcast scheme is examined. Each node has a probability p of re-broadcasting a message. The authors present some elementary results from the theory of random graphs and percolation theory. For these simple models, there is a phase transition phenomenon where a critical value of p exists. Below this critical value the flooding is non-global, but above this value there is a high probability of the flooding to reach the whole network.

However, the simple models do not apply directly to the situation of a mobile wireless network. The bimodal behaviour of percolation theory and random graphs is not found. The results for wireless networks do tend towards this situation with higher densities of nodes, however. For lower node densities, the success of broadcast is roughly linear with rebroadcast probability. Success rates of 90% can be achieved for rebroadcast probability as low as 0.65 in small/dense networks.

The authors note that there is scope for further work in this area. Varying the rebroadcast probability based on local graph topology information could improve performance. There is also scope to vary the nodes' transmission power in relation to p .

5.2.4 Gossip based Ad Hoc Routing

Gossip-Based Ad Hoc Routing ([LHH02]) applies probabilistic broadcast (Section 5.2.3) to route finding within AODV (see 2.3.2). Route requests are flooded using an alternative probabilistic scheme. A number of extensions to simple probabilistic broadcast are proposed:

- Simple probabilistic broadcast is called GOSSIP1(p). p is called the *gossiping probability*.
- Set $p = 1$ for the first k hops. Call this GOSSIP1(p, k)
- Set an increased gossiping probability for nodes with few neighbors. Specifically, GOSSIP2(p_1, k, p_2, n) uses gossip probability p_2 if a node has fewer than n neighbors.
- Listen for neighbors re-broadcasting a message. If a node has n neighbors, and overhears less than $m = pn$ rebroadcasts, then it should retry it's broadcast.

These variations are found to be able to reduce control traffic by up to 35%. The routes found by gossiping may be 10-15% longer than those found by flooding. (However, in dense networks a higher incidence of collisions often causes flooding to find longer routes also.)

5.3 Service-based Routing

Service-based routing is a relatively new area of research in ad-hoc routing. It is based on the principle that request packets are addressed to anonymous services rather than IP addresses. A service-based routing protocol can be built on top of an AODV layer [JW01], but the more interesting approach is to replace the IP routing layer with a service-based routing layer. Similar to AODV, some kind of flooding operation is required (section 2.3.2) but the possibilities for limiting the scope of the flooding are much increased as services can be replicated and distributed throughout the network, thus limiting the potential spread of a request. In theory, this should allow such networks to scale as routing request messages should be limited in their broadcast scope to the distance of the nearest replicated service. However, service request messages can include quality of service (QoS) criteria which would increase flooding in the network, so a trade-off between support for QoS and overall network efficiency must be managed. Services can be identified by a name and a set of service criteria, with examples including port numbers and service descriptors [VGPK97].

The service-based routing paradigm is particularly suitable to stigmergic routing approaches as service requests are anonymous and thus all requests for the same type of service will lay down identical pheromone trails. Stigmergic routing can be used to solve the optimisation problem of finding the shortest path to a service from any particular node. Service requests and replies are cached by intermediate nodes in routing tables with accompanying "pheromones", and based on pheromone trails the shortest path to a particular service (service request), as well as the shortest return path to a particular node (service reply), should emerge after a critical level of service request packets has been routed through the node.

Another problem of distributing replicated services throughout the ad-hoc network can be tackled either by having some static notion of the network topology or by employing a distributed algorithm that where nodes take independent decisions on which replicated service to provide. The independent decisions by nodes should produce the global effect of distributing the copies of the service such that no node is more a maximum number of hops from a replicated copy of a service. Issues such as the ability of nodes to provide a replicated copy of a service (given their location in the topology, available resources and their mobility) must also be addressed.

6 Conclusion

The integration of ad-hoc networks with fixed services is an open and interesting area of research. However, a practical solution should be designed with a realistic network scenario in mind. In particular, the interaction of routing protocols with the MAC layer should be considered. Realistic mobility and traffic patterns should be used for evaluation. The issue of where such a service resides in a network protocol stack should be considered carefully.

The use of probabilistic and swarm-intelligence approaches may be a fruitful area of research. In particular, a routing approach based on the ant colony metaphor has shown promise in telecommunications networks. Ad-hoc networks used to access services may have interesting topological properties that are not present in random network scenarios. In particular, routes will often involve a node that provides a service. An approach to routing using the 'pheromone' metaphor could be a natural way to match the effort used to maintain routes to a service with that node's popularity: the more a node communicates, the stronger a 'trail' is formed towards that node.

References

- [AGK⁺01] Payman Arabshahi, Andrew Gray, I. Kassabalidis, M.A. El-Sharkawi, R.J. Marks II, A. Das, and S. Narayanan. Adaptive Routing in Wireless Communication Networks using Swarm Intelligence. In *9th AIAA Int. Communications Satellite Systems Conf., 17-20 April 2001, Toulouse, France, 2001*.

- [AMM01] Kaizar A. Amin, John T. Mayes, and Armin R. Mikler. Agent-Based Distance Vector Routing. In *MATA*, pages 41–50, 2001.
- [Ara02] Payman Arabshahi. Swarm Intelligence Resources Page, <http://dsp.jpl.nasa.gov/members/payman/swarm/>, 2002.
- [BMJ⁺98] Josh Broch, David A. Maltz, David B. Johnson, Yih-Chun Hu, and Jorjeta Jetcheva. A Performance Comparison of Multi-Hop Wireless Ad Hoc Network Routing Protocols. In *Mobile Computing and Networking*, pages 85–97, 1998.
- [CAG02] Stuart Cheshire, Bernard Aboba, and Erik Guttman. Dynamic Configuration of IPv4 Link-Local Addresses, 2002. IETF Internet Draft.
- [CBD02] T. Camp, J. Boleng, and V. Davies. A Survey of Mobility Models for Ad Hoc Network Research. *Wireless Communications & Mobile Computing (WCMC): Special issue on Mobile Ad Hoc Networking: Research, Trends and Applications*, 2(5):483–502, 2002.
- [CD98] Gianni Di Caro and Marco Dorigo. AntNet: Distributed Stigmergetic Control for Communications Networks. *Journal of Artificial Intelligence Research*, 9:317–365, 1998.
- [CJ03] Thomas Clausen and Phillipe Jacquet. Optimized Link State Routing Protocol, 2003. IETF Internet Draft.
- [CM99] S. Corson and J. Macker. RFC 2501: Mobile Ad Hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations, January 1999. Status: INFORMATIONAL.
- [CSS02] David Cavin, Yoav Sasson, and A. Schiper. On the Accuracy of MANET Simulators, 2002.
- [CWKS97] B. Crow, I. Widjaja, J.G. Kim, and P. T. Sakai. IEEE 802.11: Wireless Local Area Networks, September 1997.
- [DACM02a] Douglas S. J. De Couto, Daniel Aguayo, Benjamin A. Chambers, and Robert Morris. Effects of Loss Rate on Ad Hoc Wireless Routing. Technical Report MIT-LCS-TR-836, MIT Laboratory for Computer Science, March 2002.
- [DACM02b] Douglas S. J. De Couto, Daniel Aguayo, Benjamin A. Chambers, and Robert Morris. Performance of Multihop Wireless Networks: Shortest Path is Not Enough. In *Proceedings of the First Workshop on Hot Topics in Networks (HotNets-I)*, Princeton, New Jersey, October 2002. ACM SIGCOMM.
- [Eco02] The Economist. Watch this Airspace, June 2002.
- [EP02] Mustafa Ergen and Anuj Puri. MEWLANA : Mobile-IP Enriched Wireless Local Area Network Architecture, 2002.
- [Fal00] K. Fall. ns Notes and Documentation. *The VINT Project*, 2000.
- [GK99] P. Gupta and P. Kumar. Capacity of wireless networks, 1999.
- [IEE99] IEEE. IEEE Std. 802.11: Wireless LAN Media Access Control (MAC) and Physical Layer (PHY) Specifications, 1999.
- [JA99] U. Jonsson and F. Alriksson. MIPMANET Mobile-IP for Mobile Ad-Hoc Networks, 1999.
- [JAL⁺00] U. Jonsson, F. Alriksson, T. Larsson, P. Johansson, and G. Maguire. MIPMANET - Mobile IP for Mobile Ad Hoc networks, 2000.

- [JLH⁺99] Per Johansson, Tony Larsson, Nicklas Hedman, Bartosz Mielczarek, and Mikael Degermark. Scenario-based Performance Analysis of Routing Protocols for Mobile Ad hoc Networks. In *MobiCom'99. Seattle WA*, August 1999.
- [JM96] David B Johnson and David A Maltz. Dynamic source routing in ad hoc wireless networks. In Imielinski and Korth, editors, *Mobile Computing*, volume 353. Kluwer Academic Publishers, 1996.
- [JMB99] D. Johnson, D. Maltz, and J. Broch. Supporting Hierarchy and Heterogeneous Interfaces in Multi-Hop Wireless Ad Hoc Networks, 1999.
- [JMB01] D. Johnson, D. Maltz, and J. Broch. DSR: The dynamic source routing protocol for multihop wireless ad hoc networks, 2001.
- [JTT99] M. Jiang, Jinyang Ti, and Y.C. Tay. Cluster Based Routing Protocol, 1999.
- [JW01] M. Zitterbart J. Wu. Service awareness and its challenges in mobile ad-hoc networks. In *Workshop der Informatik 2001: Mobile Communication over Wireless LAN: Research and Applications*, September 2001.
- [KESI⁺01] I. Kassabalidis, M.A. El-Sharkawi, R.J. Marks II, P. Arabshahi, and A.A. Gray. Swarm Intelligence for Routing in Communication Networks, November 2001.
- [Kle00] Jon Kleinberg. The Small-World Phenomenon: An Algorithmic Perspective. In *Proceedings of the 32nd ACM Symposium on Theory of Computing*, 2000.
- [KZ02] Vikas Kawadia and Y Zhang. Ad-hoc Support Library (ASL) website: aslib.sf.net, 2002.
- [LBD⁺01] Jinyang Li, Charles Blake, Douglas S. J. De Couto, Hu Imm Lee, and Robert Morris. Capacity of Ad Hoc Wireless Networks. In *Proceedings of the 7th ACM International Conference on Mobile Computing and Networking*, pages 61–69, Rome, Italy, July 2001.
- [LHH02] L. Li, J. Halpern, and Z. Haas. Gossip-based Ad Hoc Routing, 2002.
- [LLN⁺02] H. Lundgren, D. Lundberg, E. Nordström, C. Tschudin, and J. Nielsen. A large-scale testbed for reproducible ad hoc protocol evaluations. In *IEEE Wireless Communications and Networking Conference (WCNC 2002), Orlando, Florida*, 2002.
- [Loc03] LocustWorld. MeshAP, www.locustworld.com, 2003.
- [LP97] Hui Lei and Charles Perkins. Ad Hoc Networking with Mobile IP. In *Proceedings of 2nd European Personal Mobile Communication Conference*, 1997.
- [MMPS00] Frank McSherry, Gerome Miklau, Don Patterson, and Steve Swanson. The Performance of Ad Hoc Networking Protocols in Highly Mobile Environments, 2000.
- [Mon98] CMU Monarch. The CMU Monarch Project's Wireless and Mobility Extensions to NS, 1998.
- [PB94] C. Perkins and P. Bhagwat. Routing over Multi-hop Wireless Network of Mobile Computers, 1994.
- [PBR02] C. E. Perkins, E. M. Belding-Royer, and Y. Sun. Internet connectivity for ad hoc mobile networks, 2002.
- [Per96] C. Perkins. RFC 2002: IP mobility support, October 1996. Updated by RFC2290 [SG98]. Status: PROPOSED STANDARD.

- [Per97] C. Perkins. Ad Hoc On Demand Distance Vector (AODV) Routing, 1997.
- [Plu82] David Plummer. An Ethernet Address Resolution Protocol, 1982.
- [PM94] C. E. Perkins and A. Myles. Mobile IP. *Proceedings of International Telecommunications Symposium*, pages 415–419, 1994.
- [RK03] Prashant Ratanchandani and Robin Kravets. A Hybrid Approach to Internet Connectivity for Mobile Ad Hoc Networks. In *Proceedings of IEEE WCNC*, 2003.
- [RT99] E. M. Royer and Chai-Keong Toh. A Review of Current Routing Protocols for Ad Hoc Mobile Wireless Networks, 1999.
- [SCS02] Yoav Sasson, David Cavin, and A. Schiper. Probabilistic Broadcast for Flooding in Wireless Mobile Ad Hoc Networks, 2002.
- [SG98] J. Solomon and S. Glass. RFC 2290: Mobile-IPv4 configuration option for PPP IPCP, February 1998. Updates RFC2002 [Per96]. Status: PROPOSED STANDARD.
- [SHBR96] Ruud Schoonderwoerd, Owen E. Holland, Janet L. Bruten, and Leon J. M. Rothkrantz. Ant-Based Load Balancing in Telecommunications Networks. *Adaptive Behavior*, (2):169–207, 1996.
- [Sta01] William Stallings. IEEE 802.11: Moving Closer to Practical Wireless LANs, 2001.
- [TG02] Christian Tschudin and Richard Gold. LUNAR - Lightweight Underlay Network Ad-Hoc Routing, 2002.
- [TNCS02] Yu-Chee Tseng, Sze-Yao Ni, Yuh-Shyan Chen, and Jang-Ping Sheu. The Broadcast Storm Problem in a Mobile Ad Hoc Network. *Wireless Networks*, 8(2/3):153–167, 2002.
- [VGPK97] J. Veizades, E. Guttman, C. Perkins, and S. Kaplan. RFC 2165: Service location protocol, June 1997. Status: PROPOSED STANDARD.
- [Wil03] A. Williams. Requirements for Automatic Configuration of IP Hosts, 2003. IETF Internet Draft.
- [XS01] Shugong Xu and Tarek Saadawi. Does the IEEE 802.11 MAC Protocol Work Well in Multihop Wireless Ad Hoc Networks? *IEEE Communications Magazine*, pages 130–137, June 2001.
- [ZBG98] Xiang Zeng, Rajive Bagrodia, and Mario Gerla. GloMoSim: A Library for Parallel Simulation of Large-Scale Wireless Networks. In *Workshop on Parallel and Distributed Simulation*, pages 154–161, 1998.