ON THE SCALABILITY OF AD HOC WIRELESS NETWORKS

by

SULI ZHAO

A Dissertation submitted to the

Graduate School—New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Electrical and Computer Engineering

Written under the direction of

Professor Dipankar Raychaudhuri

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

January, 2008

ABSTRACT OF THE DISSERTATION

On the Scalability of Ad Hoc Wireless Networks

By SULI ZHAO

Dissertation Director:
Professor Dipankar Raychaudhuri

This dissertation considers the problem of scaling ad hoc wireless networks now being applied to urban mesh and sensor networks scenarios. Ad hoc networks involve multi-hop communication which has inherent scaling problems in that throughput per node drops as the square root of the number of nodes in the network. We investigate mechanisms for improving performance and scalability of multi-hop wireless networks, with focus on system architecture and routing protocol aspects.

First we propose a generalized multi-tier hierarchical hybrid network with three tiers of radio nodes: low-power end-user *mobile nodes* (MN) at the lowest tier, higher power radio *forwarding nodes* (FN) that support multi-hop routing at intermediate level, and wired access points (AP) at the highest level. We present an analytical model for the capacity of the proposed network and identify conditions on transmission range and node density for scalability to be maintained. From the derived upper and lower bounds, it is shown that the low-tier capacity increase linearly with the number of FN's, and the high-tier capacity grows linearly with the number of AP's in the scaling region.

The analytically obtained capacity results are validated with detailed system simulations for dense network scenarios. The simulation study also examines the allocation of separate channels to

avoid the increased protocol overhead which arises in the single channel case. A heuristic distributed channel assignment algorithm is proposed to achieve conflict-free transmissions in the network.

Next, we investigate cross-layer adaptive routing as another type of scaling mechanism. An adaptive routing framework, which allows introduction of adjustable parameters and programmable routing modules, is described. The proposed framework can support various cross-layer mechanisms including those based on integrated routing metrics that incorporate PHY and MAC information.

We investigate a PHY/MAC aware routing metric (PARMA) which incorporates physical layer link speed and MAC congestion. Design and implementation of PARMA are outlined, and simulation results for typical multi-rate 802.11 ad hoc network scenarios show that PARMA helps improve throughput and decrease congestion by selecting paths with high bit-rate links while avoiding MAC congestion areas.

Acknowledgements

I would first like to express my sincere gratitude to my thesis advisor, Professor Dipankar Raychaudhuri, for his invaluable guidance, encouragement, and support during my years as a graduate student at Rutgers University. Thank you so much for giving me the opportunity to learn from you and giving me time when I needed an intelligent discussion.

I would like to sincerely thank the other members of my thesis committee, Professor Roy Yates, Professor Wade Trappe, and Dr. Arup Acharya, for taking the time to participate in my committee and for providing constructive guidance and insight into my research.

I sincerely thank Professor Narayan Mandayam, Professor Christopher Rose, Professor Predrag Spasojevic, and Dr. Hang Liu, for all the advice they have given me in my course work and my research. I thank Mr. Ivan Seskar for sharing his expertise in wireless networking field.

Thanks are also due to fellow graduate students and the staff members at WINLAB and ECE Department for their friendship and help. They have made lab such a supportive, collaborative, and enjoyable place to work. I particularly thank Mr. Zhibin Wu collaborated with me.

I would like to express my sincere gratitude to my family for their loving support and encouragement. In particular I thank my father for all his efforts as well as instilling in my childhood the importance of education in being able to reach my dreams.

Lastly, I thank New Jersey Commission on Science and Technology (NJCST) and National Science Foundation (NSF) for their financial support. This research was supported by NJCST under Grant 03-2042-007-12 and NSF NRT under Grant ANI-0335244. Some materials in this dissertation have been presented in my publications listed in the vita.

Dedication

In memory of my father

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

Ad hoc wireless networks in which radio nodes communicate via multi-hop routing without wired infrastructure have long been considered for military communications. Early work on ad hoc networks can be traced back to the early eighties, such as the HF Intra Task Force (ITF) network [1] and the DARPA Packet Radio (PRNET) [2] and SURAN [3] programs. In the 1990's, emergence of commodity wireless local area network technology (e.g. the IEEE 802.11 [4]) led to new uses in ad hoc and multi-hop applications. At the same time, there has been a lot of work in computer science community on the problem of supporting mobility in the Internet [5], including scenarios such as wireless hotspots, hybrid cellular, and emergency ad hoc technology.

Ad hoc radio techniques are currently migrating to commercial scenarios such as sensor networks, home computing, and public wireless LAN's. Wireless mesh networks [6] are being deployed in cities to provide ubiquitous wireless coverage for the general population or in many instances, as a network for common use by the different first responder agencies such as the police, fire-fighters, or emergency medical services [7, 8].

While ad hoc wireless networks offer important rapid deployment and cost benefits, it is important that the network scales well when the node density and traffic load grow. However, the work of Gupta and Kumar [9] shows that ad hoc networks have a scalability bottleneck, i.e., the capacity of an ad hoc network tends to zero as the density of the nodes in the network approaches to infinity. This poses an important opportunity for research on ad hoc networking. With this motivation, our

work is focused on the mechanisms which can be used to improve the scaling properties of ad hoc networks. Specially, this work concentrates on the system architecture and protocol design issues involving the network and link layers (also including the physical layer), and the interactions between the network and the lower layers. We will go through the fundamental properties of ad hoc networks, review research progress in the area, and outline our approach to solve the scalability problem in ad hoc networks.

1.1 Network Architecture

The default structure formed by the physical connectivity of nodes and links of the ad hoc network is "flat", where all nodes in the network have identical functionalities. Control functions such as routing, when performed with respect to this flat network, have potential problems such as poor scalability. These problems can be overcome by self-organizing ad hoc nodes into a certain structure based on their physical connectivity. It has been shown that self-organizing control structures help increase network availability, reduce delay in responding to changes in network state, and reduce configuration errors [10]. Furthermore, changing topology and shared wireless links, which are two important properties of ad hoc networks, make cluster-based hierarchical structures a good approach for ad hoc networking. With cluster-based structures, the physical network is transformed into a virtual network of interconnected node clusters, based on which the performance of ad hoc networks is improved through reduced sensitivity to small network state changes and localized control in response to significant changes [10]. An early work proposed a distributed link cluster algorithm for discovering the network connectivity for ad hoc networks [11]. Clustering also imposes hierarchy on large ad hoc networks. Not limited to clustering, the hierarchical structure helps reduce the routing overhead without sacrificing the routing quality, as demonstrated in [10].

The hierarchical architecture is also motivated by the scalability bottleneck which arises in flat

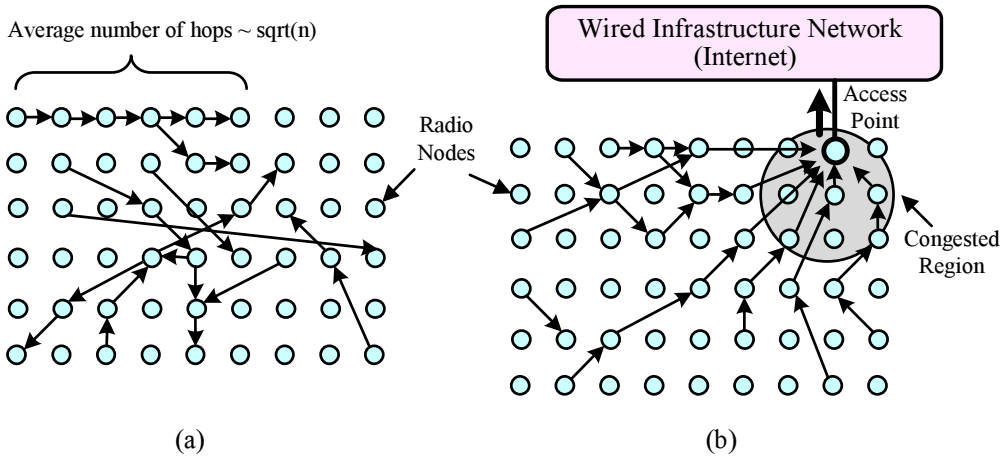


Figure 1.1: Scaling issues in flat ad hoc wireless networks with different traffic models. (a) Uniform peer-to-peer traffic model. (b) Major of traffic flows to/from the Internet.

ad hoc networks. Conventional mobile ad hoc network (MANET) architectures which use shared medium access control protocols along with multi-hop routing such as DSR (Dynamic Source Routing) [12] or AODV (Ad hoc On-demand Distance Vector) [13] suffer from decreasing throughput per node as node density increases. In particular, Gupta and Kumar's well-known theoretical result [9] for multi-hop wireless networks (illustrated in Figure 1.1(a)) indicates that achievable end-to-end per-node throughput decreases in proportion to the square root of the number of radio devices.

When considering scalability, it is also important to note that most applications involve traffic flows to and from the Internet in addition to peer-to-peer communication between radio nodes. In this case, the scaling problem in flat networks becomes even more difficult due to traffic bottlenecks around gateway nodes when a significant fraction of packets have to be routed to a correspondent host within the wired Internet, as shown in Figure 1.1(b). This kind of scenario requires effective integration of wired access points (AP's) with the ad hoc wireless network.

The above considerations motivate a network with more than one tier of ad hoc radio nodes, in which lower tiers aggregate traffic up to intermediate radio relays, while continuing to use robust ad hoc self-organization and routing protocols as in flat ad hoc networks. Several early papers have appeared on MAC and routing protocols for hierarchical radio networks [10, 14–16]. The

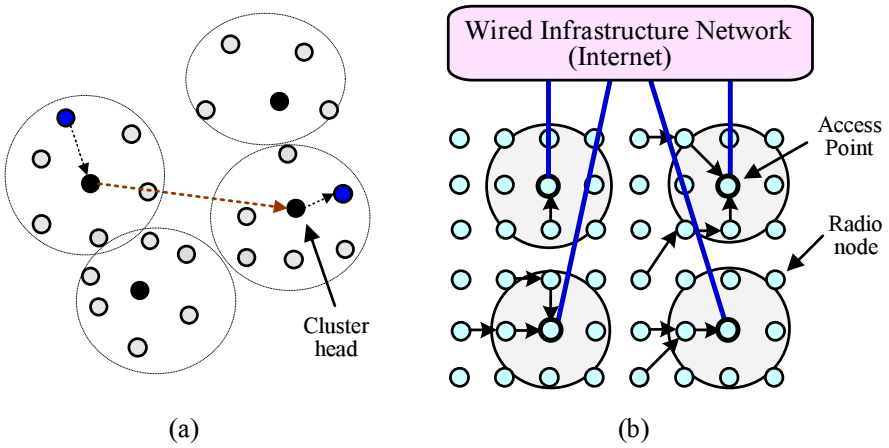


Figure 1.2: (a) MAC-layer "clustering". (b) Concept of two-tier hierarchical hybrid network.

results reported in these papers tend to indicate that scalability can be improved by introducing hierarchical structures in radio ad hoc networks. For instance, in [14], Gerla shows that a MAC-layer "clustering" (shown in Figure 1.2(a)) can improve system capacity and performance by providing a framework for code separation, channel access, and bandwidth allocation, making it possible to introduce efficient channel access procedures and localize control in response to network state changes. However, the performance gain achieved by the clustering might be limited.

Another approach is to introduce some proportion of wired or wireless "infrastructure" nodes that serve to organize the network into a hierarchy with "short cut" paths for traffic that would have required larger numbers of hops in a flat ad hoc network. Figure 1.2(b) depicts a "hybrid network" with two hierarchical tiers (i.e. radio ad hoc nodes and wired access points). In a recent paper, [17], Liu and Towsley proved that linear scaling of throughput can be approached in a two-tier hybrid network as long as the number of access points grows asymptotically faster than the square root of the number of radio nodes. In addition, results of [18, 19] have also shown that adding infrastructure nodes to ad hoc networks can effectively reduce the average number of end-to-end hops and ultimately help achieve better performance than flat ad hoc networks. All these results demonstrate that ad hoc mesh networks benefit from a hierarchical "hybrid" wired/wireless

architecture both in terms of scalability and effective integration with the Internet. However, we note that with this two-tier architecture, wired infrastructure costs can be high, especially for dense usage scenarios.

Therefore, we consider a multi-tier hierarchical hybrid wired/wireless architecture for ad hoc networks, which is designed to provide significant improvements in capacity scaling and performance relative to conventional "flat" ad hoc and two-tier "hybrid" networking approaches. The self-organizing hierarchical network architecture is introduced next.

1.1.1 Multi-Tier Hierarchical Hybrid Wireless Network Architecture

Based on the above considerations, we further generalize the two-tier hierarchy of the "hybrid network" in [17] to a K-level hierarchy with $(K-1)$ tiers of radio nodes and a top tier of AP's, as shown in Figure 1.3. Multiple radio tiers can provide further performance improvements by facilitating shorter routes between distant nodes, improving MAC efficiency via traffic aggregation and less stringent transmit power constraints, while potentially reducing the required number of wired AP's relative to the two-tier hybrid network case. A key technology enabler for the generalized hierarchical wireless network is the so-called "radio forwarding node" or "radio router" which has two or more radio interfaces to permit it to handle packets going to or from one layer of the hierarchy to another. It is also observed that each radio tier can avoid interfering with the layers above and below by orthogonal assignment of frequencies if so desired.

Note that the proposed architecture does not have a strict hierarchy in the sense that all radio nodes are allowed to access those which are more than one-tier above them; for example, a low-tier radio node can directly connect to a nearby access point when available. Also, (with the exception of the lowest-tier) radio nodes at a given level of hierarchy can communicate via paths on the same tier, or can go up and then down by one or more tiers, whichever is considered to be preferable by the

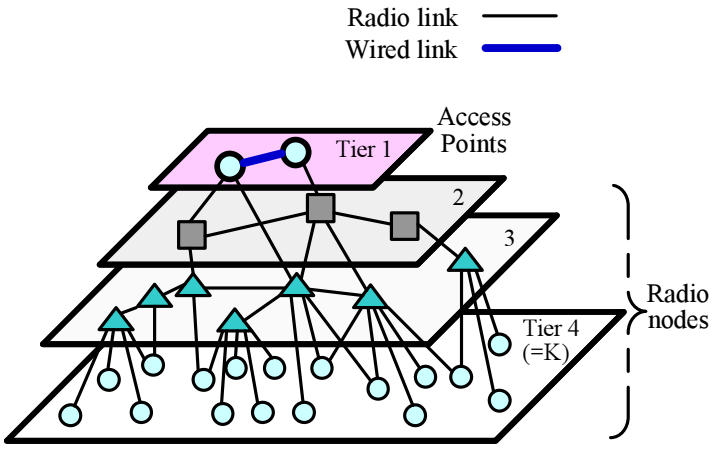


Figure 1.3: Concept of multi-tier hierarchical hybrid network.

routing algorithm. The restriction on the lowest-tier nodes is intended to avoid the need for multi-hop routing functionality at end-user radio nodes, which may be expected to have both processing and power constraints. Observe that lowest-tier nodes are allowed to associate with more than one higher tier node at the same time (i.e. "dual homing" is permitted). Radio nodes at intermediate tiers may in general have more than one radio in order to handle different bit-rate, coverage and access control needs at each level of the hierarchy. For example, a hierarchical system with three tiers may have IEEE 802.11b [4] radios at the lowest tier, dual 802.11b and 3G [20] cellular radios at the intermediate tier, and 3G base stations at the top layer to create an integrated local-area/wide-area hybrid network solution. This "network of wireless networks" architecture not only has the potential to provide scalability and improve performance, but is also a very natural framework for integrating emerging ad hoc wireless devices with existing network infrastructure. The goal is to design the system in such a way that ad hoc network advantages of dynamic self-organization and low routing overhead/complexity are retained at the lower tiers of the system, while providing the capacity and scaling advantages of a hierarchical hybrid network structure.

The key system and protocol design problems arise in conjunction with the proposed architecture include medium access control (MAC), discovery procedures, and multi-hop ad hoc routing,

and have been investigated in our group. In addition to the standard IEEE 802.11 which may be used in ad hoc mode to build a hierarchical network, MAC enhancements such as IRMA proposed in [21] can be used to improve performance via scheduling approaches that reduce MAC contention. Discovery procedures such as BEAD proposed in [22] can be used by radio nodes to identify and self-organize themselves into a network topology which considers not only connectivity but also throughput, delay and energy requirements or constraints. For ad hoc routing, algorithms which take into account the hierarchical nature, aggregated traffic and QoS requirements, as well as enhanced routing metrics such as cross-layer aware "PARMA" proposed in [23] may also be appropriate in a hierarchical network.

In this work, we focus on the system concept, capacity scaling, and network performance of the proposed hierarchical wireless network. In particular, we give the system model and outline its protocol architecture in the next section. We will analyze the capacity and scaling properties of such a hierarchical network in Chapter 2, and verify the system performance with detailed system simulations in Chapter 3.

1.1.2 Three-Tier System Model

As shown in Figure 1.4, a typical realization of the proposed system has three tiers of radio nodes: low-power end-user *mobile nodes* (MN) at the lowest tier, higher powered radio *forwarding nodes* (FN) that support multi-hop routing at the second level, and wired access points (AP) at the third and highest level. The AP's and FN's use a self-organizing discovery protocol to form a multi-hop routed wireless infrastructure network. MN's in this system simply connect to the nearest available (i.e. strongest signal) AP or FN in order to conserve power, and thus are not required to carry any intermediate multi-hop routed traffic. This architecture is applicable to a number of emerging ad hoc networking scenarios including hybrid cellular/ad hoc networks, urban mesh networks, home wireless networks, and large-scale sensor nets. In each of these scenarios, the introduction of a

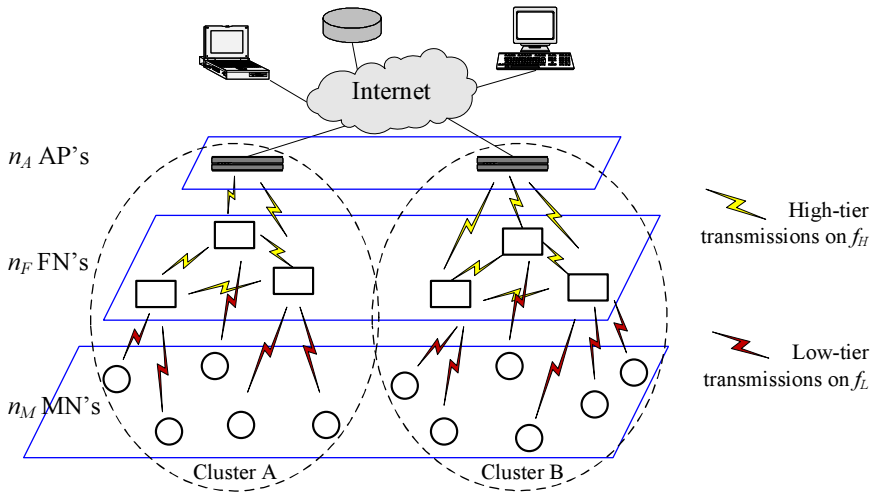


Figure 1.4: Proposed hierarchical hybrid wireless network with three tiers.

forwarding node (FN) as an intermediate radio router helps to scale network throughput and reduce infrastructure deployment cost.

It is observed that the proposed hierarchical architecture is radio independent, in the sense that different radio access technologies can be used for each tier depending on the specific scenario being addressed. For example, in a wide-area "hot-spot" data service application, the low tier nodes can use IEEE 802.11b WLAN radios, while the FN's and AP's (or base stations) can use a broadband wide-area radio standard such as WCDMA/3G [20] or WiMax/IEEE 802.16 [24]. Alternatively, a sensor net scenario could use a low-power access protocol such as Zigbee [25] for access from MN's and 802.11b for links between FN's and AP's. Of course, an all-802.11 solution is also possible for application scenarios such as mesh WLAN's for community networking. Note that when the lower and upper tiers of the hierarchical network use different radio technologies, the forwarding node must be equipped with multiple radio interfaces and be able to forward packets between them.

Each of the network entities in the proposed system are defined in further detail below:

- *Mobile Node* (MN), is a mobile end-user device (such as a sensor or a personal digital assistant) at the lowest tier of the network. The MN attaches itself to one or more nodes at the higher tiers of the network in order to obtain service using a discovery protocol. For instance,

the MN may use a single 802.11a, b or g radio operating in ad hoc mode to communicate with the point(s) of attachment. As an end-user node, the MN is not required to route multi-hop traffic from other nodes, although it does participate in the routing protocol used by the network as a whole. It is noted that as a battery-operated end-user device, the MN will typically have energy constraints.

- *Forwarding Node* (FN), is a fixed or mobile intermediate radio relay node capable of routing multi-hop traffic to and from all three tiers of the network's hierarchy. As an intermediate node without traffic of its own, the FN is only responsible for multi-hop routing of transit packets. An FN with one 802.11 radio interface uses the same radio to connect in ad hoc mode to MN's, other FN's and the higher-tier AP's defined below. Optionally, an FN may have two radio cards, one for FN-MN transmissions and the other for intra FN and FN-AP traffic flows (typically carried on different frequencies, as displayed in Figure 1.4). Specific nodes that an FN will connect to at each of the three tiers are identified using a discovery protocol that includes a distributed topology optimization algorithm. The FN is typically a compact radio device that can be plugged into an electrical outlet, but in certain scenarios, may also be a battery-powered mobile device. Thus, the FN is also energy constrained, but the cost of power is typically assumed to be an order of magnitude lower than that of the MN defined above.

- *Access Point* (AP), is a fixed radio access node at the highest tier of the network, with both a radio interface (e.g. 802.11) and a wired interface to the Internet. The AP is capable of connecting to any lower tier FN or AP within range but unlike 802.11 typical WLAN deployments, it operates in ad hoc mode for each such radio link. The AP also participates in discovery and routing protocols used in the lower tier FN's and MN's, and is responsible for routing traffic within the ad hoc network as well as to and from the Internet. Logically,

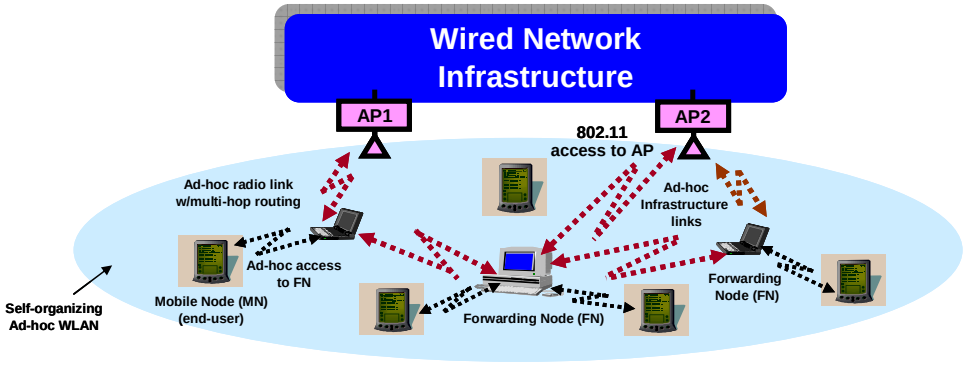


Figure 1.5: Three-tier 802.11-based network.

the AP-tier is no different from the FN-tier when routing internal ad hoc network traffic—the wired links between AP's are reflected in (generally) lower path metrics. Since the AP is a wired node, it is usually associated with an electrical outlet and energy cost is thus considered negligible.

Figure 1.5 displays a three-tier 802.11-based network which can be applicable to mesh networking. In such a network, some nodes without power constraints (network elements or terminals) such as laptops and hosts, either fixed or mobile, serve as forwarding nodes, to extend the 802.11 WLAN coverage. End-user mobile nodes attach themselves to forwarding nodes to obtain service using a discovery protocol. The links between higher-tier nodes (AP's or FN's) form ad hoc infrastructure, and are chosen by the multi-hop routing protocol to transfer packets for end-user nodes. The system evaluation of the capacity scaling and network performance of the proposed hierarchical network, presented in Chapter 3, is based on this three-tier 802.11-based network. We will show that the hierarchical structure helps to achieve scalability, effective integration with the WLAN and Internet, improved coverage, and saved power consumption at end-user radios.

Figure 1.6 depicts the protocol architecture along with control and data interfaces of an 802.11-based network consisting of MN's, FN's, and AP's defined above. The MAC layer in each of the radio devices provides the capability of discovering other ad hoc nodes as well as of resolving

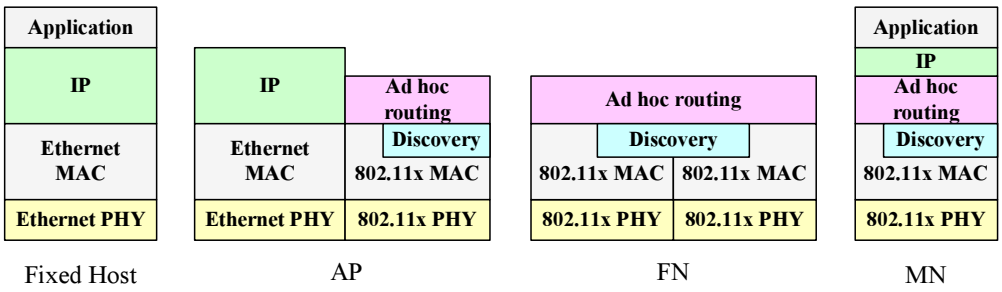


Figure 1.6: Protocol architecture of an 802.11-based hierarchical hybrid network.

contention between packets to be transmitted on the channel. Clearly, the PHY and MAC layers are radio-specific: for an all-802.11 solution, the current ad hoc mode beaconing procedure in the MAC layer needs to be modified to identify the type of node (MN, FN or AP) and its transmit power level so that other nodes can execute a suitable distributed discovery protocol to determine network connectivity. As shown in the protocol diagram, a radio-independent ad hoc routing protocol runs between the MN, FN and AP nodes, with the AP providing a gateway between the wired IP network and the ad hoc wireless subnet. Note that MN's in this system implement a simplified client version of the routing protocol in view of the fact that they typically connect to only a single FN or AP, while FN's and AP's are required to maintain complete multi-hop routing tables.

Figure 1.7 shows the main concepts behind the self-organizing network protocols described here. In particular, the system is based on augmented beacons which are periodically broadcasted by every FN and AP. FN's seek other FN's and AP's and then carry out a distributed topology establishment algorithm that achieves a suitable balance between connectivity, throughput, delay, and energy consumption. The newly connected FN then carries out routing updates to those nodes to which it has chosen to attach, thus becoming a part of the ad hoc network's routing tables. MN's which enter the system seek AP's and FN's by listening for beacons and then associate with one or more nearby upper-tier nodes based on local optimization criteria. As a related work associated with the proposed hierarchical architecture, the proof-of-concept prototype of an 802.11-based self-organizing hierarchical ad hoc wireless network (SOHAN) is presented in [26], and the details of

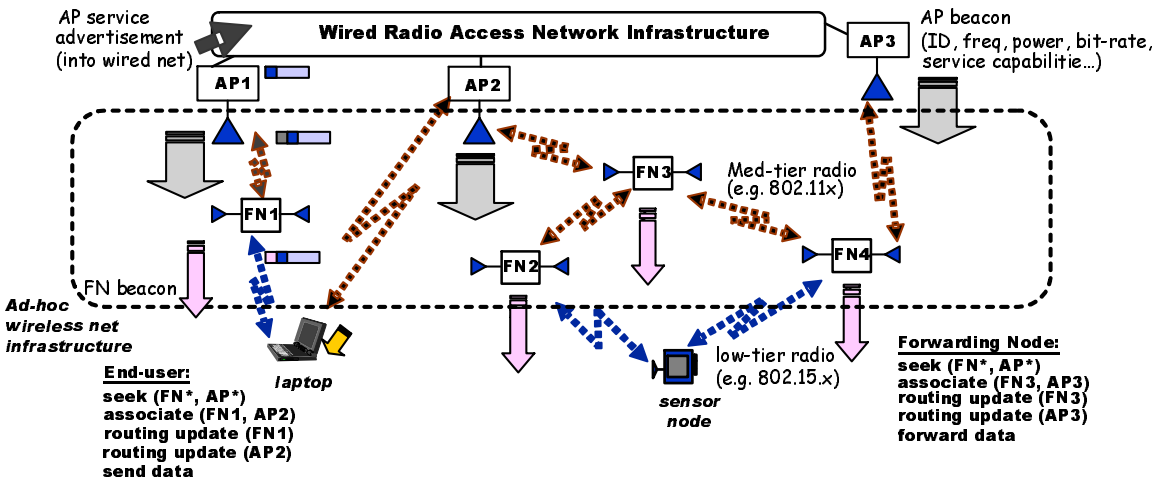


Figure 1.7: Schematic showing operation of self-organizing discovery protocol in the hierarchical ad hoc network

the enhanced MAC protocol and the related discovery protocol are described.

1.2 Protocol Design

As shown in Figure 1.6, the main protocol stack consists of MAC, discovery, and routing. Therefore, the key protocol design and evaluation problems of ad hoc networks are in these three functions.

1.2.1 Medium Access Control (MAC) and Discovery

A basic characteristic of the radio channel is its broadcast nature, which requires sharing the medium among neighboring nodes resulting in interference between transmissions from different nodes. Therefore, a multiple access protocol is needed to handle the resolution of transmission conflicts, such as Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA) [4, 27] or Time Division Multiple Access (TDMA) [28]. One of the important problems associated with ad hoc networks of current technology is the inefficient MAC in multi-hop transmissions. For instance, the CSMA/CA mechanism used in the IEEE 802.11 [4] implies “exposed node” which cannot transmit in parallel. This problem becomes more severe when the network gets larger.

Therefore, efficient MAC and MAC enhancements are indispensable for efficient ad hoc network operation. Although the standard IEEE 802.11 alone may be used in ad hoc mode, it is generally possible to improve performance via scheduling approaches that reduce MAC contention especially at the higher layers of the hierarchical network which deal with aggregated traffic flows. For instance, an integrated routing and MAC scheduling algorithm (IRMA) proposed in [21] eliminates contention between radio nodes.

Another very important characteristic of a mobile ad hoc wireless network is its changing topology [11, 29]. In particular, the network connectivity is affected by the nodes when they join, leave, and move in the network. The topological changes may also be caused by fluctuating wireless link quality and physical bit-rate adaptation [30, 31]. Under these circumstances, whether a node is a "neighbor" or not depends on a set of physical layer parameters, instead of fixed connections as in wired networks. The topological changes may be unpredictable and/or frequent, which makes the architecture and protocol design of ad hoc wireless networks an important challenge. Therefore, the multiple access protocol must include management functions that discover and maintain the network topology in spite of the network changes on connectivity and offered load, thus help maximize the medium usage and achieve good throughput and system performance.

Through the discovery procedure, radio nodes become aware of each other's presence and then automatically create a wireless network via selection of appropriate radio frequencies and nodes to associate with. The discovery process requires both protocol support (such as beacons for radio nodes to identify themselves) and distributed algorithms for self-organization of the network topology taking into account connectivity, throughput, delay, and energy requirements or constraints. Given as an example, the BEAD proposed in [22] is a beacon-assisted discovery mechanism for the proposed self-organizing hierarchical ad hoc networks. Furthermore, by making a subset of wireless links available to routing, the discovery protocol creates an efficient topology and reduces burden

on routing and routing overhead.

1.2.2 Ad Hoc Routing

The routing algorithm is used to generate a decision-making procedure for each node to select one or more of its neighbors to forward a packet on its way to the correct destination. Most ad hoc routing protocols are based on ideas from routing methods in conventional wired computer networks. Two of the most popular routing algorithms in computer networks are *distance vector* and *link state* routing [32]. Distance vector routing algorithms maintain a table in each router which gives the best known distance to each destination and the link to use to get there. Pure distance vector algorithms such as distributed Bellman-Ford [27] do not perform well in mobile networks because of slow convergence and count-to-infinity problem. Thus these algorithms need to be modified and enhanced when used in ad hoc network scenarios, and the resulting example protocols are Destination Sequence Distance Vector (DSDV) [33] and Ad hoc On-demand Distance Vector (AODV) [13]. In link state routing algorithms, each router discovers its neighbors and measures the link cost to each of them, and then distributes the link state information to all other routers and finally computes the shortest path to every other router. Optimized Link State Routing (OLSR) [34] falls into this category. Compared to link state routing, distance vector routing has the advantage of more efficient computation, and is easier to implement and requires less storage space. But link state routing records the entire path and enables nodes to gather more link state information which facilitates route selection corresponding to different criteria.

Ad hoc routing protocols may also be categorized as *proactive* (or *table-driven* e.g. DSDV) and *reactive* (or *on-demand* e.g. AODV and Dynamic Source Routing (DSR) [12]) routing protocols, and combinations thereof (e.g. Zone Routing Protocol (ZRP) [35]). Proactive routing protocols continuously compute routes to all nodes so that a route is already available when a packet needs to be sent to a particular node; while on-demand routing protocols start a route computation process only

when a packet needs to be sent to some other node. On-demand routing protocols avoid periodic route advertisements and hence save bandwidth and power consumption in mobile environments, but data packets may experience larger delay than using proactive routing protocols [36].

Location information may improve routing performance. For example, as an extension of DSR, Location-Aided Routing (LAR) [37] sends location information in all packets to decrease the overhead of a future route discovery.

In addition to the above ad hoc routing protocols which are extended from wired networks, there are some other routing approaches specifically designed for dense sensor networks, of which *directed diffusion* [38] is an example. Directed diffusion incorporates attribute-based naming, data-centric routing, and application-specific procession inside the network. In particular, each node in the sensor network names data that it generates with one or more attributes, and other nodes may express interests based on these attributes. Then each node disseminates interests based on the contents of the interest, and the path of interest propagation sets up a reverse data path for data that matches the interest.

Another routing approach designed for ad hoc networks is geographic routing. Geographic routing identifies nodes by their locations and uses these coordinates to forward packets (if possible) in a greedy manner towards the destination. This type of routing only keeps local information thus it scales well. The challenge of geographic routing is how to get through dead-ends when greedy routing fails. Greedy Perimeter Stateless Routing (GPSR) [39] solves this problem by routing around the perimeter of such regions.

While there are many results on specific classes of ad hoc routing protocols, no single routing protocol performs well across the full range of parameters associated with a complex real-world environment. Meanwhile, the potential scaling problem associated with ad hoc routing, i.e., the routing overhead would increase as node density grows, is still an open problem. Therefore, we do

not restrict our investigation to particular ad hoc routing algorithms such as AODV or DSR. Also, in order to generalize our scalability mechanisms, we do not confine to any routing algorithm designed for specific applications, e.g. directed diffusion and location-aware routing. What interests us is how to ensure the routing quality as the node density increases. Therefore, the problem may be how to adapt the routing protocol to changing network conditions while meeting different external service needs. We also need to minimize the resource consumed by routing i.e. routing overhead as a key measure of performance. Based on these considerations, we study an adaptive routing framework in this dissertation.

Again, deployment of a hierarchical structure, by which an intrinsic routing aggregation benefit is achieved, is a possible solution for the ad hoc routing scaling problem. In the hierarchical ad hoc network, routing should take into account the hierarchical nature of the network, aggregated traffic, and the QoS requirements for services such as VoIP and media streaming. Although popular ad hoc routing protocols such as DSR and AODV can be used as a starting point, modifications to routing metrics and update policies may be appropriate in a hierarchical network. In addition, there is a need to better understand and exploit the dependencies between routing, MAC, and discovery protocols. We will study the routing behavior in the hierarchical ad hoc network in Chapter 3, and discuss an adaptive routing framework that supports enhanced routing metrics in Chapter 4.

1.2.3 Cross-Layer Adaptation

In ad hoc wireless networks, interdependencies between different layers are so prominent that a holistic approach, i.e., taking a unified view across the stack of the OSI architecture, to network design is unavoidable [29]. In this work we focus on the network (i.e. routing) layer and how to jointly consider it with the lower layers. Since the physical, MAC, and routing protocols of multi-hop wireless networks depend so strongly on one another, it makes good sense to optimize them jointly. On the other hand, the interactions between them may be complex and software

implementation would become more difficult than with a layered architecture. As explained in [29], it is hard to tackle such a complex cross-layer optimization problem as a whole at once. Starting with PHY/MAC aware routing is a simple and applicable way to gradually implement cross-layer design while gaining insight towards global optimization across different layers.

The physical layer wireless link is the basis of network connectivity. As we know, whether the wireless link is "on" or not depends on many physical layer parameters such as the transmission rate, the transmission power, the modulation and coding scheme, and the channel characteristics. Among them the transmission rate and the transmission power are two readily controllable parameters. That is, we can very the network connectivity and ultimately benefit the network performance via rate and/or power control.

It is noted that, when comparing power control with rate control, adjusting transmission power would change the interference on neighboring nodes, while rate control does not affect the interference while transmitting. Meanwhile, rate control can be implemented readily, such as auto-rate fallback (ARF) [30] and receiver-based auto-rate (RBAR) [31] which are two schemes proposed for the IEEE 802.11 devices [4]. Power control is related to energy efficiency [40, 41] which can be an important design consideration in hierarchical ad hoc networks. For simplicity, our work is restricted to rate control for cross-layer adaptation, and we do not explicitly address the energy efficiency issue.

In addition to the physical layer parameters, it is necessary to take into account the MAC layer parameters in ad hoc routing design. The MAC protocol aims to resolve transmission conflicts by controlling nodes to access the shared wireless channel. The busy medium in the proximity of links with high load can also be sensed by the MAC. But the problem can not be totally solved by the MAC even if the MAC is of some adaptive feature. However, the routing layer can utilize the MAC information to guild the traffic to go around the busy region and maximize the traffic being

accommodated by the medium.

A natural way to incorporate the lower layer parameters to routing is to use a PHY/MAC aware routing metric, as routing metric provides a way to reflect the characteristics of wireless environment. Most conventional ad hoc routing protocols, including DSDV, AODV, and DSR, use the *minimum hop* (MH) as the metric to make routing decisions. This is primarily a carry-over from wired networks but may not produce optimal performance in wireless environment. Therefore, we investigate a routing metric which takes into account the physical and MAC layer parameters such as the PHY data rate and the MAC activity level. We expect that with this routing metric, packets can choose the high rate links while also avoiding congested areas in the network thus improving the system throughput and other performance metrics. In particular, we propose a PHY/MAC aware routing metric, denoted PARMA, and study its performance with DSDV. PARMA can also be deployed in the proposed adaptive routing framework.

One important issue in cross-layer design is the different time scales of the network and PHY/MAC layer variations [42]. We must be careful not to degrade the routing performance when incorporating the PHY/MAC aware routing metric. We will discuss the mechanisms for such purpose in Chapter 4.

1.3 Dissertation Outline

In this chapter, we have introduced the research status of ad hoc networking, identified the challenges of scaling ad hoc networks, and presented some directions that could be taken to solve the problem. Next, following these directions of research, we are going to investigate specific mechanisms for improving the scaling properties of ad hoc networks, with focus on system architecture and routing protocol aspects including three-tier hybrid network model, dual radio forwarding

nodes, separate allocation of frequency bands to each tier, adaptive routing framework, and cross-layer routing.

Based on the conceptual system model of multi-tier ad hoc networks give in Section 1.1, a general analytical model for the asymptotic capacity and scaling properties of the proposed three-tier network is developed in Chapter 2. Based on this analytical model, we prove the asymptotic throughput capacity, refine the linear scaling regime by providing an upper bound, and identify the conditions on transmission range and node density for scalability to be maintained. We also study the capacity of a *Random Aggregate Network* and the traffic distribution in such a network.

In Chapter 3, we verify the scaling properties of the three-tier network with detailed system simulations (for dense network scenarios) and relate the experimental results obtained with the analytical asymptotic results. The simulation model is implemented with realistic MAC and routing protocols, and is also useful for providing insight into the choice of system parameters such as the ratio of mobile nodes to forwarding nodes and access points, the traffic pattern, and the channel bandwidth allocation required for high and low tiers of the hierarchy. In addition, the simulation study also examines the allocation of separate channels to avoid the increased protocol overhead which arises in the single channel case. A heuristic distributed channel assignment algorithm is proposed to achieve conflict-free transmissions in the network.

In Chapter 4, we propose a unified adaptive routing framework which allows introduction of programmable routing modules and adjustable parameters. In this framework, protocol selection, routing algorithm parameters, and/or routing metric can be controlled in response to observed performance and external service needs. We also study a cross-layer adaptive routing metric as an integrated routing algorithm that can be deployed in the adaptive framework. The proposed PHY/MAC Aware Routing Metric (PARMA) aims to minimize end-to-end delay that includes both transmission and access times. We study PARMA working with proactive distance vector ad hoc routing

protocols (such as DSDV), and specific enhancements to the routing protocol which is originally designed working with the minimum hop routing metric. In addition, smoothing techniques for the link portion of the proposed metric is introduced to adjust the different change variations between the MAC layer and the network layer and also to improve route convergence.

Finally, Chapter 5 summarizes the main results of this dissertation and outlines the future directions of the research.

Chapter 2

Capacity of Three-Tier Hierarchical Hybrid Wireless Networks

In this chapter, we develop a general analytical model for the asymptotic capacity and scaling properties of three-tier hierarchical hybrid wireless networks. First, we briefly review the relevant notation, the related work, and some useful results established in prior work.

2.1 Background and Notation

2.1.1 Interference Model

In this work, we use the *Protocol Interference Model*. The interference model describes the successful reception of a transmission over one hop [9]. Let the common transmission range be r. A transmission from node X_i is successfully received by node X_j if

- The distance between X_i and X_j is no more than r, i.e., $|X_i - X_j| \leq r$;
- For every other node X_k simultaneously transmitting over the same channel, the distance between X_k and X_j is no less than $(1+\Delta)r$, i.e., $|X_k - X_j| \geq (1+\Delta)r$,

where $\Delta > 0$ defines the size of the guard zone.

Suppose node X_i is transmitting to node X_j and $|X_i - X_j| \leq r$. If there is another node X_k that is transmitting in the same time slot and the distance satisfies $|X_k - X_j| \leq r$, then there is a *collision* at node X_j and X_j obtains no information about the transmitted packets.

The nodes in the network can be organized into certain type of groups, e.g., sub-clusters or clusters (they will be introduced later). Similar to [9], we define that two groups are *interfering neighbors* if there is a node in one group which is within a distance of $(2 + \Delta)r$ of a node in the other group. Therefore, in the Protocol Interference Model, if two groups are not interfering neighbors, a transmission from one group does not collide with a transmission from the other group, i.e., simultaneous transmissions from these two groups are allowed.

2.1.2 Time Scheduling

In our wireless network model, time is divided into slots of fixed durations. A node is scheduled to send data in each time slot on a wireless channel. A slotted packet scheduling is used to eliminate transmission collisions and interference for nodes in the interfering neighborhood.

2.1.3 Feasible Throughput

A throughput of $\lambda(n)$ bits per second for each source node is *feasible* if there exists a spatial and temporal scheme of transmissions, such that each source node can transmit $\lambda(n)$ bits per second on average to its destination node.

The *throughput capacity* of a wireless network is of order $\Theta(f(n))$ bits per second if there are deterministic constants $c_1 > 0$ and $c_2 < \infty$ such that

$$\lim_{n\to\infty} \text{Prob}\left(\lambda(n) = c_1 f(n) \text{ is feasible }\right) = 1,$$

$$\lim_{n\to\infty} \text{Prob}\left(\lambda(n) = c_2 f(n) \text{ is feasible }\right) < 1.$$

The *aggregate throughput capacity* is measured as the sum of the throughputs furnished to the nodes of a network (or cluster). Let $\lambda_i(n)$ bits per second be the throughput of node X_i. The aggregate throughput capacity of a wireless network, i.e. $\sum_{i=1}^{n} \lambda_i(n)$, is of order $\Theta(f(n))$ bits per

second if there are deterministic constants $c_1 > 0$ and $c_2 < \infty$ such that

$$\lim_{n\to\infty} \text{Prob}\left(\sum_{i=1}^{n} \lambda_i(n) = c_1 f(n) \text{ is feasible}\right) = 1,$$
$$\lim_{n\to\infty} \text{Prob}\left(\sum_{i=1}^{n} \lambda_i(n) = c_2 f(n) \text{ is feasible}\right) < 1.$$

2.2 Related Work

In [9], Gupta and Kumar obtain the capacity of multi-hop wireless networks with n identical randomly located nodes, each capable of transmitting at W bits per second, using a fixed range and under the Protocol Interference Model, which is $\Theta\left(\frac{W}{\sqrt{n \log n}}\right)$ bits per second per node for randomly chosen destinations.

Capacity scaling laws for two-tier hybrid wireless networks have been studied in [17–19,43,44]. Suppose a hybrid network consist of n ad hoc nodes and m base stations (BS's) or AP's. In [17], Liu and Towsley have proved that, for the deterministic routing strategy with regular placement of BS's for the Protocol Interference Model, the aggregate throughput capacity grows linearly with m when $m = \Omega(\sqrt{n})$ and all bandwidth is allocated to the traffic that goes through the infrastructure.

When ad hoc nodes and AP's are randomly distributed, a throughput capacity of $\Theta\left(\frac{W}{\log n}\right)$ can be achieved, provided that the number of AP's scales linearly with n [18]. In this work, the transmission range depends on n and m, and ensures an optimal routing scheme in which each ad hoc node is only one hop away from a certain infrastructure node.

The results on the capacity of hybrid networks obtained in [43] are similar to those reported in [17], but a spatial-diversity scheme is used to achieve the maximum throughput which is fairly shared among the ad hoc nodes.

Employing power control allows better scaling of capacity in wireless networks. In [45], the authors show that a $1/\sqrt{n}$ rate is achievable in flat ad hoc networks of randomly located nodes. In [44], it is shown that a $\Theta(1)$ throughput to some $\Theta(n)$ nodes can be guaranteed in a hybrid

network.

Linear scaling of capacity can be achieved when $\sqrt{\frac{n}{\log n}} \lesssim m \lesssim \frac{n}{\log n}$ in a hybrid network with arbitrarily placed infrastructure nodes [19]. This scaling upper bound of $\frac{n}{\log n}$ is similar to that obtained in our work, but in [19] linearity is achieved by letting ad hoc nodes communicate directly via infrastructure nodes thus requiring power control. However, we use a common power level with regularly placed AP's.

In order to facilitate system evaluations with practical model and realistic protocols, in our analysis we employ bandwidth partitioning and consider the Internet traffic and local traffic separately (similar to [17]). This is because bandwidth partitioning can be implemented with the multi-channel capability provided by current radio devices, and different traffic types have different scaling behaviors. As distributed power control is a complex design problem in wireless networks, we do not consider it here. At the highest tier, multi-hop wireless transmissions are allowed between AP's and FN's, which does not impose restrictions on transmission power as in [19]. Also, we do not consider the capacity gain that could be achieved from node mobility [46], or other approaches such as node cooperation and MIMO communication [47,48]. Generally, we use an analytical model that is closely related to practical systems. Note also that the approach of bandwidth partitioning can be easily applied to the K-tier hierarchical network with $K > 3$ that is introduced in Section 1.1.

It is observed that our results on the capacity of Random Aggregate Networks are similar to those on the capacity of cluster networks [43]. However, the proof presented here is simpler and by our approach we can derive the traffic distribution of such a network (see Appendix A), which provides a basis for designing scheduling algorithms to overcome the capacity bottleneck at the hotspots in close proximity to aggregation nodes.

2.2.1 Useful Results

Some useful results established in [17] are summarized here. In the two-tier hybrid network of n ad hoc nodes and m base stations (BS's) [17], the regularly placed BS's divide the network into a hexagon tessellation, in which each hexagon is called a cell. Assuming that each node can transmit at W bits per second over the wireless channel, and further the channel is divided into two sub-channels to carry intra-cell traffic and inter-cell traffic with bandwidth W_a and W_i respectively. Under the deterministic routing strategy and the Protocol Interference Model, the following lemmas hold:

Lemma 2.1. *If $m = o(\sqrt{n})$, the aggregate throughput capacity contributed by intra-cell traffic is*

$$T_a = \Theta\Big(\sqrt{\frac{n}{\log(n/m^2)}}\, W_a\Big).$$

The aggregate throughput capacity is maximized when $W_a/W \to 1$. And the corresponding capacity is:

$$T = \Theta\Big(\sqrt{\frac{n}{\log(n/m^2)}}\, W\Big).$$

Lemma 2.2. *If $m = \Omega(\sqrt{n})$, the aggregate throughput capacity contributed by intra-cell traffic is*

$$T_a = O(\sqrt{n}\, W_a).$$

The aggregate throughput capacity is maximized when $W_i/W \to 1$. And the corresponding capacity is:

$$T = \Theta(m\, W).$$

2.3 Analytical Model

We model the three-tier wireless network, which is introduced in Section 1.1.2, as follows.

2.3.1 System Modeling

As shown in Figure 1.4, we suppose there are n_A AP's, n_F FN's, and n_M MN's ($n_A < n_F < n_M$) in a disk of unit area on the plane. MN's are independently and uniformly distributed, while AP's and FN's are placed in a regular pattern. Through the ad hoc network discovery procedure, each MN is associated with the nearest FN via direct transmission (assuming there is always at least one FN within its transmission range), and each FN is associated with its nearest AP via one-hop or multi-hop transmissions. After associations, all nodes in the disk form clusters, each of which consists of one AP, its associated FN's, and their associated MN's (as in Figure 1.4). Also each FN and its associated MN's form a sub-cluster. Note that lower tier nodes are only allowed to associate with one higher tier node at a given time, i.e., "single homing" is deployed in the analytical model.

The wireless channel is divided into two sub-channels: one is carried on frequency f_L and the other f_H. f_L is used for transmissions to and from MN's (denoted *low-tier transmissions*); f_H is used for transmissions not involving MN's (denoted *high-tier transmissions*). We assume that each node, if working on frequency f_L (or f_H), is capable of transmitting at W_L (or W_H) bits per second (we simply state that these two sub-channels are with bandwidth W_L and W_H respectively). Each FN is equipped with two radios, and can thus participate in both low- and high-tier transmissions using different radios. All nodes working on a frequency are assumed to have a common transmission range, but the range could be different from the transmissions on the other frequency. Let the transmission range on f_L and f_H be r_L and r_H respectively, and the guard zone size be Δ_L and Δ_H respectively. The network model of a planar layout is shown in Figure 2.1.

AP's are interconnected through an infrastructure network of infinite capacity. We do not consider the mobility of MN's, and therefore there is no capacity improvement brought in by node mobility as discussed in [46].

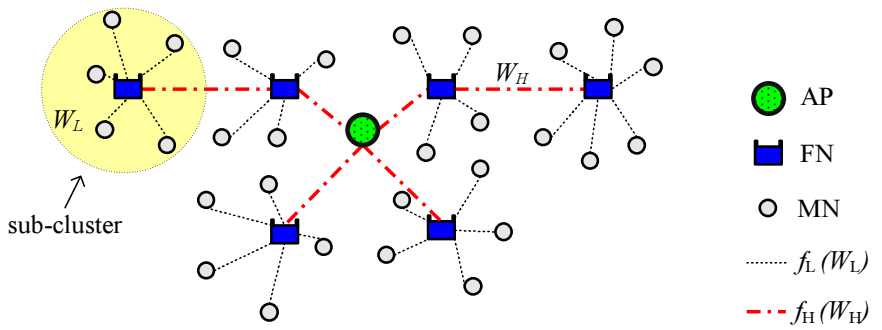


Figure 2.1: Analytical system model.

2.3.2 Traffic Pattern

In the three-tier hierarchical network under consideration, MN's are at the lowest tier and perform end-user functions such as mobile computing or sensing/actuation. MN's do not forward packets for others but send out and receive their own packets. FN's do not have any traffic needs of their own, but forward packets for other nodes and work as the relay tier between the MN-tier and AP-tier. As discussed before, we need to consider both traffic to/from the Internet and peer-to-peer traffic. Accordingly, we define the *Internet traffic* as traffic between MN's and AP's, and the *local traffic* as traffic between MN's.

2.3.3 Routing

We assume that the Internet traffic must go through the FN-tier, even if the MN is only one hop away from the AP-tier. For the local traffic, depending on the relative locations of the communicating pairs, it is either intra-cluster or inter-cluster traffic: it is intra-cluster traffic if the source and destination MN's are in the same cluster, otherwise it is inter-cluster traffic. We assume that intra-cluster traffic is relayed by the FN's of the same cluster via ad hoc mode transmissions; while inter-cluster traffic always goes through the infrastructure using a mix of multi-hop wireless links and wired infrastructure paths.

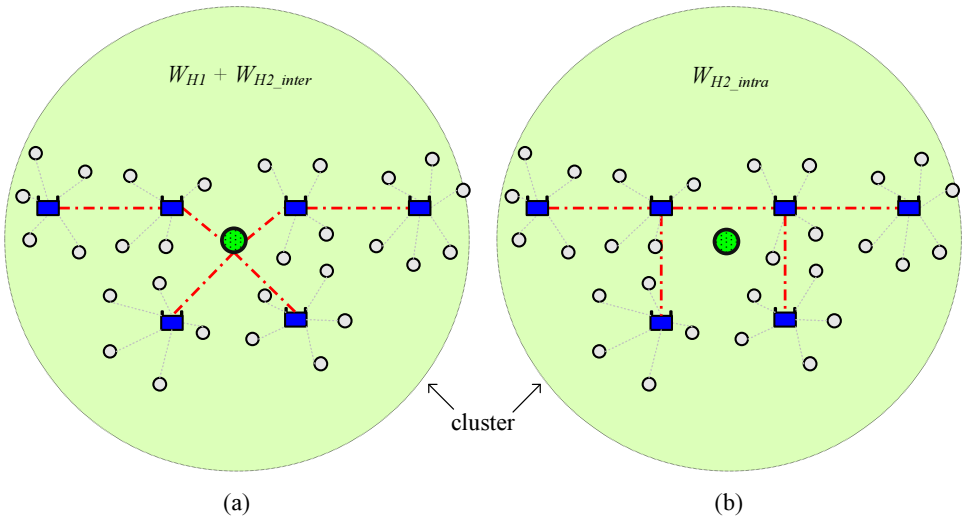


Figure 2.2: High-tier transmissions in a cluster. (a) The Internet traffic and the inter-cluster local traffic at the high-tier. (b) The intra-cluster local traffic at the high-tier.

2.3.4 Capacity Separation

Since low-tier and high-tier transmissions use different sub-channels, and there is no interference between them, we can look at capacity contributed by transmissions at these two tiers separately (denoted *low-tier capacity* and *high-tier capacity* respectively). Figure 2.1 shows a sub-cluster and the low-tier transmissions in it. High-tier transmissions are illustrated in Figure 2.2, where Figure 2.2(a) depicts a network of a cluster with the Internet traffic and inter-cluster local traffic at the high-tier, and Figure 2.2(b) displays a network of a cluster with the intra-cluster local traffic at the high-tier.

Before the capacity analysis, we model the *Random Aggregate Network* for deriving the aggregate throughput capacity of the network in Figure 2.2(a) and the required transmission range.

2.4 Random Aggregate Networks: Randomly Located Nodes and Aggregate Traffic Pattern

In a random aggregate network scenario, n source nodes are independently and uniformly distributed in a disk of unit area on the plane. We randomly place a node in the disk as the destination.

Thus all n source nodes have a common destination, to which each of them wishes to transmit packets at rate $\lambda(n)$ bits per second.[1] We further assume that all nodes employ the same transmission range or power, and are capable of transmitting at W bits per second. Also we use the Protocol Model for interference.

Theorem 2.3. *For a Random Aggregate Network of n source nodes and one destination node on a planar disk in the Protocol Interference Model, the order of aggregate throughput capacity is*

$$T = \Theta(W) \quad \text{bits per second.} \tag{2.1}$$

And the transmission range r is chosen as

$$r = \frac{80}{\sqrt{\pi}}\sqrt{\frac{\log n}{n}}.$$

Proof: Similar to [9], we use a Voronoi tessellation [49] of the planar disk to prove this theorem. The Voronoi tessellation of a set of nodes in the disk is plotted in Figure 2.3. Assuming the edge effects are ignored. It can be shown that the special properties of the Voronoi tessellations of the surface of the sphere, described as Lemma 4.1 in [9], hold for the planar disk here. The lemma can be rewritten as follows:

Lemma 2.4. *For every $\epsilon > 0$, there is a Voronoi tessellation of the disk on the plane with the property that every Voronoi cell contains a disk of radius ϵ and is contained in a disk of radius 2ϵ.*

Let $R_v = \sqrt{\frac{100\log n}{\pi n}}$. Lemma 2.4 implies that there exists a Voronoi tessellation $\mathcal{V}_n$ such that each Voronoi cell $V \in \mathcal{V}_n$ contains a disk of radius R_v and is contained in a disk of radius $2R_v$. Define *adjacent cells* as the Voronoi cells that share a common point. Set the transmission range

$$r = 8R_v = \frac{80}{\sqrt{\pi}}\sqrt{\frac{\log n}{n}}.$$

[1] In the "Random Networks" scenario [9], the destination is randomly chosen for each source node.

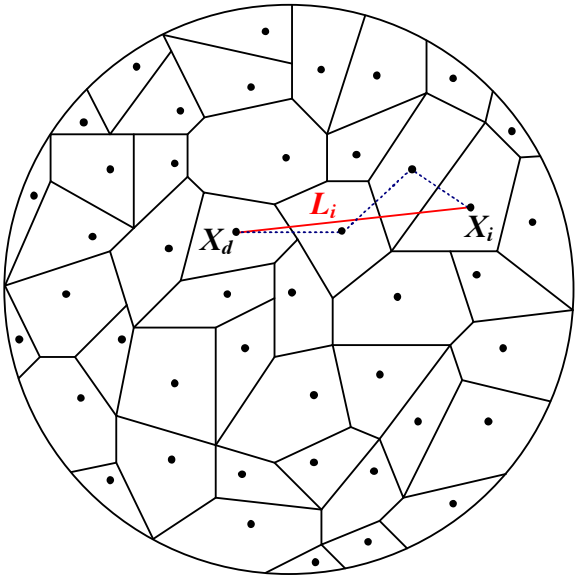


Figure 2.3: A Voronoi tessellation of a set of nodes in the disk.

The following lemma on the properties of the constructed Voronoi tessellation $\mathcal{V}_n$ has been proved in [9]:

Lemma 2.5. *For the chosen R_v and r, the constructed Voronoi tessellation $\mathcal{V}_n$ has the following properties:*

- *Each Voronoi cell in $\mathcal{V}_n$ contains at least one node (with probability approaching 1 as $n \to \infty$).*
- *Every node in a Voronoi cell is within a distance r from every node in its own cell or adjacent cells.*
- *Every Voronoi cell in $\mathcal{V}_n$ has no more than c_3 interfering neighbors, and c_3 depends only on the guard zone size $\triangle$.*

The first two properties of Lemma 2.5 imply that, there always exists direct communication between adjacent cells, and it is feasible to relay traffic from one Voronoi cell to one of its adjacent cells.

Now let X_d be the common destination node, X_i be a source node, and L_i be the straight-line segment connecting X_i to X_d, for $i = 1, \ldots, n$, as shown in Figure 2.3. We choose the routes to approximate the straight-line segments, which intersect the Voronoi cells in $\mathcal{V}_n$. More specifically, each packet is relayed from the source cell to the *destination cell* (the Voronoi cell contains the destination node X_d, denoted C_d) in a sequence of hops (as the 3-hop dotted lines in Figure 2.3). According to Lemma 2.5, in each hop, the packet can be transferred from one Voronoi cell to an adjacent cell along the straight-line segments. In the last hop, the packet is sent to the destination from an adjacent cell of the destination cell C_d. The traffic generated at any source node that is located within C_d is directly transferred to the destination. The above multi-hop relaying scheme is feasible with probability one when n is large.

Next we compute the mean number of routes served by the destination cell C_d. Since C_d contains the final destination node and each multi-hop route from the source to the destination is feasible, it follows

$$\text{Prob}(L_i \text{ intersects } C_d) = 1,$$

and therefore

$$\text{E}[\text{number of lines in } \{L_i\}_{i=1}^{n} \text{ intersecting } C_d] = n.$$

Suppose each line L_i carries traffic of rate λ bits per second. Since the traffic handled by a Voronoi cell is proportional to the number of lines passing through it, the rate at which the destination cell needs to serve is $n\lambda$.

From the vertex coloring of graphs [50], we know that a graph of degree no more than D can have its vertices colored by using no more than $1 + D$ colors in such a way that no two adjacent vertices have the same color. According to the third property of Lemma 2.5, all Voronoi cells in $\mathcal{V}_n$ can be colored with no more than $1 + c_3$ colors such that the colors of interfering neighboring

are different. By equating time slots with colors, we can design a schedule for transmitting packets in such a way that each Voronoi cell gets one time slot to transmit in every $1 + c_3$ slots and no transmission collisions occur. Thus the rate at which each Voronoi cell gets to transmit is $W/(1+c_3)$ bits per second.

Note that the traffic can be accommodated if it is less than the rate available. Based on the scheduling scheme as described, the traffic handled by the whole network is restricted by the destination cell C_d, i.e. $n\,\lambda \leq W/(1 + c_3)$. Therefore, the rate per source node

$$\lambda \leq \frac{W}{(1 + c_3)\, n} \tag{2.2}$$

is feasible with probability one when n is large. Equation (2.2) gives an achievable rate per source node (i.e., the lower bound of the per source node capacity). On the other hand, since C_d can only handle data at rate of W bits per second, the aggregate capacity is upper bounded by W. Therefore, the order of the aggregate throughput capacity is given by

$$T = \Theta(W),$$

when the transmission range is chosen as $r = \frac{80}{\sqrt{\pi}}\sqrt{\frac{\log n}{n}}$.

□

We note that the aggregate throughput capacity of the Random Aggregate Network given in (2.1) has the same form as that of a single-hop random aggregate network, e.g. in a cell of a cellular system in which all transmissions are between the base station and terminals and are single-hop. This implies that the asymptotic capacity of a random network having the aggregation traffic is independent of the number of hops required to reach the aggregation node as well as the location of the aggregation node.

The result on capacity has the same expression as the achievable throughput of a randomly deployed flat network using multi-hop transmission for any-to-one communication obtained in [51].

In our analytical approach, the aggregation node can be randomly placed, and we can obtain not only the asymptotic throughput capacity, but also the traffic distribution in the network (see Appendix A). The traffic distribution could be helpful in designing scheduling algorithms for such a network.

The results on aggregate throughput capacity hold for the network in Figure 2.2(a), in which the source nodes are regularly placed. This is because the multi-hop relaying scheme described before is feasible in a deterministic manner. Note that for regular networks, the optimal transmission range can be reduced as long as

$$r = \Omega(1/\sqrt{n}) \tag{2.3}$$

is satisfied for maintaining connectivity.

2.5 Capacity of Three-Tier Hierarchical Hybrid Wireless Networks

2.5.1 Low-Tier Capacity and Transmission Range

Low-tier transmissions use frequency f_L with bandwidth W_L. The disk is divided into n_F sub-clusters, each of which consists of one FN and its associated MN's. Each MN communicates with the nearest FN via one-hop transmissions, as shown in Figure 2.1. We observe that all low-tier transmissions have to go through the associated FN's, and each FN can only handle data at rate of W_L bits per second at any time. Therefore, the per sub-cluster throughput capacity, η_L, is upper bounded by W_L. For the lower bound, since each MN is one-hop away from its associated FN, there is a schedule for each MN to communicate with its associated FN in a round robin fashion, resulting in a throughput of W_L. Hence, it follows

$$\eta_L = \Theta(W_L).$$

There may exist interference among sub-clusters. Note that, in the Protocol Interference Model, if two sub-clusters are not interfering neighbors, transmissions in one sub-cluster do not interfere

with transmissions in the other sub-cluster.

The regular placement of FN's results in a hexagon tessellation with each hexagon corresponding to the area of a sub-cluster. We can prove the following lemma.

Lemma 2.6. *Each sub-cluster (or hexagon) has no more than c_L interfering neighbors, and c_L is a constant that depends only on Δ_L, when the transmission range satisfies*

$$r_L = O(1/\sqrt{n_F}). \tag{2.4}$$

Proof: Let $b = c\,r_L$ denote the length of each side of the hexagon. Since the disk is divided into n_F hexagons, we have $b = \Theta(1/\sqrt{n_F})$. Combined with the condition $r_L = O(1/\sqrt{n_F})$, it is implied that

$$c = \Omega(1). \tag{2.5}$$

Under this setting, each hexagon contains a disk of radius $\frac{\sqrt{3}}{2}c\,r_L$ and is contained by a disk of radius $c\,r_L$.

Now, we consider a hexagon H. In the Protocol Interference Model, every interfering neighbor of H has (at least) one point within a distance of $(2+\Delta_L)r_L$ of some point in H. Therefore, all the interfering hexagons of H must be contained by a disk D of radius $3c\,r_L + (2+\Delta_L)r_L$. Since the area of each hexagon is larger than the area of the contained disk, the number of hexagons contained in disk D is bounded by:

$$\begin{aligned} c_L &= \frac{\pi\big((3c\,+2+\Delta_L)r_L\big)^2}{\pi\big(\frac{\sqrt{3}}{2}c\,r_L\big)^2} \\ &= \frac{4}{3}\Big(3+\frac{2+\Delta_L}{c}\Big)^2. \end{aligned}$$

Combining with (2.5), we have the desired result.

□

According to the vertex coloring result of graph theory [50], there is a transmission schedule such that each sub-cluster gets one time slot to transmit in every $1 + c_L$ time slots. Therefore, the aggregate throughput capacity contributed by low-tier transmissions, denoted T_L, is given as

$$T_L = \Theta(n_F \, \eta_L) = \Theta(n_F \, W_L).$$

Suppose each MN carries traffic of rate λ_M bits per second on average, then

$$T_L = n_M \, \lambda_M = \Theta(n_F W_L). \tag{2.6}$$

It is observed that the aggregate throughput capacity contributed by low-tier transmissions increases linearly with the number of FN's, when the transmission range satisfies $r_L = O(1/\sqrt{n_F})$. Traffic carried by MN's instead of their exact number counts in (2.6), which implies that the system's scaling behavior does not depend on the number of ad hoc nodes. This is a great improvement upon flat and two-tier wireless networks, achieved by adding FN's to aggregate traffic for ad hoc nodes, especially when the density of ad hoc nodes is high. It also suggests that we can increase either the number of FN's or the bandwidth allocated to low-tier transmissions to accommodate the traffic of the network.

2.5.2 High-Tier Capacity and Linear Scaling Regime

We assume that the bandwidth allocated to the Internet traffic and the local traffic carried by high-tier transmissions (denoted *high-tier Internet traffic* and *high-tier local traffic*) are W_{H1} and W_{H2}, respectively, and $W_{H1} + W_{H2} = W_H$.

High-Tier Internet Traffic

The results of Random Aggregate Networks, given in Theorem 2.3, can be applied to each cluster with the high-tier Internet traffic. Therefore, the per cluster throughput capacity contributed by the

high-tier Internet traffic is given by

$$\eta_{H1} = \Theta(W_{H1}),$$

where W_{H1} is the bandwidth allocated to the high-tier Internet traffic.

There are n_A clusters in the hierarchical network. Notice that the regular placement of AP's result in a hexagon tessellation with each hexagon corresponding to the area of a cluster. By applying Lemma 2.6, the number of interfering neighbors of each cluster is bounded by a constant which depends only on Δ_H, if the high-tier transmission range satisfies

$$r_H = O(1/\sqrt{n_A}). \tag{2.7}$$

Thus, there is a scheduling such that each cluster gets one slot to transmit in every constant number of time slots. Therefore, the aggregate throughput capacity contributed by high-tier Internet traffic is given by

$$T_{H1} = \Theta(n_A W_{H1}). \tag{2.8}$$

According to (2.3), the high-tier transmission range for regularly placed FN's is chosen as

$$r_H = \Omega(1/\sqrt{n_F}). \tag{2.9}$$

Equations (2.7) and (2.9) imply

$$n_A = O(n_F), \tag{2.10}$$

which, we will see later, is the upper bound of the scaling regime for the high-tier capacity to scale linearly with the number of wired AP's.

In case that FN's are randomly placed, according to Theorem 2.3, the high-tier transmission range has to be chosen as

$$r_H = \frac{80}{\sqrt{\pi}}\sqrt{\frac{\log n_F}{n_F}}. \tag{2.11}$$

And the corresponding upper bound is

$$n_A = O(n_F/\log n_F). \tag{2.12}$$

High-Tier Local Traffic

Suppose the bandwidth allocated to high-tier inter-cluster and intra-cluster traffic are W_{H2_inter} and W_{H2_intra} respectively, and $W_{H2\_inter} + W_{H2\_intra} = W_{H2}$.

(1) Inter-Cluster Local Traffic

Inter-cluster traffic is assumed to always go through the infrastructure. In particular, the inter-cluster traffic enters the infrastructure at the source FN's associated AP and leaves it at the destination FN's associated AP. Note that this may not be the optimal route. For instance, if the source and destination FN's are neighbors, it may be preferable to have direct FN-FN connections rather than using the infrastructure. However, under the uniform traffic assumption (destination node uniformly chosen within the disk), the fraction of such local connections (source and destinations are close neighbors) goes to zero when the number of FN's goes to infinity. So in this case, the adverse effect of the routing protocol is negligible asymptotically.

According to our routing assumption, in the wireless network, one inter-cluster communication can be decomposed into two parts: one is from FN to AP in the source cluster, and the other is from AP to FN in the destination cluster. Only one part is counted in the throughput capacity.

For each cluster, we can apply Theorem 2.3 and obtain the aggregate throughput capacity contributed by high-tier inter-cluster local traffic as follows:

$$T_{H2\_inter} = \Theta(n_A W_{H2\_inter}), \tag{2.13}$$

when r_H is chosen as in (2.7).

(2) Intra-Cluster Local Traffic

Suppose AP's do not participate in transferring this kind of traffic. There are n_A clusters in the network, and n_F/n_A FN's in each cluster. According to the scaling property of n_A with respect to n_F, there are two cases: $n_A = o(\sqrt{n_F})$ and $n_A = \Omega(\sqrt{n_F})$. We apply Lemmas 2.1 and 2.2 to the analysis below.

(i) $n_A = o(\sqrt{n_F})$*:* In this case, according to Lemma 2.1, the aggregate capacity contributed by high-tier intra-cluster local traffic is given as

$$T_{H2\_intra} = \Theta\left(\sqrt{\frac{n_F}{\log(n_F/n_A^2)}}\, W_{H2\_intra}\right).$$

Taking into account (2.13) and applying Lemma 2.1 again, the aggregate throughput capacity contributed by all high-tier local traffic is maximized when $W_{H2\_intra}/W_{H2} \to 1$. The achieved capacity is given as:

$$T_{H2} = \Theta\Big(\sqrt{\frac{n_F}{\log(n_F/n_A^2)}}\, W_{H2}\Big), \tag{2.14}$$

when $n_A = o(\sqrt{n_F})$. Hence in this case it is more beneficial to assign bandwidth to intra-cluster traffic.

(ii) $n_A = \Omega(\sqrt{n_F})$*:* According to Lemma 2.2, in this case, the aggregate throughput capacity contributed by high-tier intra-cluster local traffic is given as

$$T_{H2\_intra} = O(\sqrt{n_F}\, W_{H2\_intra}).$$

When $W_{H2\_inter}/W_{H2} \to 1$, the aggregate throughput capacity contributed by all high-tier local traffic is maximized, and the achieved capacity is

$$T_{H2} = \Theta(n_A\, W_{H2}), \tag{2.15}$$

when $n_A = \Omega(\sqrt{n_F})$. This shows that it is more effective to allocate bandwidth to carry inter-cluster traffic in this case, and the achieved capacity increases linearly with n_A.

Therefore, when $n_A = \Omega(\sqrt{n_F})$, the capacity improvement achieved by adding wired infrastructure nodes is significant. Note that $n_A = \Omega(\sqrt{n_F})$ is the lower bound of the linear scaling regime, while (2.10) gives the upper bound for the chosen transmission range.

Capacity of High-Tier Transmissions

From (2.8), (2.10), (2.14), and (2.15), the achieved aggregate throughput capacity contributed by all high-tier transmissions, denoted T_H, can be achieved as follows:

- For $n_A = o(\sqrt{n_F})$,

$$\begin{aligned} T_H &= T_{H1} + T_{H2} \\ &= \Theta\big(n_A W_{H1}\big) + \Theta\Big(\sqrt{\frac{n_F}{\log(n_F/n_A^2)}}\, W_{H2}\Big). \end{aligned} \tag{2.16}$$

- For $n_A = \Omega(\sqrt{n_F})$ and $n_A = O(n_F)$,

$$T_H = \Theta\big[n_A(W_{H1} + W_{H2})\big] = \Theta(n_A W_H). \tag{2.17}$$

These results are obtained when r_H is chosen as $r_H = O(1/\sqrt{n_A})$.

Equation (2.17) reveals that the capacity improvement achieved by adding AP's can be significant, if n_A grows asymptotically faster than $\sqrt{n_F}$ but slower than n_F, and the achieved capacity has the linear relationship with the number of AP's. This capacity is shared among the nodes whose packets are routed through the infrastructure as determined by the defined routing scheme.

2.6 Conclusions and Discussions

In this chapter, we present an analytical model for the capacity of the proposed three-tier hierarchical hybrid wireless network, and identify conditions on transmission range and node density for scalability to be maintained. It is shown that in a three-tier network of n_A AP's, n_F FN's and n_M

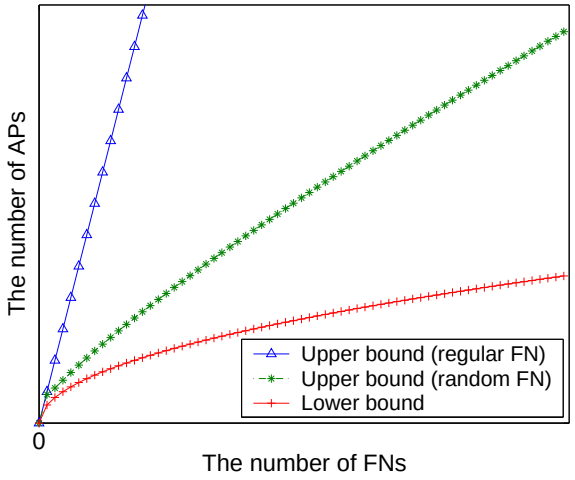


Figure 2.4: The asymptotic upper and lower bounds of the linear scaling regime.

MN's, the low-tier capacity increases linearly with n_F when the low-tier transmission range satisfies $r_L = O(1/\sqrt{n_F})$. Also, the high-tier capacity increases linearly with n_A in the scaling regime that satisfies $n_A = \Omega(\sqrt{n_F})$ and $n_A = O(n_F)$, when the high-tier transmission range satisfies $r_H = O(1/\sqrt{n_A})$. The asymptotic upper and lower bounds of n_A as a function of n_F is plotted in Figure 2.4. Also note that the upper bound is closely related to the chosen transmission range, and the linear scaling regime can be extended by reducing the transmission range when FN's are regularly placed. In this identified linear scaling regime, the capacity improvement achieved by adding AP's can be significant. The upper bound implies that further investments in the infrastructure do not lead to improvement in capacity scaling.

It is observed that in the three-tier network, AP's do not need to cover the whole area, thus the network coverage is improved. The number of AP's required is related to the number of FN's rather than that of ad hoc nodes (i.e. MN's) for scalability to be maintained, and the system's scaling behavior does not depend on the number of ad hoc nodes. Therefore, compared to the two-tier hybrid network model in [17], the same amount of traffic can be served by adding a new tier of FN's and reducing the number of AP's. Since the investment and recurrent wired access cost for AP's or BS's is significantly higher than that for FN's, system costs can be significantly reduced with the

three-tier approach, especially when the density of ad hoc nodes is high.

By dividing wireless channels into sub-channels to carry the traffic of different tiers of the hierarchy, the proposed analytical model can be generalized to a multi-tier hybrid wireless network with more than one relay tiers. Under the deterministic routing strategy (i.e. strictly hierarchical routing), the linear scaling regions of the capacities of higher tiers can be determined accordingly in the generalized model.

The above asymptotic capacity and scaling regime are obtained by deploying a specific system model and the deterministic routing approach, similar to the two-tier model in [17]. There are other studies on two-tier hybrid networks, which use different system models and routing schemes, leading to different asymptotic bounds for different scaling regimes [18, 19]. However, these two-tier hybrid networks have the same scaling behavior as that in [17] in terms of: first, infrastructure investments need to be high in order to obtain a significant capacity gain; second, the scaling regime depends on the number of ad hoc nodes. It can be shown that the capacity improvement and infrastructure cost saving have similar properties when extending these results to our proposed three-tier architecture.

Since both the low-tier and high-tier transmissions are involved in each traffic flow, numerically low-tier capacity needs to be equal to high-tier capacity. Our analysis suggests that, when designing the three-tier hierarchical network, we can adjust n_A and n_F (or W_L and W_H) such that FN's can accommodate network traffic and the scaling properties can be achieved by satisfying $n_A = \Omega(\sqrt{n_F})$ and $n_A = O(n_F)$. In the next chapter, we will study the scaling properties of a three-tier 802.11-based network via detailed system simulations, and show how protocol overhead affects the system's scaling behavior.

Chapter 3

Three-Tier Hierarchical System Evaluation

The theoretical analysis presented in Chapter 2 gives the asymptotic capacity of the three-tier hierarchical hybrid network. We note that the analytical results on capacity were obtained by assuming perfect spatial and temporal scheduling and ignoring medium access and routing overhead. However, these assumptions are not true in practice. Taking the IEEE 802.11 specification as an example, the backoff mechanism would cause idle times, and the virtual carrier sense mechanism (used to overcome the hidden terminal problem) would result in some part of the space not being fully used. The overhead of wireless medium access and routing control in system and protocol design can be quite high and will generally tend to degrade system performance. Hence, we investigate the system performance and scaling properties of the three-tier network system with MAC and routing protocols typical of a real implementation. It would be of interest to see whether the analytically obtained scaling relationships hold for system simulations with realistic protocol and traffic assumptions.

Meanwhile, the analytical results prove that the aggregate throughput capacities of both tiers have linear relationships with the allocated bandwidth. But they do not reveal explicitly how to allocate system bandwidth to each tier of the transmissions to accommodate network traffic in a balanced way. In this chapter, we validate the analytically obtained capacity results and further study system performance with detailed system simulations. For this purpose, we set up an 802.11-based hierarchical hybrid network simulation model using the *ns*-2 network simulator [52]. The example three-tier 802.11-based network consisting of mobile nodes (MN's), forwarding nodes

(FN's), and access points (AP's) is shown in Figure 1.5.

Firstly we compare the hierarchical system performance of dense network scenarios with a conventional flat ad hoc network in order to estimate the potential performance improvement with the three-tier hierarchy. As discussed in Section 2.6, the performance improvements and capacity gain with the three-tier hierarchy come at the expense of increased investment in multi-radio forwarding node hardware and total system bandwidth. This motivates us to next study the scaling behavior of such a network as a function of key parameters such as relative node densities, traffic pattern, and channel bandwidth allocation at each tier of the network. We also apply different ad hoc routing protocols with appropriate modifications to the hierarchical system and evaluate how different routing protocols work in the hierarchical mode.

Combined with the analytical results, these experimental results are intended to provide a more complete understanding on how hierarchy, infrastructure, and multi-channel capability help in the design of a scalable network.

3.1 Methodology and Simulation Model

We assume a network in which all nodes use the IEEE 802.11 MAC protocol with Distributed Coordination Function (DCF) [4]. A static discovery procedure pre-computes a well-balanced hierarchical network topology, and also maintains and optimizes the topology in response to node movements and varying network traffic. We consider different ad hoc routing protocols, including Dynamic Source Routing (DSR) [12], Ad hoc On-demand Distance Vector (AODV) [13], and Destination Sequence Distance Vector (DSDV) [33], which are modified appropriately for use in the hierarchical network.

3.1.1 Hierarchy Construction and Node Modeling

First, ad hoc nodes (i.e. MN's) are randomly distributed in the network. Then FN's and AP's are added to the network to form the cluster-based hierarchy. In particular, AP's are placed in a regular pattern. MN's and FN's are associated to a nearest AP based on their geographic distance from the AP's, assuming each node has the location information of itself and all the AP's in the network. As in the analytical model described in Section 2.3, each cluster consists of one AP and an arbitrary number of FN's and MN's, and the AP works as the gateway of the cluster to the infrastructure; "single homing" is used, i.e. any FN or MN can only belong to one cluster. This assumes that the discovery protocol supports identification of gateway AP's and association of related FN's and MN's in each cluster.

As discussed in Section 1.1.1, the hierarchical system can be designed flexibly. Using an appropriate discovery protocol with the MAC protocol, an "optimized" self-organizing hierarchical topology can be maintained over time. For example, an optimized hierarchy may be achieved by introducing dynamic node reassociations to balance workload and manage node mobility over clusters. Under this assumption, we can simply implement the hierarchical network by dividing the simulated site into a certain number of clusters, with a gateway AP located at the center of each cluster and approximately the same number of nodes and same traffic load in each cluster. The nodes in movement are assumed to be moving within the original cluster sites. In this way, simulations can always be conducted over well-balanced hierarchical topologies.

MN's are modeled as simple wireless nodes without routing capability offered to any other node. FN's offer multi-hop routing capability to the nodes of their own clusters. AP's provide both wireless and wired access, and are fully connected by 100 Mbps high-speed wired links. The delay caused by the wired link is defined as the packet transmit time plus propagation delay, where the former is the ratio of the packet size to the link bandwidth, and the latter is 2 microseconds (μs)

(which is equivalent to a cable of length 500 meters). With *ns*-2 wired-cum-wireless scenarios [52], each communication between clusters goes through AP's, as our analysis assumes.

We have implemented dual-radio nodes with two network interfaces, which allow simultaneous transmissions over two non-interfering channels. In particular, two network interfaces with separate protocol stacks of link layer, MAC layer, and PHY layer are created below the routing agent in the "mobilenode" object of *ns*-2. The routing protocol decides the network interface by which each data packet is transmitted. In order to disseminate routing information over the network, routing messages are sent via both network interfaces.

Although the analysis suggests use of optimal transmission range by satisfying (2.4) or (2.7), for simplicity we choose a common transmission range of 250 meters (m) for both tiers. The interfering range is 550 meters.

3.1.2 Traffic Pattern

We consider both the Internet and local traffic, which have been defined in Section 2.3.2, and only uplink is considered for the Internet traffic. Therefore, all traffic in the network is originated at the MN's. The relative proportions of these two types of traffic can be adjusted parametrically. For example, in a sensor network scenario, 80% of the traffic is assumed to be bound for a server within the Internet, accessed through an AP; the remaining 20% of MN traffic is assumed to be routed to other MN's in the network, accessed via AP's and/or FN's.

At MN's, traffic is generated according to an exponential on/off model [52], in which both the "on" (burst) and "off" (idle) periods are taken from the exponential distribution with an average of 500 milliseconds (ms). Packets are sent at a specific rate only during "on" periods, and this packet generation rate per source node (in bits per second) is varied as an input parameter in order to gradually increase the offered load to the network. The packet size is 64 bytes. Each MN can

simultaneously support up to two traffic flows to different destinations.

3.1.3 Performance Metrics

We use the following performance metrics for system evaluation:

- *Packet delivery fraction*: measured as a ratio of the number of data packets delivered to their eventual destinations and the number of data packets generated by sources.

- *Average end-to-end delay*: includes all possible delays before data packets arrive at their destinations.

- *Normalized routing overhead*: measured as the number of routing packets transmitted (in the wireless network) per data packet delivered at destinations. Each wireless hop is counted as one transmission for both routing and data packets sent over multi-hop paths.

- *System throughput*: measured as the total number of bits of data received at destinations over simulated time (in bits per second i.e. bps).

- *MAC throughput*: measured as the total number of bits of data sent by all nodes per second, including forwarded bits. Only data packets delivered at destinations count. MAC throughput is used as a measure of traffic load in Section 3.4.

The simulations are run for multiple independent replications [52] with different ad hoc node placements or source-destination pair distributions. Each simulation result represents an average of 5 independent runs lasting 550 seconds of simulated time.

Table 3.1: Simulation parameters for baseline comparison

Simulation dimension	$1000\text{m}\times1000\text{m}$
Number of clusters; AP's; FN's; MN's	4; 4; 20; 100
Packet generation rate per source (Kbps)	1,4,8,12,16,24,32
Packet size	64 bytes
Fraction of the Internet traffic	100%
Number of communication pairs	20/40/60
MAC	802.11b ad hoc mode
Radio data rate; Radio range	1 Mbps; 250 meters
AP-AP wired link speed	100 Mbps

3.2 Baseline Comparison

3.2.1 Simulation Parameters

Our baseline simulations are for an example dense sensor network deployed over a square geographical area with dimension $1000\text{m}\times1000\text{m}$. We divide the coverage area into four $500\text{m}\times500\text{m}$ smaller squares, each corresponding to a cluster with one AP and several FN's and MN's. FN's and MN's are randomly placed within the clusters with a nominal uniform density of 20 FN's and 100 MN's spread over the entire coverage area. FN's move according to the random waypoint model [12] with a randomly chosen speed (uniformly distributed between 0 and 1 m/s) and a pause time of zero (i.e. FN's do not stop during their journey). Half of the MN's are static; the remaining half move according to the same random waypoint model as FN's. In order to measure the system improvement achieved by the three-tier hierarchy, we use single radio FN's for baseline comparison as the single-frequency system may be considered as an imperfect type of dynamic allocation via the MAC protocol. Also we only consider the Internet traffic here. The key parameters are summarized in Table 3.1.

We assume that there are separate entries (i.e. AP's) into the Internet, each of which serves some set of MN's. In the flat ad hoc network, each packet uses multi-hop wireless path to its assigned AP without the help of infrastructure. Furthermore, there is no FN in the flat network, and MN's deploy

the same distribution and mobility pattern as in the hierarchical peer.

3.2.2 Results and Discussions

Simulations with DSR

From Figure 3.1(a) which shows throughput as a function of offered load from MN's for 40 communicating pairs with DSR, we see that the hierarchical system begins to saturate when the packet generation rate per source reaches 16 Kbps; while the flat system saturates at about 4 Kbps. For the 802.11b bandwidth of 1 Mbps used here, system capacities are found to be around 320 Kbps for the hierarchical case and about 77 Kbps for the flat case, respectively.

It is observed that, for a specific network model with 4 AP's, the system capacity roughly increases by a factor of 4 if the proposed hierarchical architecture is adopted. This is a significant scaling increase over the relatively low 77 Kbps obtained with the flat network. The increase factor is consistent with the number of AP's deployed. We will look at different number of AP's in Section 3.3. Meanwhile, the average end-to-end delay, packet delivery fraction, and routing overhead curves are illustrated in Figure 3.1(b), (c), and (d). The improvement of packet delivery fraction shows that the three-tier hierarchy helps deliver packets in mobility scenarios. Figure 3.2 demonstrates delay-throughput curves which summarize system capacity and performance as a whole. The simulations have been repeated for two other cases corresponding to 20 and 60 communication pairs and results similar to the 40-pair case have been observed. This verifies that the system capacity is independent of the number of MN's, as the analytical result (2.6) implies.

In the cluster-based hierarchy, each MN communicates through a few FN's and a gateway AP, thus the infrastructure serves as shortcut for the traffic over long distance to reduce the average number of hops required to reach the destination, which is the Internet where 100% packets from MN's have their destinations here. In addition, MN's do not join the full distribution of routing

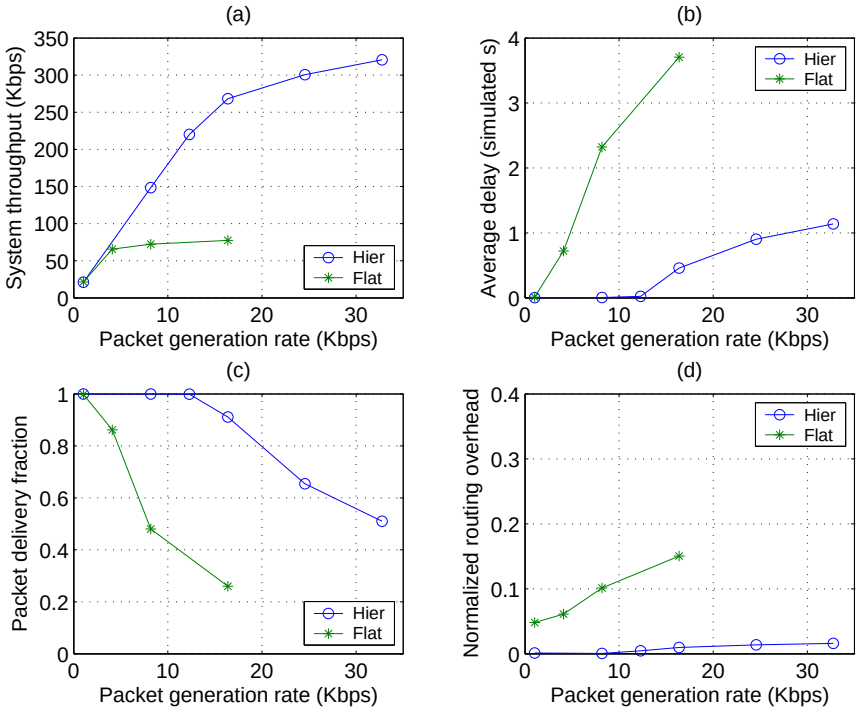


Figure 3.1: Baseline comparison. DSR case, 40 communication pairs. (a) System throughput. (b) Average end-to-end delay. (c) Packet delivery fraction. (d) Normalized routing overhead.

messages, thus reducing routing overhead significantly. Of course, the capacity increase comes at the expense of increased hardware (FN's and AP's) relative to a flat network, and in that sense it is not an "apples-to-apples" comparison.

Simulations with AODV

We replace DSR with AODV routing and repeat the simulations in order to study the impact of routing protocols. The *ns*-2 AODV implementation is from Uppsala University [53]. We use the same system model and parameters as for the DSR case described earlier. The simulation results obtained are shown in Figure 3.3 and Figure 3.4.

From these curves it is observed that the system capacity and performance improvement obtained with AODV are comparable to those obtained with DSR. In particular, the throughput gain is also approximately 4. This is mainly because of the deployment of three-tier hierarchy and wired

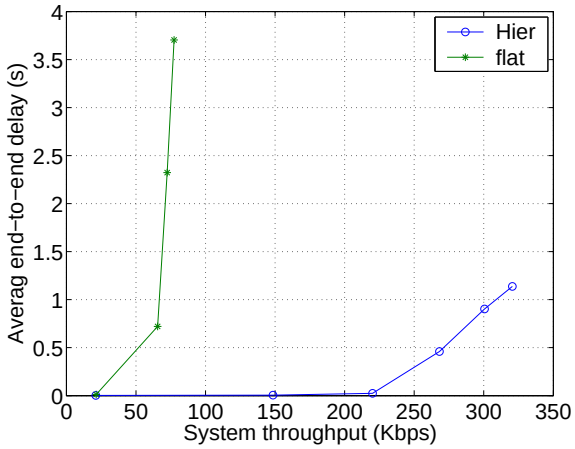


Figure 3.2: Baseline comparison. Delay vs. throughput. DSR case, 40 communication pairs.

integration, as discussed in the DSR case.

When comparing the results, it is observed that the three-tier hierarchy provides about 4 times higher saturated throughput (capacity) as the flat network, and both have similar delay vs. offered load characteristics. The capacity of AODV is found to be marginally higher, and the average packet delays are also correspondingly lower. The AODV case does show a poorer fraction of packets delivered, possibly due to higher rates of link delivery failure. Previous simulation studies show that AODV generally has a higher routing load than DSR [54]. This is because DSR's caching is very efficient at low speed, which is our case here. AODV's routing load is dominated by route request packets. In the hierarchical network, MN's do not join in the flooding of route requests, so AODV can achieve a low routing overhead comparable to DSR. At the same time, fewer routing packets decrease the chance of link delivery failure, which results in higher data packet delivery fraction in the hierarchical mode.

We conclude that the performance and achievable throughput of the three-tier network are relatively insensitive to the choice of routing protocol between DSR and AODV. AODV appears to have marginally higher capacity and lower delay than DSR, but with lower packet delivery fraction for the particular scenario and configuration parameters. The hierarchical system works well with

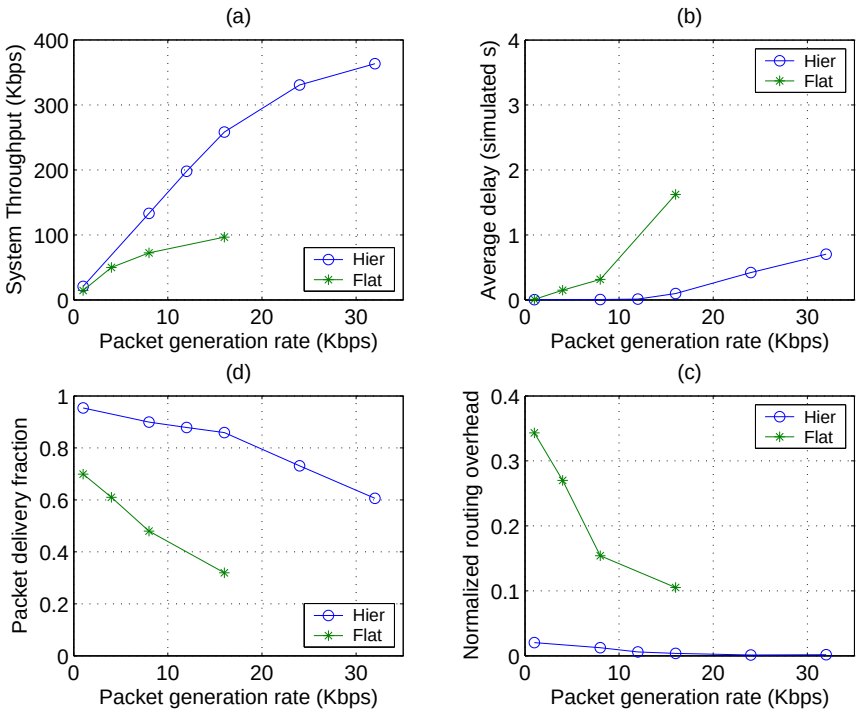


Figure 3.3: Baseline comparison. AODV case, 40 communication pairs. (a) System throughput. (b) Average end-to-end delay. (c) Packet delivery fraction. (d) Normalized routing overhead.

on-demand routing protocols with low mobility FN's. We will look at system performance with another category of routing protocols, proactive routing protocols, taking DSDV as the example in the next section.

3.3 Scaling Behavior: Impact of Node Density

Our experiments for example dense network scenarios with nominal parameters chosen above have shown that the system throughput increases significantly with a scaling factor approximately equal to the number of AP's deployed. It is also verified that the system throughput is independent of the number of ad hoc nodes. As identified by the analytical results given in Chapter 2, the precise capacity scaling factor depends upon several factors including the topology, the spatial distributions of FN's and AP's, the ratio of FN's to AP's, and the traffic pattern. Making the same assumptions on the topology and the spatial distributions of FN's and AP's, we investigate the impacts of the

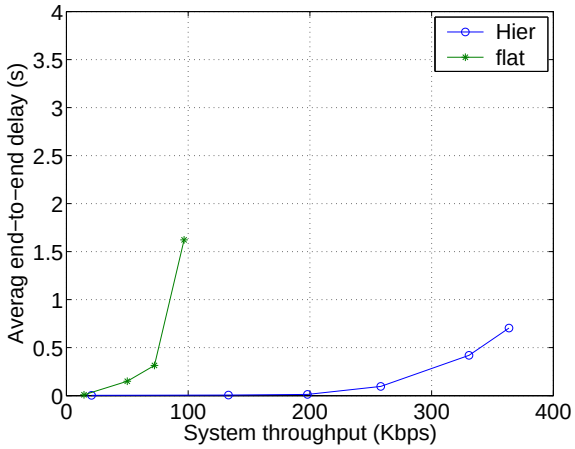


Figure 3.4: Baseline comparison. Delay vs. throughput. AODV case, 40 communication pairs.

relative densities of FN's and AP's and the traffic patter on capacity scaling in this part of work.

3.3.1 Simulation Parameters

For the three-tier hierarchy under consideration, we observe that the core wireless network is formed by FN's and AP's, while MN's feed traffic into nearby FN's or AP's, as illustrated in Figure 3.5(a). As a result, the key parameters for the system's scaling behavior include the offered traffic load density from MN's (denoted λ_S, in bps/m^2), the density of FN's (denoted $\mathcal{X}_{FN}$), and the density of AP's (denoted $\mathcal{X}_{AP}$). Clearly, the numbers of FN's and AP's cannot be selected arbitrarily given that FN's and AP's must cover the entire service area to ensure that all MN's are reachable. Note also that in the three-tier network, an AP represents a significantly higher investment than an FN, so we adopt a heuristic approach of covering the entire service area with FN's and then determining the right number of AP's necessary for the network to scale in a balanced way. We measure the normalized system throughput (per unit area) as a function of the offered load density λ_S while varying the density of AP's $\mathcal{X}_{AP}$, given a fixed value of the density of FN's $\mathcal{X}_{FN}$, and obtain the normalized system capacity as the maximum throughput.

We consider a square simulated region, where AP's and FN's are placed in a regular pattern.

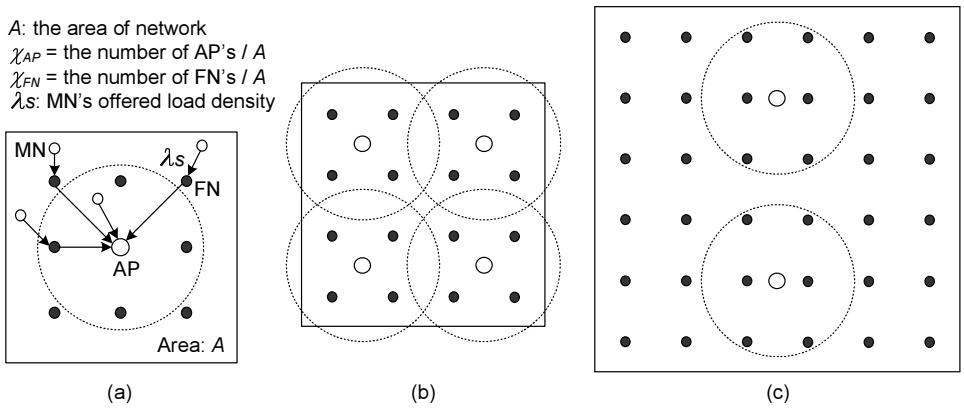


Figure 3.5: Parameter definitions and the regular planar network topology. (a) FN's and AP's form the ad hoc core network. (b) A grid topology consisting of 16 FN's and 4 AP's (MN's are not given in the plot). (c) A grid topology consisting of 36 FN's and 2 AP's.

Although this regular spatial model is expected to produce optimistic results relative to random spatial model, it is considered useful for estimating the achievable capacity [55]. The distances between neighboring FN's are all 200 meters. With the specified 802.11b transmission range of 250 meters, this separation is likely to yield close to the maximum throughput [56], and this FN density provides full coverage over the simulated region. We use two simulation cases with dimension of $1200\text{m}\times1200\text{m}$ and $800\text{m}\times800\text{m}$ respectively in order to explore the sensitivity to physical assumptions. With identical FN density, there are a total of 16 FN's for the $800\text{m}\times800\text{m}$ case and 36 FN's for the $1200\text{m}\times1200\text{m}$ case, as depicted in Figure 3.5(b) and (c) (suppose 4 AP's in (b) and 2 AP's in (c)), where MN's are not plotted in the figures.

MN's are randomly distributed in the network with an identical density. In order to investigate the scaling behavior affected by the three-tier hierarchy, we use single radio for FN's. The number of AP's is varied for each dimension to see how the hierarchical throughput changes with the ratio of FN's to AP's. We consider with both DSR (as an on-demand routing protocol) and DSDV (as a proactive routing protocol) for routing. The mobility pattern of MN's and the traffic pattern are set differently for each simulation and will be described separately in later sections. Table 3.2 lists the key parameters. For other parameters please refer to Table 3.1.

Table 3.2: Parameters for scaling simulations

Simulation dimension	1200m×1200m	800m×800m
Number of AP's	1, 2, 3, 4, 6, 9, 12	1, 2, 4, 6
Number of FN's	36	16
Number of MN's	90	40
Packet generation rate (Kbps)	1, 4, 8, 12, 16, 24, 32, 40	
Fraction of the Internet traffic	20%, 50%, 80%, 100%	
Distance between neighboring FN's	200 meters	

3.3.2 Results and Discussions

Experiments with DSR

When DSR is used, MN's move according to the same random waypoint model as in the baseline comparison, and only the Internet traffic is considered.

Simulation results for the 1200m×1200m case are shown in Figure 3.6. Figure 3.6(a) shows that the normalized system throughput increases when the number of AP's increases from 1 to 9. Observe that once the number of AP's reaches 4, the system throughput tends to increase at a slower rate. As expected, the highest capacity is obtained with 9 AP's, since 9 AP's give almost the full coverage over the simulated site in our model. When the number of AP's is greater than 9, the overlapping of the coverage areas of the neighboring AP's becomes a factor, which results in interference between AP's thus the normalized throughput begins to decrease. The normalized throughput with 12 AP's is even less than that with 4 AP's for the light offered load, but it starts to outperform others (except for the 9 AP's case) when the offered load increases. We also observe that the average end-to-end delay decreases when the number of AP's increases.

Figure 3.6(b) summarizes the normalized capacity with different numbers of AP's. We observe that the achievable end-to-end throughput increases almost linearly before the number of AP's increases to 4, while the curve saturates rather rapidly as the number of AP's is increased further. This saturation phenomena verifies that the achievable throughput grows linearly with the number

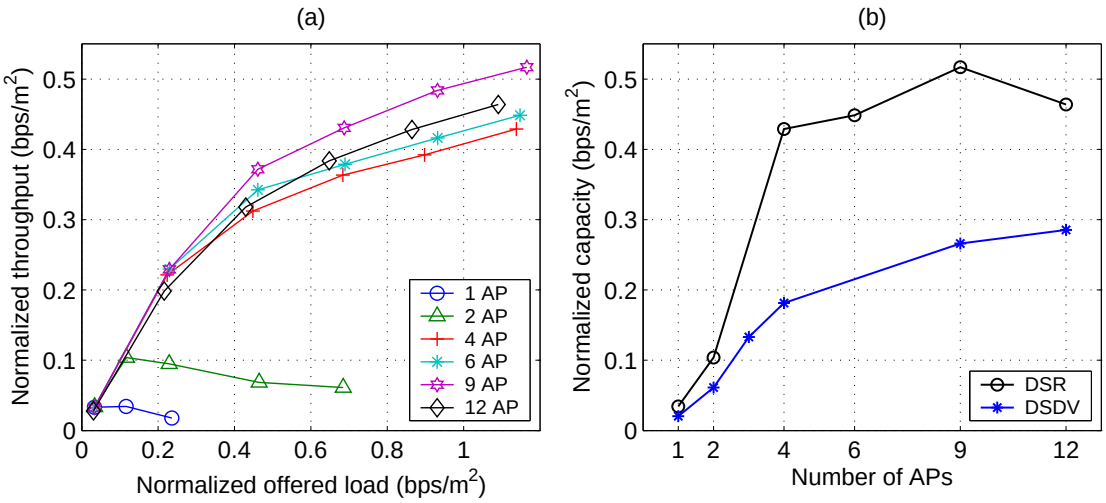


Figure 3.6: Impact of AP density. DSR case, 100% Internet traffic, 1200m×1200m field, 36 FN's. (a) Normalized throughput vs. normalized offered load. (b) Normalized capacity vs. the number of AP's. DSDV curve given for comparison purpose.

of AP's only when AP's are in the right proportion to FN's. Clearly, the "knee" of this curve is a good operating region for system designers because it achieves near-maximum network capacity with a modest investment in wired AP's. In this case, a system designer should aim to provision the network with about 4–6 AP's for a region that requires around 36 FN's for full coverage.

We repeat the experiments for the 800m×800m case and observe similar results as shown in Figure 3.7. In this case, the knee of the capacity curve is reached with about 3–4 AP's. It is also observed that the normalized capacities of these two cases are comparable, as might be expected as the simulated region grows larger.

Experiments with DSDV

For the DSDV case, we assume MN's to be stationary. This setting is used to avoid the influence of node movement, which would tend to degrade system performance in a proactive routing protocol. We adjust the fraction of the Internet traffic to see the impact of traffic pattern, and assign local traffic to be either inter-cluster or intra-cluster traffic with equal probability.

Figure 3.8 depicts the normalized throughput curves of different numbers of AP's for different

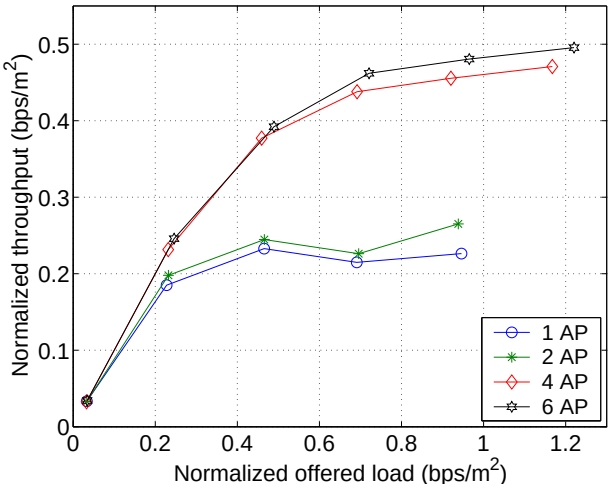


Figure 3.7: Normalized throughput vs. normalized offered load. DSR case, 100% Internet traffic, 800m×800m field, 16 FN's.

Internet traffic fractions, and Figure 3.9 shows the achievable throughput vs. the number of AP's for different Internet traffic fractions. It is observed that the throughput grows as the Internet traffic fraction grows from 20% to 100% for the 2, 3, 4, 9, and 12 AP's cases, but it is reversed for the 1 AP case. The analytical result suggests that, for the local traffic, it is more effective to allocate bandwidth to carry inter-cluster traffic for large number of AP's (i.e. when $n_A = \Omega(\sqrt{n_F})$), and to carry intra-cluster traffic for small number of AP's (i.e. when $n_A = o(\sqrt{n_F})$). Since the Internet traffic has the same scaling property as the inter-cluster local traffic, these simulation results are consistent with the analysis in terms of the traffic pattern.

The capacity curves in Figure 3.9 have similar behavior as the DSR case given earlier, and all the curves approach saturation in the region of 4. When the number of AP's is greater than 9, the throughput hardly increases any further. The saturation phenomena verify a linear scaling regime, in which n_A cannot grow faster than some function of n_F for the linearity to be maintained. Otherwise, the number of interfering neighbors of each AP would increase as the AP density increases, invalidating the linearity. Moreover, a rough square-law relationship between the density of FN's and AP's (i.e. $\mathcal{X}_{AP} \cong \sqrt{\mathcal{X}_{FN}}$) around the knees of the capacity curves may be inferred.

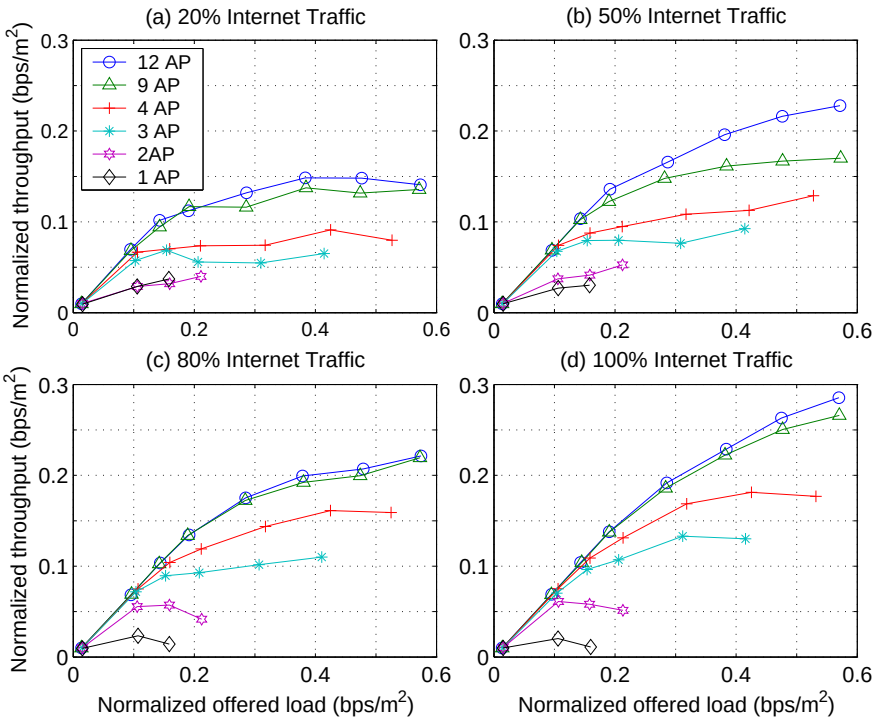


Figure 3.8: Normalized throughput for different number of AP's and different Internet traffic fractions. DSDV case, 1200m×1200m field, 36 FN's. (a) 20% Internet traffic fraction. (b) 50% Internet traffic fraction. (c) 80% Internet traffic fraction. (d) 100% Internet traffic fraction.

The analysis assumes transmission range chosen as in (2.4) or (2.7). Therefore, throughput is expected to be improved by employing an optimal transmission power. Also, the saturation region is expected to move to a larger number of AP's provided that the transmission range is adjustable. The analysis also assumes perfect medium access (including scheduling) and routing. But in practice, the protocol overhead of medium access and routing would degrade system performance. These factors cause deviations from the theoretical bound obtained earlier. Our experiment study provides understanding on the design of hierarchical systems, and we do not look for the exact knees.

Additionally, it is observed that the hierarchical system achieves higher throughput with the on-demand routing protocol than with the proactive routing protocol (see Figure 3.6(b)), as in the flat ad hoc network [36]. This is because the routing overhead can be reduced with the on-demand routing protocol.

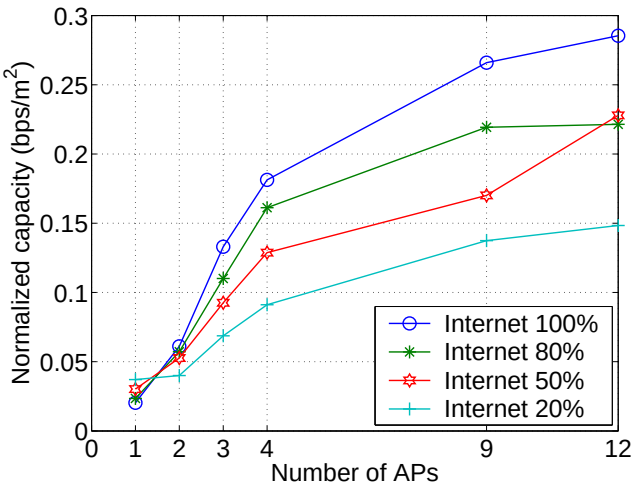


Figure 3.9: Normalized capacity vs. the number of AP's. DSDV case, different Internet traffic fractions, 1200m$\times$1200m field, 36 FN's.

3.4 Scaling Behavior: Impact of Channel Bandwidth Allocation

As proved in Chapter 2, the low-tier capacity grows linearly with n_F and W_L, and the high-tier capacity grows linearly with n_A (in the region shown in Figure 2.4) and W_H. Hence, we might increase either the node density or channel bandwidth to improve capacity. Note that different AP or FN densities require different optimal transmission ranges based on (2.4) or (2.7), which might require an appropriate power control algorithm. As distributed power control is a complex problem in wireless networks and would affect network topology, alternatively we may choose to allocate dedicated frequency bands to different tiers of the network.

When considering the problem of bandwidth allocation, we note that the analytical results do not explicitly reveal how to allocate system bandwidth to each tier of the transmissions to accommodate network traffic. Therefore, we conduct some additional simulations to investigate this issue in this part of work.

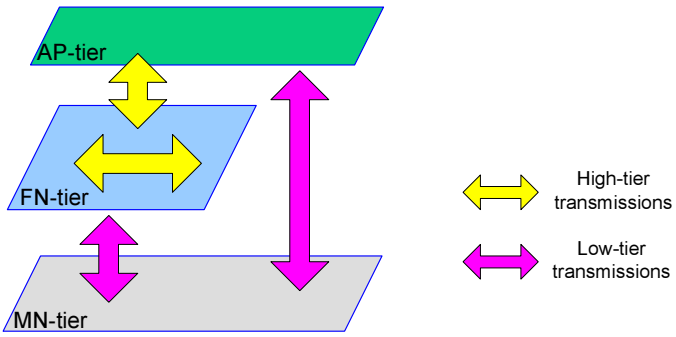


Figure 3.10: Wireless transmissions at two tiers.

3.4.1 System Model and Bandwidth Allocation

Since both the low-tier and high-tier transmissions are involved in each traffic flow, numerically the high-tier capacity is required to be equal to the low-tier capacity. This suggests that when the total network bandwidth is given, it should be allocated in suitable proportions to each tier in order to achieve a good system capacity.

Dual-Frequency System Model

In the analytical model, only FN's are equipped with dual radios, while AP's are assumed to have a single radio. However, this assumption results in an avoidable system performance degradation due to extra hops through FN's (i.e. more wireless transmissions and higher medium access overhead) even when direct MN-to-AP connections are possible. The above limitation of single radio AP's has been verified by simulations for the scenario in Section 3.3.2 showing that there is very little gain from deploying two dedicated channels to two tiers of transmissions and doubling the total system bandwidth. Therefore, we employ a system model which deploys dual radios at both FN's and AP's.

In this model, the high-tier capacity is numerically equivalent to the low-tier capacity less any traffic which does not go through the FN-tier. The wireless transmissions at the high-tier and the low-tier are demonstrated in Figure 3.10, where the low-tier transmissions involve MN's at one end

and higher-tier nodes (either AP's or FN's) at the other end, and the high-tier transmissions do not involve any MN's. This dual-frequency system model has a useful feature that packets choose the same routes as in the single-frequency system.

Bandwidth Allocation

Bandwidth should be allocated to two tiers in proportion to the traffic load at each of the two tiers of the transmissions. The MAC throughput (or one-hop throughput in [56]) can be used as a measure of traffic load, and is defined as the total number of bits of data sent by all nodes per second, including forwarded bits. According to the definition, it can be obtained from the product of the end-to-end throughput and the average hop count of data packets delivered at destinations. Therefore, the bandwidth allocation depends on the ratio of average hop count at two tiers. Note that the dual-frequency system defined here chooses the same route for each packet as the single-frequency system. Therefore, when using the average hop count as a measure of traffic load, we can obtain the values of the average hop counts at two tiers of the dual-frequency system from the single-frequency system.

According to the routing assumption, the low-tier transmissions are always one hop. So the low-tier hop count is 1 for the Internet traffic and is 2 for the local traffic (see Figure 3.10). Suppose the measured average hop count of a single-frequency system is $\bar{h}$, then in the corresponding dual-frequency system, the average hop count at the high-tier is $\bar{h}-1$ for Internet traffic and $\bar{h}-2$ for local traffic. If the Internet traffic fraction is 80%, the average hop counts of the dual-frequency system at the low and high tiers can be estimated as $0.8+2*0.2=1.2$ and $\bar{h}-1.2$, respectively. Considering the single-frequency system scenario of the DSDV and 80% Internet traffic case in Section 3.3.2 as an example, we measure its average hop count, and estimate the average hop counts at two tiers of the corresponding dual-frequency system with results given in Table 3.3. We have verified that these estimated values are consistent with the measured values.

Table 3.3: Average hop count at two tiers

The number of AP's	1	2	3	4	9	12
High-tier average hop count	1.35	1.1	0.4	0.35	0.06	0.0
Low-tier average hop count	1.2					

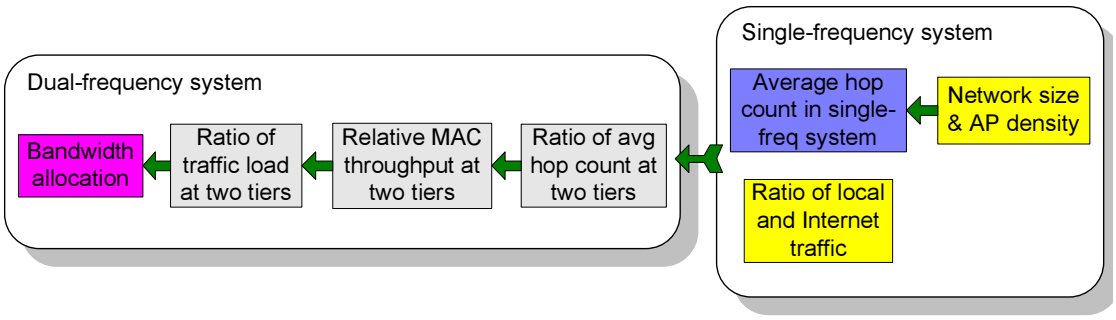


Figure 3.11: Bandwidth allocation determined by the network configuration and the traffic pattern.

Therefore, the bandwidth allocation can be determined by the average hop count of the single-frequency system and the ratio of the local and the Internet traffic, as summarized in Figure 3.11.

MAC Throughput

As a more direct illustration of traffic load, in Figure 3.12 we plot the MAC throughput at the low and high tiers as a function of the number of AP's in the network for the DSDV and 80% Internet traffic scenario used in Section 3.3.2. As expected, low-tier network throughput increases with the number of AP's due to increased direct MN-AP connections. We also observe the fact that network throughput increases rapidly with just a few AP's, and for the case with 4 AP's, the high-tier throughput contributes quite nicely to the overall network throughput which is only a few percentage points lower than the maximum system throughput obtained with 9 AP's. This justifies the use of FN's to replace wired AP's as motivated by the analytical results obtained earlier.

The average hop count as a function of the offered load is plotted in Figure 3.13. It is observed that the average packet hop count decreases when the number of AP's grows from 1 to 12. Also when the offered load gets heavier, the average packet hop count decreases. This is because that,

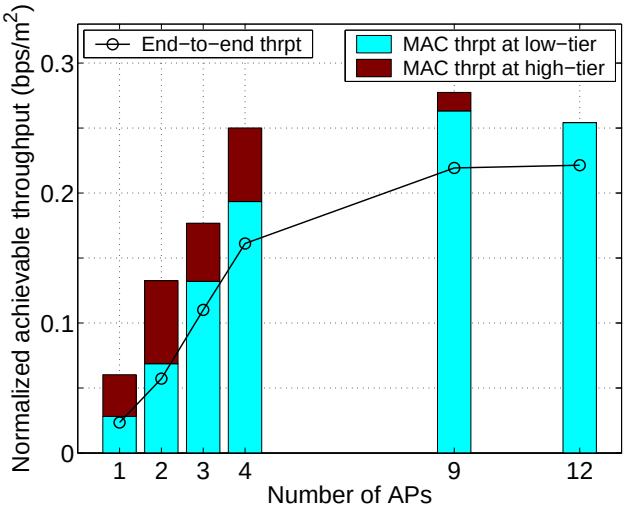


Figure 3.12: End-to-end throughput and MAC throughput. DSDV case, 80% Internet traffic fraction, $1200\text{m}\times1200\text{m}$ field.

out of the packets delivered, the fraction of single-hop packets increases. Less the average hop count, more packets the network can accommodate and larger end-to-end throughput the system can achieve. Therefore, it implies that most of the bandwidth is used to carry packets over short distances when the maximum throughput is achieved, resulting in short average hop counts in Table 3.3. This indicates some unfairness between traffic routed over different distances. In the case that most traffic has to go though the gateway AP's, it suggests that most of the bandwidth is used by the nodes in close proximity to the infrastructure nodes.

3.4.2 Results and Discussions

Based on the estimated ratios of average hop count at two tiers, we define several possible bandwidth allocation ratios between high and low tiers in Table 3.4. Single-frequency allocation to both high and low tiers, which may be considered as an imperfect type of dynamic allocation via the MAC protocol, is also considered for comparison purposes.

Figure 3.14 shows the achievable system throughput for each of the bandwidth ratios defined in Table 3.4 and also for the single frequency cases (for channel bandwidth 1 Mbps and 2 Mbps). The

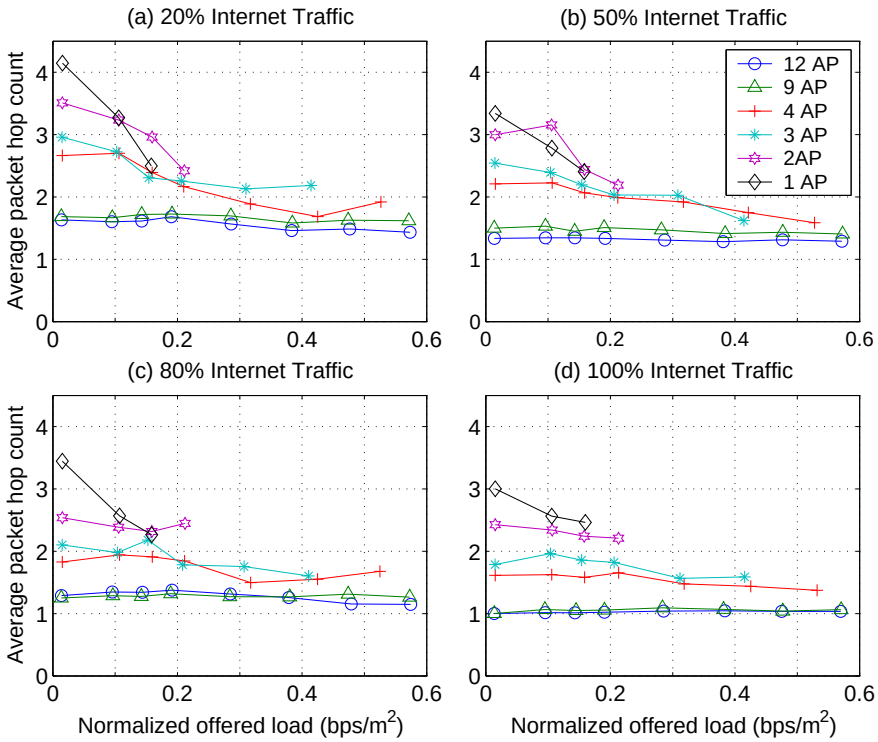


Figure 3.13: Average packet hop count. DSDV case, different Internet traffic fractions, 1200m$\times$1200m field, 36 FN's. (a) 20% Internet traffic fraction. (b) 50% Internet traffic fraction. (c) 80% Internet traffic fraction. (d) 100% Internet traffic fraction.

Table 3.4: Dual-frequency channel bandwidth allocation schemes

Bandwidth allocation scheme index	0	1	2
High-tier bandwidth (Mbps)	1.5	1	0.5
Low-tier bandwidth (Mbps)	0.5	1	1.5

detailed throughput vs. offered load curves are displayed in Figure 3.15. From the results shown, we first observe that the achievable throughput for the case with 1 AP and single frequency is limited by access to the wired network, and there is very little gain from doubling the available bandwidth from 1 to 2 Mbps. The throughput gain with increased bandwidth is better in the 4 and 9 AP single frequency cases, but still not very significant even though the channel bandwidth has been doubled. This can be attributed to the fact that 802.11 MAC involves significant and non-linear overhead for MAC layer control and routing (both of which use the basic rate 1 Mbps) as data rate increases, along with the fact that our sensor node traffic model uses relatively short 64-byte packets. This

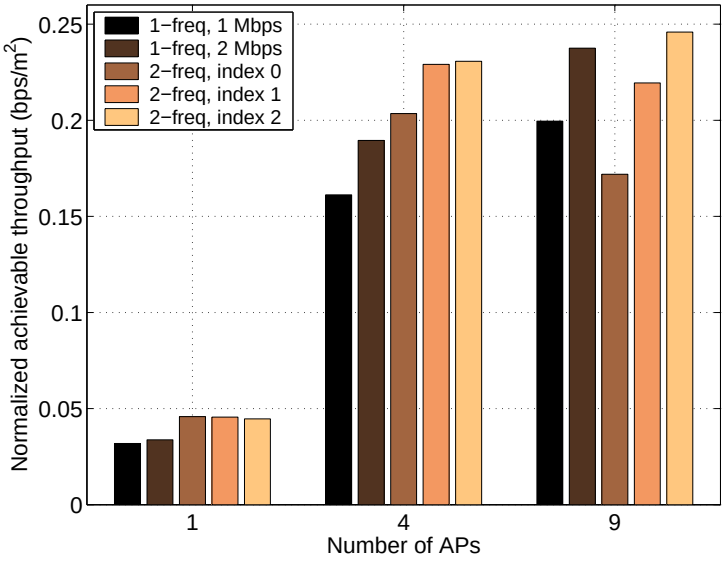


Figure 3.14: Impact of channel bandwidth allocation. DSDV case, 80% Internet traffic fraction, $1200\text{m}\times1200\text{m}$ field.

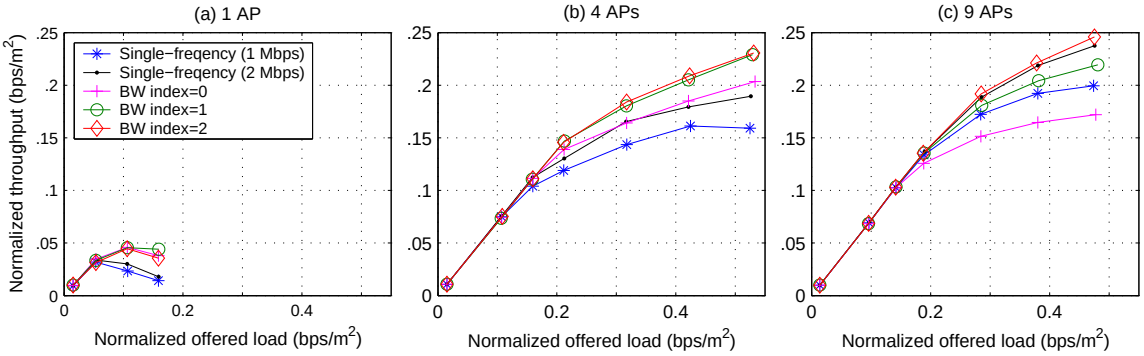


Figure 3.15: Throughput vs. offered load. DSDV case, 80% Internet traffic fraction, $1200\text{m}\times1200\text{m}$ field. (a) 1 AP. (b) 4 AP's. (c) 9 AP's.

motivates allocation of separate frequency bands to high and low tiers of the network to avoid the increased MAC layer overhead and scheduling inefficiency which arise in the single channel case.

Considering the two-frequency allocation cases shown in Figure 3.14, we observe that total throughput can be increased further provided the right ratio of high-tier and low-tier bandwidth is maintained. For the 1 AP case, as expected, the gains are small irrespective of single frequency or dual frequency assignment since the system is limited by the AP's capacity to handle traffic to the Internet. Figure 3.14 also shows that for the practically interesting (i.e. low cost, high performance)

case with 4 AP's, dual frequency assignment with the right bandwidth allocation (i.e. bandwidth allocation index 1 and 2) can provide significant gains (about 30%) over the baseline. On the other hand, when the number of AP's is large (i.e. the 9 AP case), system capacity does not benefit from dual frequency assignment since most of the traffic goes directly between MN's and AP's (i.e. is at the low-tier). We can observe that the throughput increases as the low-tier bandwidth increases. All of these verify our approach of bandwidth allocation based on the average hop count ratio described before.

The above results show that, in a hierarchical system with realistic MAC and routing protocols, the throughput of a single-frequency system does not increase linearly with bandwidth due to increased MAC/routing overhead. This motivates another mechanism for scaling ad hoc networks, by cross-layer adaptation, to reduce protocol overhead which has great impact on system scalability. We will focus on cross-layer adaptation in the next chapter.

Meanwhile, in a dual-frequency system, the highest throughput is obtained when the bandwidth is allocated according to the relative traffic load carried at two tiers. We have found that this load ratio can be computed from the average number of hops of the single frequency system as well as the ratio of local and Internet bound traffic. Since the average hop count of the single frequency system can be roughly estimated from network size and AP density, this approach (as demonstrated in Figure 3.11) can be used to allocate bandwidth to two tiers in a general and static way.

However, in a real system, the number of average hop count may be expected to change with variations in user density and traffic load. Thus, for more advanced and more accurate approaches, a hierarchical network with dual radios and separate bandwidth allocation for high and low tiers will need an adaptive algorithm to estimate parameters and allocate bandwidths accordingly. Specially, a network could be designed to have dynamic allocation of bandwidth between tiers, with control of channel allocation according to the resource utilization at different tiers. The algorithm

could be tied to inter-tier packet rates and relative queue lengths at the radio interfaces inside an FN. Note that in 802.11 networks, allocation of bandwidth will correspond to approximate partitioning of the channel space between high and low tiers with individual nodes selecting randomly from the assigned set of channels. Smooth allocation of the right amount of bandwidth to each tier of a hierarchical mesh network should become possible with future software-defined radios [57]. Additionally, multi-channel mesh, in which dedicated channels are assigned to wireless transmissions to avoid interference and collisions, is another scaling approach. With Orthogonal Frequency Division Multiplexing (OFDM) [58], variable bandwidth could provide enough sub-carriers for all the nodes in an ad hoc network to transmit in a non-interfering matter. We will propose a simple distributed channel assignment scheme based on two-hop edge coloring in Appendix B, that may be of independent interest.

3.5 Conclusions

In this chapter, we have investigated the scaling properties and system design principles for a hierarchical hybrid network via simulations with realistic MAC and routing protocols. The performance of an example three-tier 802.11-based network is evaluated and significant improvements in performance and system throughput relative to flat ad hoc networks are demonstrated. We have also studied the impact of the relative densities of AP's, FN's, and MN's as well as the traffic pattern on the achievable throughput and compared our results with the analysis. The simulation results demonstrate a well-defined saturation effect as the density of AP's is increased relative to a fixed density of FN's, thus leading to simple guidelines for determining the number of nodes at each tier of the hierarchy. Additionally, we have investigated the impact of the relative channel bandwidth allocated to two tiers on the achievable throughput of the dual-frequency hierarchical system. The results verify that dedicated allocation of frequency bands to each tier of the network can further

improve throughput provided that the bandwidth of high and low tiers is allocated according to the relative traffic load carried at two tiers.

Our experimental study has given results in confirmation with our analysis in Chapter 2, and shown that protocol overhead has great impact on system's scaling behavior. Our simulation results demonstrated that system throughput can be scaled by using the right proportions of FN's and AP's, making it possible to design high-capacity, low-cost mesh networks with a moderate number of radio FN's and only a few wired AP's.

Chapter 4

Scaling via Cross-Layer Adaptive Routing

This chapter investigates cross-layer adaptive routing as another mechanism to improve the system performance and scaling properties of ad hoc wireless networks. We propose an adaptive routing framework which allows introduction of adjustable parameters and programmable routing modules. In this framework, routing metric, routing algorithm parameters, and/or protocol selection can be controlled in response to observed performance and external service needs, and a global distributed policy manager is responsible for the adaptive operations at the nodes of the network.

The proposed architecture can support two types of adaptive mechanisms: the first involves switched selection between a set of routing protocols options or metrics, and the second is based on an integrated routing algorithm which incorporates adaptation of key network state parameters such as link speed or congestion. As integrated adaptive routing algorithms have the advantages of reduced adaptation overhead and flexibility, we further investigate this kind of algorithm, taking a cross-layer routing metric as an example.

The cross-layer routing metric is also motivated by the fact that cross-layer adaptation facilitates global optimization across different layers, and reduces protocol overhead which has strong impact on ad hoc network system scalability. The proposed PARMA (Phy/mac Aware Routing Metric for Ad hoc networks) routing metric incorporates state variables including physical layer link speed and MAC congestion. The PARMA routing metric is implemented with a proactive ad hoc routing protocol in the *ns*-2 simulator. The simulation results show that PARMA helps increase network

throughput and reduce network congestion by selecting paths with high bit-rate links while avoiding areas of MAC congestion.

4.1 Policy-Based Adaptive Routing Framework

4.1.1 Introduction

In mobile ad hoc networks (MANET's), the topology may change frequently and unpredictably due to node movement and fluctuating wireless link quality. These characteristics make the development of dynamic routing protocols with good bandwidth and power efficiency an important design challenge. While there are many results on specific classes of ad hoc routing protocols such as DSR, DSDV, AODV, or Temporally-Ordered Routing Algorithm (TORA) [59], no single routing protocol performs well in a complex real-world environment. For example, previous work [60] shows that DSDV, as a proactive routing protocol, is preferable for latency-sensitive traffic; but DSR, as an on-demand routing protocol, outperforms DSDV in high mobility environment.

The above considerations motivate the use of adaptive routing for ad hoc networks such that routing can dynamically adapt to changing network topology and external service needs. One approach is to combine proactive and reactive strategies. For example, the Sharp Hybrid Adaptive Routing Protocol (SHARP) [61] is a hybrid routing protocol that automatically finds the balance point between proactive dissemination and reactive discovery of routing information such that it can dynamically adapt to changing network characteristics and traffic behavior. Also a strategy presented in [62] dynamically combines table-driven and on-demand routing and is adaptive to node mobility. Tuning routing algorithm parameters can also help achieve adaptive behavior. The adaptive zone routing protocol (AZRP) [63] is an example which uses variable zone radius and controllable route update interval. Some channel adaptive routing schemes suggest choosing routes with high date rates [23, 64].

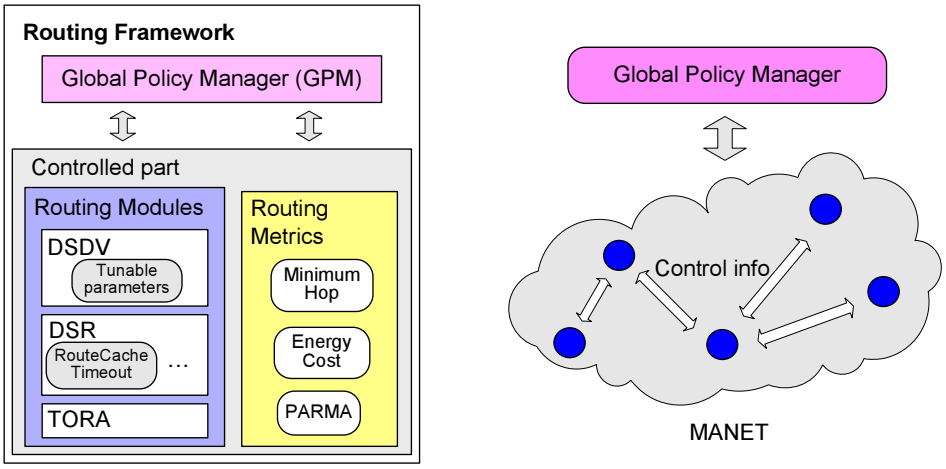


Figure 4.1: (a) Adaptive routing framework. (b) Distributed global policy manager.

Most prior work on ad hoc routing is specific to a certain routing protocol or approach. In this work, we focus on system design and propose a unified adaptive routing framework, in which various adaptive mechanisms can be deployed. We also investigate the use of adaptive routing policies in ad hoc networks to achieve improved routing performance over conventional methods. Example adaptive algorithms, in particular a cross-layer adaptive routing metric and its implementation and performance, are presented.

4.1.2 Adaptive Routing Framework

The architecture of the adaptive routing framework is shown in Figure 4.1(a). The architecture implements self-adaptation by a control loop, which collects the information about routing states from the system, makes decisions, and adjusts the system as necessary. The control loop consists of two parts: the controlling part and the controlled part. The Global Policy Manager (GPM) is the controlling part, which implements particular adaptation operations such as selecting the routing module, tuning the routing algorithm parameters, and adjusting the routing metric variables. The routing modules and routing metrics are the controlled elements.

Several routing modules are available in the framework. The routing module which would

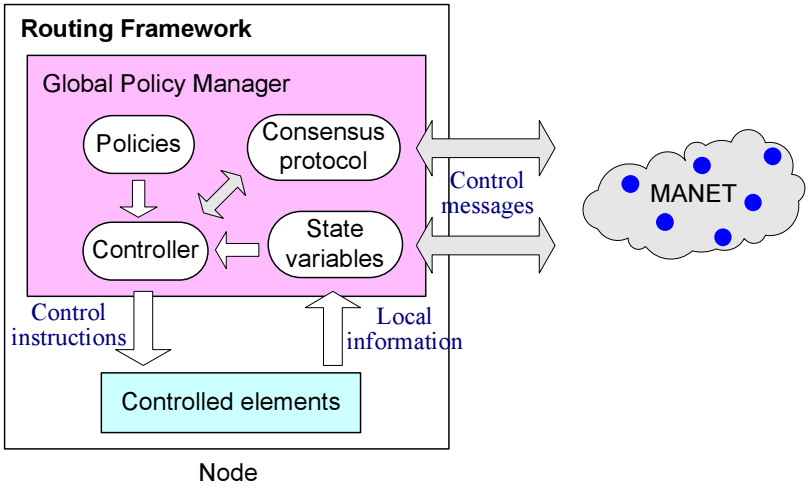


Figure 4.2: The Global Policy Manager (GPM) at a node.

produce the best desired performance is selected by the GPM. The parameters of the selected routing module can also be tuned by the GPM. According to service requirements and traffic behavior, the GPM decides a routing metric and its variables.

In order to achieve global optimization, the GPM, when making decisions, needs not only local information but also information from other nodes of the network. The control information, including state variables and management information, is disseminated through the network. Thus the GPM entities of all the nodes in the network construct a distributed system and perform the global controlling functionalities, instead of being isolated, as shown in Figure 4.1(b).

4.1.3 Global Policy Manager (GPM)

Achieving global optimization in a distributed system is a complicated goal, especially when the system conditions are under changes. As shown in Figure 4.2, in the GPM of each node, the controller monitors local behavior (e.g. channel utilization around the node), interacts with information from other nodes (via control messages), analyzes data based on routing policies, and decides control actions. Control instructions produced by the controller are sent to the controlled elements to complete adaptation functions.

The control information needed to form the global information could be sent either explicitly or implicitly. If it is sent explicitly, extra control messages are necessary. If a consensus protocol is used to achieve consensus over all the nodes in the network on how to adjust their settings to globally adapt to the network conditions, it can provide a way to transfer control information. If the control information is sent implicitly, it can be piggybacked on the existing routing messages to reduce adaptation overhead.

Routing Policy

A routing policy specifies the criteria that the GPM uses to accomplish a particular adaptation operation, designed as an algorithm to analyze data (e.g. routing state variables and service needs) and produce control processes.

Routing State Variable

The routing state variables can be categorized as follows:

- *Traffic behavior* which is dynamically changed and can be described by wireless medium access delay and/or packet dropped rate. Medium access delay (or channel busy degree) can be estimated by measuring the occupancy of the channel [23]. Packet dropped rate is the fraction of packets dropped due to full queues at nodes.

- *Link characteristics* which can be represented by link quality and link data rate. Link quality can be measured as the expected number of transmissions using probes [65]. Link data rate can be obtained from the link layer [23].

- *Mobility metrics* which evaluate the relative difficulty of routing due to mobility in ad hoc networks and is related to many factors such as the mobility pattern, movement speed, and

thus is not easy to evaluate. The example mobility metrics are *Geometric Mobility Metric* and *Minimal Route-change Metric* [66].

- *Service needs* which can be obtained from the application profile, for example, latency-sensitive and/or burst traffic.

- *Other known network conditions* which could be network size, stationary or moving nodes, power sensitive requirement etc.

When the state variables are from layers other than the network layer, cross-layer adaptation can be supported by the routing framework.

4.1.4 Adaptive Mechanisms

Corresponding to different controlled elements, the proposed architecture can support three adaptive mechanisms: (1) selecting routing module; (2) tuning routing algorithm parameters; and (3) adjusting routing metric. These mechanisms are based on specified adaptive routing policies, and can be used either individually or together with any others. In this section, after describing each mechanism, we review some existing adaptive techniques and discuss how they can be supported by the proposed adaptive routing framework.

Selecting Routing Module

When the routing service is initialized, the routing module is selected by the GPM according to the application requirements (e.g. latency sensitive or real-time service) and the observed network conditions. The routing module may be re-selected accordingly if the environmental conditions are changed during run time; this is dynamic selection. In order to accomplish routing module selection, an additional distributed mechanism is required to reach a consensus.

Switching routing module requires resetting and reconfiguring the routing service and may cause service discontinuity, so it is preferred for long-time changes. Another approach is to use a hybrid routing protocol combining several routing strategies into one protocol and dynamically balancing the tradeoff among these parameters corresponding to the environment. The SHARP protocol is such an example [61].

Tuning Routing Algorithm Parameters

Tuning of key routing algorithm parameters is another step towards designing an adaptive system. When the optimal values of the parameters are related to the network conditions under changes such as node movement speeds, it is preferred that these parameters can be dynamically adjusted accordingly, rather than fixed to the predetermined values.

An example parameter is the *cache timeout* of DSR [67]. In DSR, the routing information stored in the cache can be used to avoid route rediscovery for each individual packet. However, it may be out-of-date information. To overcome this potential drawback, each link in the cache has a timeout associated with it, to allow that link to be deleted if not used within this timeout. The simulation studies in [66] show that this timeout value is closely related to the performance metrics, and depends on complex factors including the mobility scenario, the node movement speed, the contents of the cache, and the routing protocol's reaction on it. Taken these factors as the state variables, the cache timeout parameter can be adjusted in the proposed routing framework according to the routing policy which is designed as a cache timeout algorithm.

Adjusting Routing Metric

The routing metric consists of path selection criteria . The minimum hop routing metric, which is a carry-over from the wired network, might not be suitable in wireless environments when bandwidth, medium congestion or power usage are concerned. In wireless networks, different routing metrics

can be used to trade off among system performance metrics such as throughput, latency, and power. For example, when power is considered, the energy cost [68] can be used as the link metric to reduce power consumption at nodes by trading off against the throughput. Therefore, according to the requirement of power, the GPM can decide the routing metric for the routing module to be used.

The routing metric can also be switched between available metrics when service needs change. Like re-selecting the routing module, the metric switch requires explicit control messages to reach consensus and may cause service discontinuity. Adaptive routing metrics represent a different approach, which can provide adaptive features via routing metric variables that adjust to the dynamically changing environment and traffic load. In particular, cross-layer adaptive routing metrics incorporate parameters which reflect the observed network state changes (e.g. the link speed and MAC congestion) such that the routes which would produce the best system performance are chosen.

4.1.5 Example Algorithms

According to the implementation of adaptation operations, the adaptive mechanisms can be classified into two types: one is switching between available routing modules or routing metrics, and the other is implementing an integrated adaptive algorithm to control a particular routing element, such as a routing metric or a routing algorithm parameter. For each of these two types, we give examples of routing policies and adaptive algorithms that can be deployed in the routing framework.

Switching Routing Module

As an on-demand routing protocol, TORA can quickly create and maintain loop-free multipath routing for packets, while reducing routing overhead [59]. However, it is shown in [60] that TORA would fail to converge because of congestion collapse when the number of communication pairs

Table 4.1: Algorithm for switching routing module

1	Check *channel busy degree* ρ during run time
2	**if** $\rho \geq 0.6$ && TORA **then**
3	Trigger the Consensus Protocol for switching to DSDV
4	**else if** $\rho \leq 0.4$ && DSDV **then**
5	Trigger the Consensus Protocol for switching to TORA

increases, while DSDV has approximately consistent performance regardless of the number of communication pairs. Based on these characteristics, we could switch the routing protocol from TORA to DSDV when the number of communication pairs increases. We assume that the offered traffic load at each source node does not change over time, then MAC congestion can be used as a measure as the increased number of communication pairs.

In particular, the channel utilization factor (or *channel busy degree*, refer to Section 4.2.3), ρ, is used as the state variable, and can be estimated at each node by sensing the occupancy of the channel. In Section 4.2.3, Figure 4.7 shows that the network begins to saturate when the channel busy degree approaches 0.6. Since switching routing module is worth the overhead it brings only when the change is long-term, we need to observe the channel busy degree for some period of time. When we measure $\rho \geq 0.6$ for some period of time, the routing module is switched to DSDV; when $\rho \leq 0.4$, it switches back to TORA. Based on this policy, the adaptive algorithm is given in Table 4.1, which can be used to trigger a transition from one protocol to another.

Integrated Routing Metric

Integrated adaptive algorithms provide an approach in which the controlled and controlling parts of the self-adaptation function are implemented together in a distributed algorithm. The control information including time-varying state variables can be propagated over the network by routing

messages, and each node makes decision based on its local information without a consensus protocol. Therefore, the integrated adaptive algorithms are practically flexible to implement with reduced adaptation overhead and do not involve service interruptions.

Integrated adaptive algorithms could be adaptive routing protocols (e.g. SHARP), adaptive routing parameters (e.g. the cache timeout of DSR) or adaptive routing metrics (e.g. PARMA). We will focus on the PARMA routing metric as a cross-layer adaptive algorithm that could be applied to the routing framework.

4.2 PARMA: A PHY/MAC Aware Routing Metric for Ad Hoc Wireless Networks with Multi-Rate Radios

4.2.1 Introduction

Most conventional ad hoc routing protocols, including DSDV, AODV, and DSR, use the minimum hop (MH) as the metric to make routing decisions. This is primarily a carry-over from wired networks where the transmission rate of a link does not dynamically change and the link rate is independent of the physical transmission range. However, in case of wireless networks, the MH metric tends to choose paths with fewer hops. And each hop in paths chosen by MH will tend to have a longer physical span and also be associated with a lower bit rate than an alternative path with more hops. In order to take advantage of the wireless multi-rate capability and make better use of available network capacity, it is clear that transmission rate needs to be incorporated into the routing metric.

In addition to the transmission rate, we observe that it is also possible to provide an awareness of congestion at each node in order to avoid bottleneck regions with high link utilizations. The wireless link is usually shared with other links in the same neighborhood, while in a wired network, links operate independently of each other and channel access on one link has no effect on any of the

adjacent links. Thus, it makes sense to devise metrics that account for both congestion and rate in a combined manner. For example, a link may provide for a high transmission rate, but could appear congested because neighboring links have a high link utilization. Thus, if we account for only the link rate, this link would show up as a "good" link but when combined with a congestion metric, it may turn out to be just the reverse, which is a more accurate reflection of the PHY/MAC layer.

The above considerations motivate cross-layer adaptation for better scaling, which is based on a PHY/MAC aware routing metric and optimizes PHY, MAC, and routing jointly. In particular, this integrated PHY/MAC aware routing metric is related to both the PHY bit-rate and MAC congestion information. Taking DSDV as the routing protocol baseline, we study the distance vector routing behavior under a multi-rate PHY, and with the PARMA metric. The problem is studied in detail with an *ns*-2 simulation model [52].

Related Work

A routing algorithm incorporating multi-rate PHY is investigated in [69], which proposes the Medium Time Metric (MTM) to find paths such that the total transmission time is minimized. MTM is a static solution which only handles the multi-rate capability and leaves out the wireless medium access contention factor. A "full interference" assumption is used in its theoretical model.

De Couto *et al.* observe that using the shortest path would result in poor throughput [70]. They propose the expected transmission count metric (ETX) to incorporate the effects of link loss ratios [71]. ETX introduces extra routing overhead of the dedicated link probe packets to measure the delivery ratio. Also ETX is independent of network load and does not attempt to route around congested links.

Incorporating both the link loss rate and the link speed, the Expected Transmission Time (ETT) is used as the weight associated to each link [72]. The individual link weights are combined into a

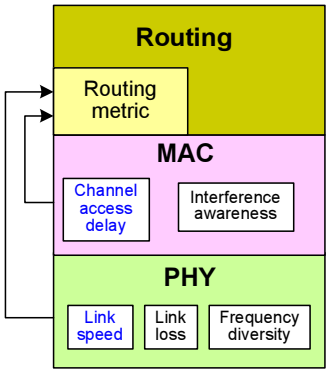


Figure 4.3: PHY and MAC information incorporated in cross-layer routing metric.

path metric called Weighted Cumulative ETT (WCETT) that explicitly accounts for the interference among links that use the same channel which tends to choose channel-diverse paths. The WCETT metric is proposed for multi-radio multi-hop wireless networks with stationary nodes, aiming to provide high-throughput pathes for packet transfer.

The airtime link metric is specified in the IEEE 802.11s draft [6]. Note that after taken away the constant factors such as the channel access overhead, the protocol overhead, and the packet size of test frame, this metric leaves the link bit rate and the link error rate. Therefore, it is in fact equivalent to ETT.

In [73], a path weight function, called Metric of Interference and Channel-switching (MIC), is proposed to provide interference-aware and multi-channel/multi-radio aware load balancing for mesh networks. In [74], the MAC layer utilization, which is averaged over a 10-second period, is used to adjust the behavior rather than metric of routing protocols.

The state variables collected from the PHY and MAC layers which can be incorporated into the adaptive routing metric are demonstrated in Figure 4.3. In this work, we incorporate both the PHY link speed and the MAC congestion, where the latter is not considered in any of the cross-layer routing metrics discussed in this section. The channel access delay is estimated via short-term measurements, and used to avoid the busy area without higher layer congestion avoidance schemes.

4.2.2 Rate-Adaptive PHY

The widely used IEEE 802.11x standard uses adaptive selection of physical layer bit rate as a function of observed channel quality. 802.11b radios can choose different physical rate (1, 2, 5.5, 11 Mbps) while 802.11a/g radios select between 6, 9, 12, 18, 24, 36, 48, or 54 Mbps as the physical channel rate. This automatic PHY bit-rate adaptation feature is considered to be useful in most systems because it permits end-users to take advantage of good-quality short-range links when available. When such multi-rate radios are used to build ad hoc networks, the network topology and link speed change more dynamically than in radio networks with a single mode radio with fixed bit rate and range.

Figure 4.4 depicts the way in which an 802.11b radio device experiences different bit rates when connecting to its neighbors at various distances. As shown in the figure, if a node wants to use rate 11 Mbps, only nodes in the inner-most circle can decode its frame correctly with sufficient Signal-to-Noise Ratio (SNR). However, if it chooses to use the lower 1 Mbps rate, the transmission range would be much larger. The outer-most circle indicates the threshold of carrier sense in 802.11 MAC. If there is a radio outside this circle, then the signal level of this radio's transmission received at the central node is not large enough so that the central node would sense the channel as "idle". Note that in the above description, a two-ray path loss channel model [75] is assumed and the received signal strength is simply compared to a series of fixed thresholds.

Networks may benefit from connections with multiple short-range, high-speed links relative to a single low bit-rate hop that spans a longer distance. In Figure 4.5, 10 stationary nodes are placed in a straight line. Assume each node can reach its immediate neighbor with a fast 11 Mbps link but can only reach the node next to the immediate neighbor with a 1 Mbps link. Therefore, for packet transfer from node 3 to node 7, node 3 can either choose a 4-hop route 3-4-5-6-7 with 11 Mbps rate used for each hop, or a 2-hop route 3-5-7 with 1 Mbps each hop. This diversity of route selection will

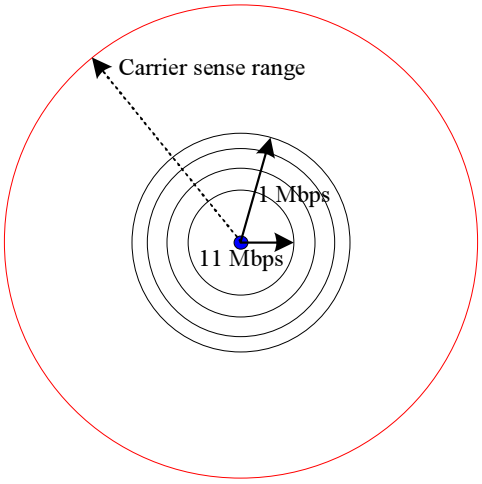


Figure 4.4: Transmission range of a multi-rate radio.

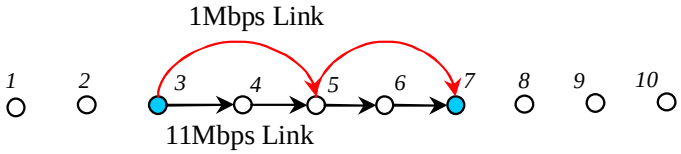


Figure 4.5: Chain topology and different links.

not occur in single-rate networks. With the simulation in Section 4.2.5, it will be demonstrated that short high-speed links are actually better than long-range slow links under certain circumstances.

Auto Multi-Rate Mechanisms

The IEEE 802.11 standard does not specify how to choose PHY rate based on varying channel conditions, leaving this as an implementation detail. Previous studies have proposed some schemes to select rate adaptively. For instance, with the Auto Rate Feedback (ARF) scheme [30], a node simply steps up to a new rate whenever 10 consecutive transmissions succeed or no failure occurs during a time interval, and steps down to a lower rate if one or more ACK's are missing. But the ARF scheme only works when a sender communicates with one single receiver. Otherwise, the ARF fallback would make wrong judgments of rates to different receivers. The Receiver Based Auto Rate (RBAR) scheme is proposed to solve this problem [31]. Basically, a node first uses the lowest rate to send an RTS frame to a receiver. Based on the measured signal strength, the receiver chooses

an appropriate rate for the sender and piggybacks this information to the sender in a CTS frame. Then the sender sends the DATA frame with the selected bit-rate. The RBAR scheme overcomes the problem of ARF, but requires an RTS/CTS exchange before every DATA transmission.

When constructing an ad hoc network with muti-rate 802.11x or other similar radios, PHY rate-adaptation can be applied on a packet-by-packet basis for each communicating neighbor. This can be implemented with a table at the MAC layer for recording the selected physical rate to each neighbor, and the rate is based on the SNR measurements of the received packets. For each outgoing packet, the device driver looks up the table for the next hop and obtains a suitable rate for it. Note that continuous transmissions are required for maintaining up-to-date SNR measurements in the rate-neighbor table. For the purpose of our research, we use this SNR-based auto-rate scheme and implement it in the *ns*-2 model.

4.2.3 MAC Channel Congestion

The MAC channel congestion can be measured by the channel access delay. The channel access delay correlates to the traffic at the MAC layer taking into account both the locally offered traffic and that forwarded by neighboring nodes. Because the wireless medium is shared, whether a packet can access the channel immediately is determined by not only the states of the two end nodes of the link, but also the states of all neighboring nodes.

To measure this effect, a “virtual access delay” estimation based on physical layer information is introduced. In order to avoid unnecessary overhead introduced by periodic probes, we propose a passive estimation method. Every node records every channel event (i.e. transmission) sensed from the physical channel and makes an estimation of the “expected delay if a packet has to be sent”. Suppose an M/M/1 queuing system [27] with the common channel as the server, where packets arrive from the nodes in the neighborhood of the channel to obtain service (i.e. access the channel

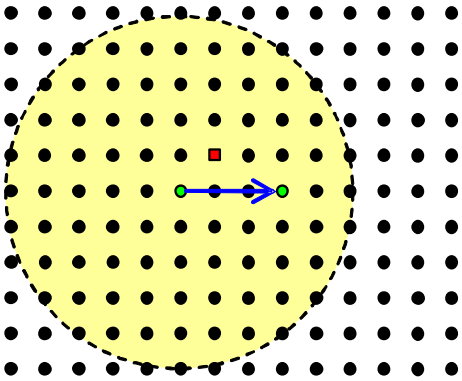


Figure 4.6: A 14 by 11 grid with a 3-hop flow.

and get transmitted). According to the results of queueing theory, the average waiting time in queue, i.e., the *channel access delay*, is given by

$$T_W = T_S \frac{\rho}{1-\rho}, \tag{4.1}$$

where ρ represents the utilization factor of server (i.e. the *channel busy degree*), and T_S is the service time for the channel event, corresponding to the packet transmission time. Each node can estimate ρ by sensing the occupancy of the channel. The estimated T_W is given to the routing protocol for use in the routing metric.

To evaluate the above channel access delay estimation method, we use a grid topology having 154 nodes with spacing of 350 meters between adjacent nodes, as shown in Figure 4.6. The carrier sense range is 1783 meters (slightly greater than a distance of 5 nodes). We have a 3-hop constant bit rate (CBR) flow running in the center of the grid. The bit rate used for each hop is 11 Mbps, and the corresponding transmission range is 399 meters. The channel busy degree ρ when simulated with saturated load is shown in Figure 4.7. Obviously, the congested area is much larger than the specific region through which the flow passes in this example. Nodes with highest busy degrees actually lose the ability to support any flows further. Note that the maximum channel busy degree shown in the figure is only around 0.77.

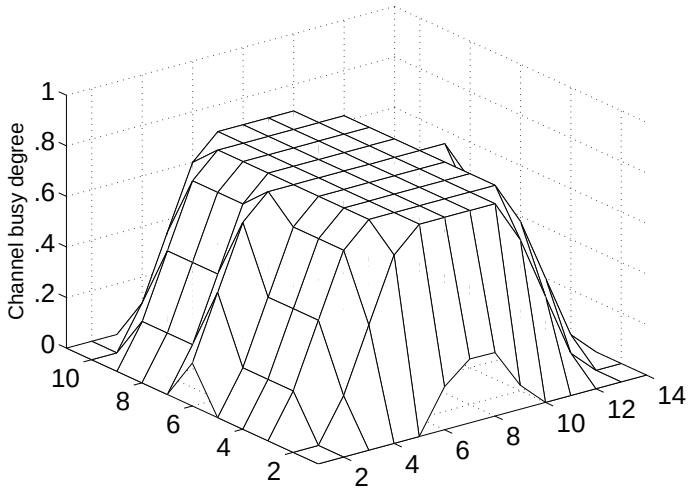


Figure 4.7: Channel busy degree across the grid.

We also plot the channel access delay, estimated at one specific node (shown as the square-shaped node around the center of Figure 4.6), as a function of the offered load. Figure 4.8 shows that the estimated access delay increases monotonically when the offered load is increased. However when the network is congested, the access delay estimate does not increase any further but holds at a steady level. This is because that queuing effects cannot be monitored by the physical layer. Although the queuing delay will increase dramatically if the congestion is not eased, it cannot be estimated from the channel access delay alone. We also compare the access delay estimates with the corresponding measurements and find that the estimates are very close to the measurements[1]. All these verify that the channel access delay obtained by the above estimation method represents the expected waiting time for a packet to get transmitted at the MAC layer, and thus can be used as a state variable from the MAC layer.

[1]A minor difference is that the access delay estimate is usually smaller than the actual measurements. There are some additional delays introduced by the IEEE 802.11 MAC, such as SIFS, DIFS, and backoff intervals. As those delays cannot be monitored, the delay estimate is expected to be an underestimate using the proposed method.

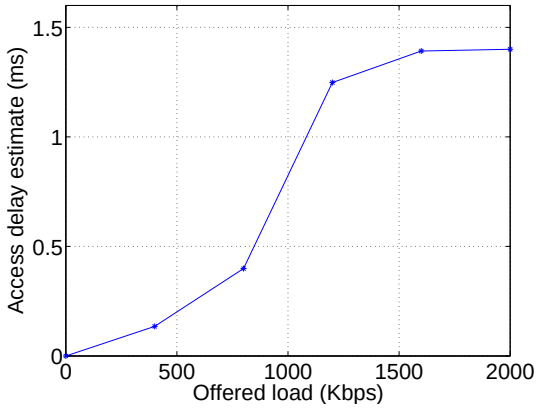


Figure 4.8: Channel access delay estimated at the square-shaped node in the grid.

4.2.4 PHY/MAC Aware Route Selection

Routing Metric

Previous work has showed that in an ad hoc wireless network, minimizing the number of hops would choose routes with a small number of links with relatively long physical span and hence lower bit-rate and worse link quality in terms of packet error rate [69, 70]. This motivates investigation of routing metrics which take into account the PHY layer bit-rate and other MAC layer information.

We propose a routing metric which aims to optimize the packet end-to-end delay. The end-to-end delay of a packet of size L_{pkt} transversing a path p_i is calculated as

$$Delay_{\ p_i} = \sum_{\forall links \in p_i} (T_{transmit} + T_{access} + T_{queuing}), \tag{4.2}$$

where $T_{transmit}$ denotes the packet transmission time in the link, T_{access} the medium access time spent by the packet getting access to the link, and $T_{queuing}$ the queuing time required for the packet waiting before trying to access the channel.

The packet transmission time can be calculated as

$$T_{transmit} = N_{transmit} \times \frac{L_{pkt}}{R_s}, \tag{4.3}$$

where R_s is the link speed, which would be one of the rates the multi-rate devices provide, and $N_{transmit}$ is the number of transmissions, including retransmissions, needed for the packet to be

received correctly. When the link quality is poor, packet retransmissions will be carried out by the MAC protocol. With adaptive multi-rate PHY, around 90% of packets get transmitted successfully in the first attempt [76]. Hence $N_{transmit}$ can be set to 1 as an approximation.

The medium access time, T_{access}, is used to indicate the medium busy level around the sending node of the link. When the medium is busy, it takes a relatively long time for a packet to get the chance to transmit. Incorporating the medium access time to the routing metric, the routing algorithm can choose a route with light traffic load in addition to high speed links, thus spread the traffic over the network to achieve load balance, avoid congestion, and increase effective bandwidth. Note that T_W, the access delay estimated by the MAC layer, can not be directly used as T_{access}. This will be discussed later in Section 4.2.4.

A large access delay reflects a growing interface queue length when the network is congested. When a system below saturation is considered, $T_{queuing}$ can be omitted[2].

With the above assumptions and simplifications, the routing metric computation can be summarized as

$$Delay_{\,p_i} = \sum_{\forall links \in p_i} \left(\frac{L_{pkt}}{R_s} + T_{access} \right). \tag{4.4}$$

It is clear that this routing metric is both PHY rate-aware and MAC load-aware. In the following, we study this kind of routing metric, in conjunction with the class of distance vector routing algorithms, taking DSDV as a specific example. Our study also reveals potential problems in the cross-layer design of ad hoc wireless networks.

Implementation

Our rationale for choosing DSDV as the routing protocol to study the proposed PHY/MAC aware routing metric is as follows. The periodic routing updates in distance vector protocols can exchange

[2]Through the experiments we verify that the route selection is not affected if the queuing delay is taken into account in the routing metric.

PHY/MAC information via the metric, so that the dynamic network conditions can be known over the network in a timely fashion. Thus packets can switch their routes whenever better routes are available. For on-demand routing protocols, extra control messages have to be added in order to initiate route discovery procedures to find a better route before the current one is broken. Moreover, with distance vector routing, the SNR-based auto-rate scheme is facilitated by adjusting the physical rates stored in the rate-neighbor table at the MAC layer according to the current SNR measurements of the received periodic updates. The channel busy degree can also be measured through continuously transmitting DSDV control messages, if necessary. With the above implementation, extra routing overhead, for instance, that introduced by using probe packets to make measurements [71], is minimized.

(1) DSDV Operation

DSDV uses a sequence number (SN) which is originated by the destination to indicate the freshness of the routing information and prevent routing loops. In addition, each entry in the routing table is associated with the weighted average settling time, which is the length of time between the arrivals of the first route and the best route to a particular destination with the same SN. Since the routing information broadcasts are asynchronous, some fluctuation of routing updates can occur. To solve this problem, and also reduce the number of rebroadcasts, advertisement of routes is delayed until the route has stabilized, i.e., twice the average settling time has passed since the first route is received.

With these two techniques implemented, the best routes are easy to achieve when the MH metric is used. But problems arise when the best route is not the shortest one. This is because, in DSDV, the time for a route from the destination to reach the other end depends on the settling time at each intermediate node in the path. The more the number of hops transversed, the greater the total settling time required. If the best route is not the shortest one, it is likely to arrive late, and even worse, it may arrive later than some routes with a new sequence number. Without the correct settling time, it

is difficult to get the best routes.

(2) Enhancement to Achieve Correct Settling Times

As discussed before, DSDV uses the delay-advertisement approach. Another approach, called the delay-use, is proposed in [71]. In the delay-use approach, two routing tables are used at each node. A route is not used until it is allowed to be advertised. Before the route can be used and advertised, if there is a route to the same destination with the new sequence number arrived, the old route is moved to the second table and the new route is stored in the current table. We use this modification in our implementation. However, still, the correct settling time for each entry of the routing tables is critical for both the delay-advertisement and the delay-use approaches.

We propose an approach to quickly adjust and achieve the correct settling time. In our approach, the received routes with the last old sequence number are handled instead of being ignored as in the current protocol. This avoids missing the best route when it arrives later than the first route of the next new sequence number. In particular, a route with the last old sequence number is chosen if it has a better metric than the one stored in the second table. Meanwhile, the average settling time of the route to the same destination in the current table is updated accordingly. With this enhancement, the settling time converges quickly and the best routes can be guaranteed before twice the settling time has passed. Our simulations show that 99% of the routes are the best ones with this enhancement, while only one third are the best without this enhancement.

We note that the enhanced DSDV with two routing tables works well in small and medium networks. When the network becomes large and the variations of different route arrival times increase, the overlap between routes to the same destination but with different sequence numbers will increase. More routing tables are required to store the routing states and prevent missing the best routes. In this sense, DSDV suffers from a scalability limitation when used with PHY/MAC aware metrics.

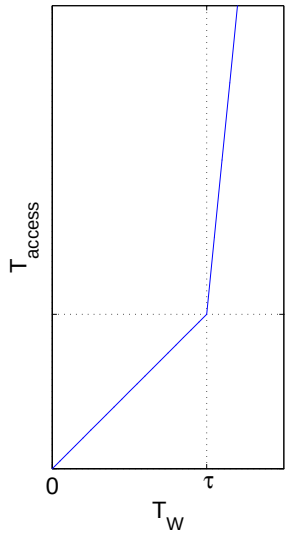


Figure 4.9: A non-linear mapping from T_W to T_{access}.

(3) Enhancement to Smooth Link Layer Changes

Because of the contention access nature of 802.11 MAC, the channel access delay is a random variable. In order to accurately reflect the channel busy degree, a smoothing window is used to get the average value over a time duration. Additionally, observing the non-linear behavior in Figure 4.8, we propose a non-linear mapping between the estimated T_W and T_{access} in the routing metric. In particular, a time threshold τ is defined according to the network condition. If the averaged access delay estimate is less than τ, it is used directly for T_{access}; if it is greater than τ, it is enlarged (e.g., ten times of its value) and then used for T_{access}. If τ is chosen appropriately, the area with high busy degree can be prevented from being chosen to support more traffic and congested. Figure 4.9 plots an example of non-linear continuous mapping from T_W to T_{access}, where the gradient is 1 before the threshold and is 10 after the threshold.

Moreover, due to the different time scales of the network and PHY/MAC layer variations, we must be careful not to degrade routing performance when incorporating the PHY/MAC aware routing metric. Note that if routing updates are too frequent, routing overhead is large and even worse, route convergence might be hard to achieve; but if routing updates are too slow, route exchanges

might not trace link changes. We propose that only significant changes, such as a delay having 20% increase, trigger routing updates; non-significant changes are advertised by periodic updates. Thus, the number of triggered updates are reduced while significant link changes can be advertised over the network. Such a cautious approach is required for this type of cross-layer design [42] in order to balance factors such as settling time, overhead, and routing performance.

4.2.5 Simulation Results

The system performance with the proposed PARMA metric is compared with the MH and MTM [69] metrics using the *ns*-2 network simulator [52]. The performance is evaluated in terms of system throughput, packet delivery ratio, and average end-to-end delay.

Simulation Parameters

In our simulations with DSDV, the time period between two consecutive periodic updates is 15 seconds, the minimum time between two consecutive triggered updates is 1 second. An update must be heard from each neighbor in every 45 seconds, otherwise the neighbor is regarded as unreachable.

Our simulation study considers constant bit rate (CBR) as the traffic generation model [52]. Packets have a constant size of 512 bytes and are sent at a deterministic rate. The sending rate is varied as an input parameter to gradually increase the offered load to the network.

Multi-rate 802.11b is used with four rates: 1, 2, 5.5, and 11 Mbps. The transmission power is fixed at 15 dbm. RTS/CTS is disabled[3]. ACK's are transmitted using the basic rate of 1 Mbps. The receiver thresholds and the corresponding effective distances are shown in Table 4.2, which are obtained by using the two-ray ground reflection propagation model [52].

[3]The experiment results with RTS/CTS enabled are similar to those with RTS/CTS disabled, though with slightly lower improvements. This is because the basic rate is used for sending the RTS and CTS frames, and this would increase the MAC overhead, especially for high rate links.

Table 4.2: Receiver thresholds and distances of multi-rate radio

	Receiver threshold	Distance
Carrier sense	-108 dBm	1783 meters
1 Mbps rate	-94 dBm	796 meters
2 Mbps rate	-91 dBm	669 meters
5.5 Mbps rate	-87 dBm	532 meters
11 Mbps rate	-82 dBm	399 meters

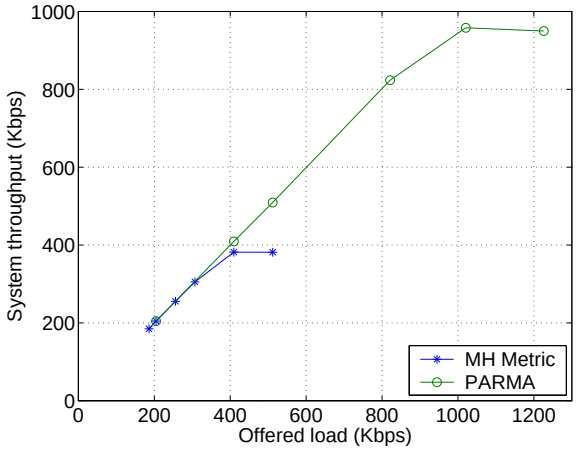


Figure 4.10: Throughput vs. offered load. Scenario I.

Scenario I: Chain Topology

First we try a simple topology with 10 stationary nodes placed in a straight line with a distance of 350 meters between the neighboring nodes, as shown in Figure 4.5. In this linear scenario, a short link with distance of 350 meters can provide 11 Mbps data rate, while a longer link with distance of 700 meters only gives 1 Mbps. There is one flow, originated from one node to the destination which is 1400 meters away. As observed, the MH metric tends to choose the shortest path with two 1 Mbps links, and the PARMA metric chooses the route with four 11 Mbps links. The simulation results are shown in Figure 4.10 and Figure 4.11.

The results show that in the chain topology, system performance improves with the PARMA metric. In this scenario, there is a significant factor of 2.5 times improvement in system throughput. Also, the packet delivery ratio and the average end-to-end delay are improved. In this scenario with

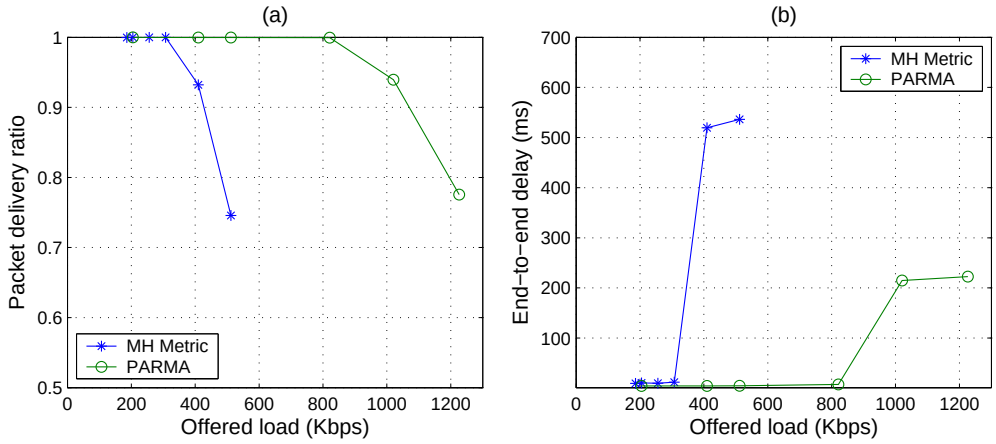


Figure 4.11: Simulation results of Scenario I. (a) Packet delivery ratio. (b) End-to-end delay.

one traffic flow, the PARMA metric has the same behavior as the MTM metric.

Scenario II: Grid Topology

A 6 by 7 regular grid topology is used as the second evaluation scenario. The distances between neighboring nodes in both the horizontal and vertical directions are 350 meters. There are three possible rates in the topology: 1, 5.5, and 11 Mbps. All nodes are stationary. There are two traffic flows in the network: flow 1 is from node 2 to node 3, and flow 2 is from node 30 to node 35, as shown in Figure 4.12.

We choose flow 1 to start transmitting packets earlier than flow 2. The traffic generation rate of flow 1 is fixed to be around 3 Mbps in order to keep the medium around flow 1 busy enough such that the congestion can occur when flow 2 is added; and the traffic generation rate of flow 2 is varied as the input parameter to gradually increase the offered load. As shown in Figure 4.12, node 2 and node 32 are in the carrier sense range of each other, so there is interference between these two flows if flow 2 takes a route including the link from node 32 to 33. Then flow 1 and flow 2 will compete for the medium when both have started. If flow 1 is close to saturation and there is a congested area around it, an ideal routing protocol will guide flow 2 to go around this congested area and achieve

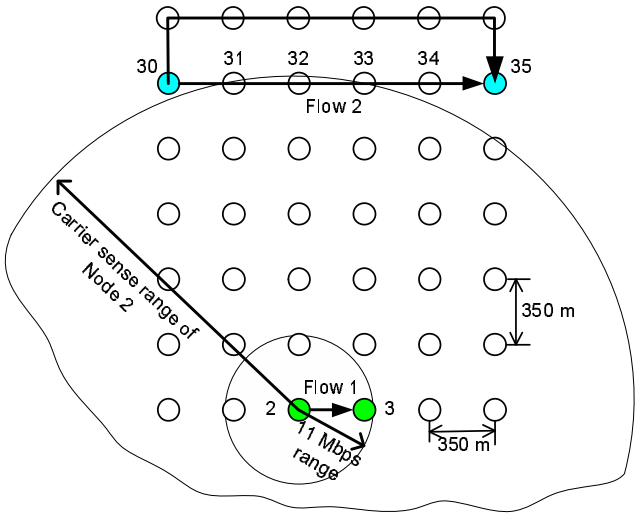
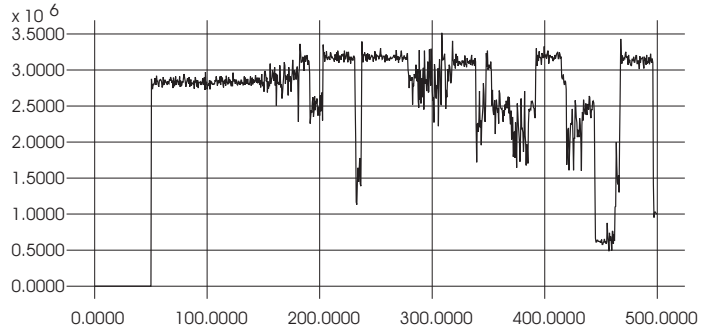


Figure 4.12: Grid topology and two flows.

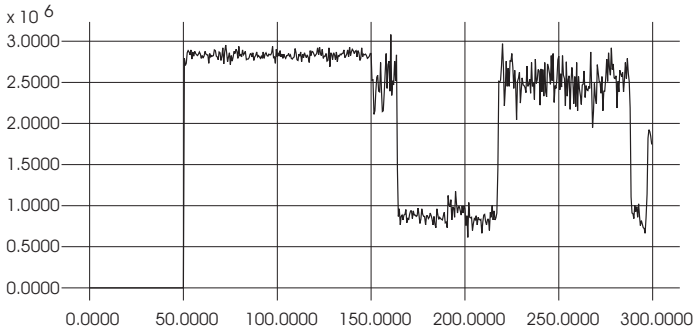
load balancing that prevents the whole system from becoming congested.

We study how the system handles interfered flows and avoids the congested area when using three routing metrics: MH, MTM, and PARMA. Figure 4.13 shows how the total throughput changes with time. In each run, flow 1 and flow 2 start at 50 second and 150 second respectively after the simulation starts. Figure 4.13(a) and (b) imply that broken links exist for MH and MTM due to interference, which results in packet drops. Figure 4.13(c) shows that with PARMA there are metric changes when flow 2 starts transmissions. After some period of settling, a new route which avoids interference from flow 1 has been used (an alternative route is shown in Figure 4.12) and thereafter the system throughput starts to go up. Figure 4.13(c) thus reflects the efficacy of our algorithm under dynamic traffic conditions.

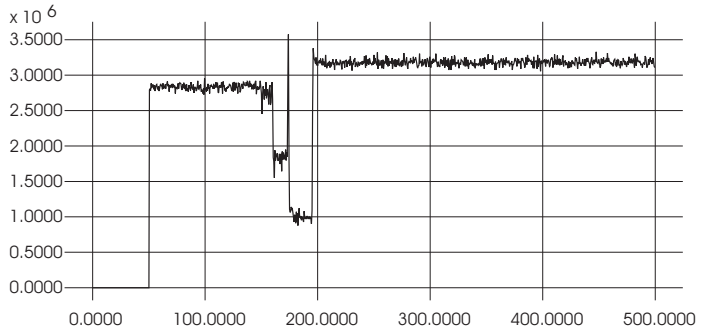
Figure 4.14 gives the throughput achieved using different routing metrics. The x-axis indicates the offered load of flow 2. We observe that using the MTM metric, system throughput drops when the offered load is increased to 550 Kbps. System throughput also drops for the MH metric. However, the throughput curve of the PARMA metric keeps increasing slowly, and is expected to increase until the system becomes excessively congested. For flow 2, MH switches among the



(a) MH



(b) MTM



(c) PARMA

Figure 4.13: Total throughput vs. time.

possible 3-hop routes, and MTM always chooses the 5-hop horizontal path from node 30 to 35. If there is no interference, PARMA will work like MTM. But due to the potential interference between these two flows if flow 2 takes any horizontal path, PARMA guides flow 2 to avoid the congested area, using routes with more hops or low rate links. Figure 4.14 also shows that the throughput of flow 2 using the PARMA metric is lower than using the MTM metric, since low rate links are used with PARMA.

The simulation results also show that PARMA can achieve the end-to-end delay comparable to

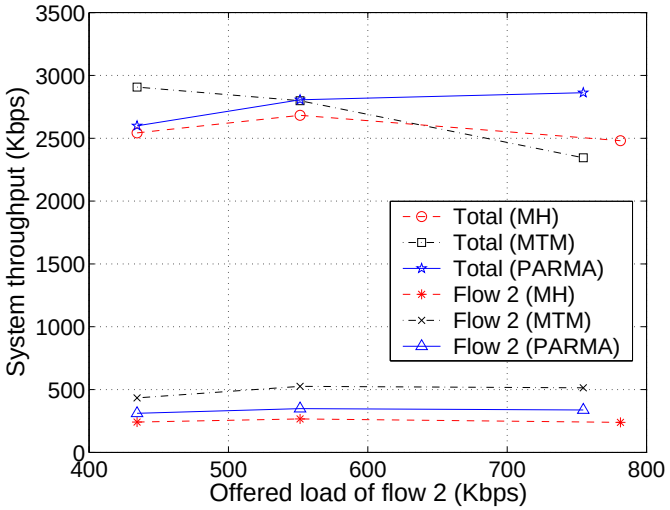


Figure 4.14: System throughput vs. offered load.

MTM. But when the system is close to saturation, PARMA has slightly larger delay. This is because PARMA helps improve the packet delivery ratio thus some packets finally arrive at their destinations but with longer latency than the average value (i.e. more hops are used).

4.3 Conclusions

In this chapter, we have investigated adaptive routing as another type of scaling mechanism for ad hoc networks. In particular, we have proposed an adaptive routing framework which is controlled by a global distributed policy manager. Control information can be disseminated over the network by flooding of control messages, or piggybacking on routing messages. While switching of routing protocols and routing metrics results in extra overhead and service discontinuity, introducing an integrated adaptive algorithm to a particular routing element such as routing metric is a more practical alternative which can be implemented more easily with smaller overhead. By taking the PHY/MAC information as state variables, the proposed framework can support cross-layer mechanisms including those based on integrated routing metrics.

Next, we have studied a specific PHY/MAC aware routing metric working with distance vector

routing protocols such as DSDV. In order to make DSDV work well with the PHY/MAC aware routing metric, we have proposed specific enhancements to the routing protocol. In addition, smoothing techniques for the link portion of the proposed metric are introduced to adjust the different change variations between the PHY/MAC layer and the network layer and also to improve route convergence. Our simulation results show that with both the PHY rate and the MAC occupancy level taken into account in the routing metric, packets can choose high rate links while avoiding congested areas in the network.

Chapter 5

Conclusions and Future Directions

In this dissertation, we have investigated mechanisms for improving the system performance and scalability of multi-hop wireless networks. We have studied some issues related to the system architecture, protocol design, and performance evaluation of multi-hop wireless networks.

We have introduced the concept of a multi-tier hierarchical hybrid wireless network, and then developed a general analytical model for the asymptotic throughput capacity and scaling properties of the proposed network. It has been proved that, in a three-tier hierarchical network with n_A access points, n_F forwarding nodes, and n_M mobile nodes, low-tier throughput capacity increases linearly with n_F, and high-tier throughput capacity increases linearly with n_A when $n_A = \Omega(\sqrt{n_F})$ and $n_A = O(n_F)$. Also, the transmission ranges are chosen as $r_H = O(1/\sqrt{n_A})$ and $r_L = O(1/\sqrt{n_F})$ for the results to hold. In the scaling regime of n_A in terms of n_F, linear scaling of throughput capacity can be achieved with a smaller number of AP's relative to the two-tier network, and the system's scaling behavior does not depend on the number of ad hoc nodes. The upper bound implies that further investments in the infrastructure do not lead to improvement in capacity scaling.

We have investigated the scaling properties and system design principles for such a hierarchical hybrid network via simulations with realistic MAC and routing protocols. The performance of an example hierarchical ad hoc network with the IEEE 802.11b radios was evaluated using *ns*-2 simulations. It is shown that modified ad hoc routing protocols including DSR, DSDV, and AODV seem to work well for the hierarchical network, and significant improvements in performance and

system throughput relative to flat ad hoc networks have been demonstrated.

We have also studied the impact of the relative densities of AP's, FN's, and MN's as well as the traffic pattern on the achievable throughput in practical systems and compared the results with those predicted by the analysis. The simulation results have demonstrated a well-defined saturation effect as the density of AP's is increased relative to a fixed density of FN's, thus leading to simple guidelines for determining the number of nodes at each tier of the hierarchy. Furthermore, we have investigated the impact of the relative channel bandwidth allocated to the two tiers of transmissions on the achievable throughput of the dual-frequency hierarchical system. The results verify that dedicated allocation of frequency bands to each tier of the network can further improve throughput provided that the bandwidths of high and low tiers are allocated according to the relative traffic load carried at two tiers.

For more advanced channel assignment schemes, a network could be designed to have dynamic allocation of bandwidth between tiers, with control of channel allocation according to the resource utilization at different tiers. Additionally, multi-channel mesh, in which dedicated channels are assigned to wireless transmissions to avoid interference and collisions, is another scaling approach. We have proposed a simple distributed channel assignment scheme based on two-hop edge coloring in Appendix B, as which may be of independent interest.

Our simulation results show that protocol overhead has great impact on system's scaling behavior, and underscore the importance of designing efficient channel access and packet scheduling protocols. Moreover, as extra wireless transmissions would degrade performance with inefficient MAC, they should be avoided in multi-channel systems. Therefore, it is particularly important to avoid contention effects at the MAC layer, and integrated scheduling and routing techniques such as IRMA [21] may be considered as a possible solution. It may also be necessary to pay attention to fairness issues because high network throughput tends to favor paths with shorter number of hops.

Overall, this part of work on hierarchical architecture has indicated that it is possible to scale network capacity quite well with a mix of several (lower-cost) radio FN's and just a few wired AP's. Both the analytical and simulation results have demonstrated the value of adding FN's to improve scaling behavior and reduce the required number of AP's relative to the two-tier case. We believe that these results are particularly relevant for emerging mesh and sensor network deployments where reducing the number of wired access points is a priority.

As another important scaling mechanism, cross-layer adaptive routing has also been explored. We have proposed an adaptive routing framework, which is controlled by a global distributed policy manager. Control information can be disseminated over the network by flooding of control messages, or piggybacking on routing messages. In this framework, routing metric, routing algorithm parameters, and/or protocol selection can be controlled in response to observed performance and external service needs. The framework supports various adaptive mechanisms, among which the integrated cross-layer routing metric incorporates adaptation of key state parameters such as link speed and congestion and is practically flexible to implement with reduced adaptation overhead and without involving service interruptions.

We have proposed a new cross-layer routing metric that takes into account both the physical bit-rate as well as the estimated medium congestion level. We have studied the proposed PHY/MAC Aware Routing Metric (PARMA) working with proactive distance vector ad hoc routing protocols such as DSDV. In order to make DSDV work well with the PHY/MAC aware routing metric, we have proposed specific enhancements to the routing protocol. In addition, smoothing techniques for the link portion of the proposed metric has been introduced to adjust the different change variations between the MAC layer and the network layer and also to improve route convergence. Our simulation results for typical multi-rate 802.11 ad hoc network scenarios have shown that routing metrics which only consider the number of hops may not achieve high throughput in multi-rate networks.

Using a metric only based on the medium transmission time would have the effect of guiding traffic to high speed links but this could cause MAC layer congestion in some area. With both the PHY rate and the MAC congestion taken into account in the routing metric, packets can choose the high rate links while also avoiding congested areas in the network thus improving throughput and decrease network congestion. It is also shown that, in cross-layer adaptation, careful design is needed to handle different change variations between different layers and avoid unintended effects.

5.1 Future Work

Both the analytical and experimental results have demonstrated that hierarchical system throughput can be scaled by using the right proportions of FN's and AP's, making it possible to design high-capacity, low-cost mesh networks with moderate number of radio FN's and only a few wired AP's. In this section, we will discuss possible algorithms for further work that could be designed explicitly for and be incorporated with hierarchical networks to further improve system throughput and other performance.

In our proposed architecture, the middle-tier (FN's) provides more economical scaling regimes for the growth of infrastructure nodes (AP's). Meanwhile, by aggregating the traffic of ad hoc nodes, the packet flows demonstrate a regular and localized manner of aggregation within each cluster, which could provide further gains by making it possible to design algorithms explicitly for the hierarchy, for instance, packet aggregation.

- Routing Optimization

  First, routing optimization could be applied in the hierarchical network. Prior work has shown that the on-demand routing protocol is more scalable than the table-driven routing protocol. For instance, suppose a flat multi-hop wireless network with n ad hoc nodes, the overhead of DSDV grows as $O(n^2)$ [77], while the overhead of AODV is linear with n [77, 78]. Our

simulations also have demonstrated that the system performance in terms of throughput and routing overhead is better with on-demand routing protocols. Meanwhile, although a strict hierarchical routing algorithm is used in the study of this work, by designing a more advanced routing protocol and incorporating an appropriate topology discovery procedure, a non-strict hierarchy can be used to provide flexibility for improving performance. For example, a non-strict hierarchical routing allows radio nodes at a given tier communicating via paths on the same tier, or going up and then down by one or more tiers. Routing optimization also allows for direct transmissions between neighboring FN's even if they do not belong to the same cluster.

- Packet Aggregation: Combine-and-Forward

  It is noted from the dual-frequency system model (as in Section 3.4.1) that system performance is degraded as the number of wireless transmissions increases as more wireless transmissions cause more interference and higher medium access overhead. Suppose the total number of packets delivered in the system is K, with a constant packet size of S bits per packet, and the average packet hop count is $\bar{h}$. For simplicity, only the successfully delivered packets count, and the MAC overhead only counts once for each hop of data transmission. Then the network traffic due to MAC overhead grows as $O(K\bar{h})$. Note that the system throughput is given as $O(KS)$. This also shows that system throughput can be improved by decreasing the number of wireless transmissions.

  The above observations and considerations motivate a "combine-and-forward" mechanism to improve system throughput. It is an approach of combining several packets, which may come from different sources and/or target to different destinations, to one single packet in the network layer, as long as they have the same next hop and the combined packet size does not exceed the packet size limitation. In the IEEE 802.11, a packet with a size exceeding

the limit, which is 2346 bytes defined by *dot11FragmentationThreshold*, would cause MAC Service Data Unit (MSDU) fragmentation [4]. In the hierarchical network, most packet flows are transferred in a regular manner within the cluster: upstream or downstream to or from the gateway AP. In this case, the benefit of combining packets would be greater than that of a flat network.

The "combine-and-forward" mechanism benefits not only in scenarios where small packet traffic tends to dominate, but also when system bandwidth is increased to accommodate increased traffic. From Figure 3.14 we observe that the throughput does not double when the total bandwidth doubles. When the bandwidth is increased, given the constant packet size, the number of packets that can be delivered at destinations increases. Note that the fraction of bandwidth used to transmit useful data packets decreases as the data rate increases [69]. The increased number of packets delivered implies an increased penalty due to overhead. Therefore, the throughput gain cannot achieve linearity with bandwidth. Combine-and-forward can improve the efficiency of bandwidth utilization.

In the wired Internet, "car pooling" as an aggregation mechanism for combining small packets to the same destination into a larger packet has been proposed in [79]. In the wireless environment, we need to trade off throughput against delay according to the application characteristics and service needs while designing appropriate aggregation mechanisms.

Appendix A

Traffic Distribution in the Random Aggregate Network

In the Random Aggregate Network, the destination node is not necessarily located at the center of the disk. For simplicity, we place the destination at the center of the disk and compute the mean number of routes served by each Voronoi cell. Let O denote the center of the planar disk, and D_V denote the outer disk of radius $2R_v$ which contains a given Voronoi cell $V \in \mathcal{V}_n \setminus \{C_d\}$, where C_d is the destination cell. We bound the probability that a segment L_i intersects V by computing the probability that L_i intersects its outer disk D_V.

Lemma A.1. *For segment L_i and Voronoi cell $V \in \mathcal{V}_n \setminus \{C_d\}$,*

$$\textit{Prob}(L_i \textit{ intersects } V) \leq \frac{c_4}{x}\sqrt{\frac{\log n}{n}}(1 - \pi x^2), \tag{A.1}$$

where x is the distance of the center of D_V from O, the center of the disk.

Proof: Suppose Voronoi cell V lies at a distance x from O. Let α denote the angle subtended at O by D_V and $A(\alpha)$ be the area of the sector formed by α. As shown in Figure A.1, the geometric property implies that

$$\alpha \leq \frac{c_5}{x}\sqrt{\frac{\log n}{n}} \quad \text{and} \quad A(\alpha) \leq \frac{c_6\alpha}{2\pi},$$

where c_5 and c_6 are constants. D_V is inscribed in this sector and divides the sector into three parts: one is the disk itself, the second is the area close to O, and the third is the area to the other side

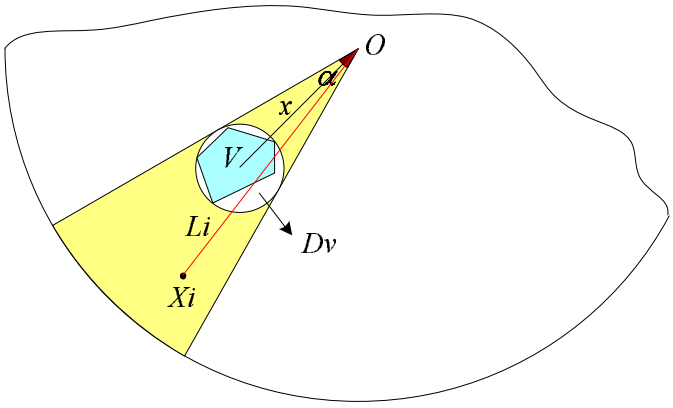


Figure A.1: Voronoi cell V.

of O. If node X_i lies in the third part of the sector, the segment L_i which connects X_i and O will intersect D_V, as shown in Figure A.1.

The sum of the areas of the first two parts are greater than the area of a sector of radius x. Thus the area of the third part is no more than the area of the whole sector subtracted by the area of the sector of radius x, i.e., $\frac{c_6\alpha}{2\pi}(1-\pi x^2)$. Therefore, it gives (A.1). □

There are n lines $\{L_i\}_{i=1}^n$, each of which connects X_i with O, for $i = 1, \ldots, n$. Since X_i is independently and uniformly distributed (i.i.d.) in the disk, the mean number of lines passing through a Voronoi cell which is at a distance x from O and not the center cell is bounded as

$$\begin{aligned}&\text{E[Number of lines in } \{L_i\}_{i=1}^n \text{ intersecting V]}\\ &\qquad \leq \frac{c_4}{x}\sqrt{n\log n}(1-\pi x^2).\end{aligned}$$

It is observed that the mean number of lines intersecting a given Voronoi cell is the function of the distance of this cell from the center, and the lines passing through get denser when the Voronoi cell is closer to the center. The traffic handled by a Voronoi cell is proportional to the number of lines passing through it.

For Voronoi cells close to the center, $x \ll (1/\sqrt{\pi})$ holds, thus $1-\pi x^2 \cong 1$. Then we have,

when $x \ll (1/\sqrt{\pi})$,

$$\text{E[Number of lines intersecting V]} \leq \frac{c_4}{x}\sqrt{n \log n}. \tag{A.2}$$

It is implied from (A.2) that the number of traffic handled by a Voronoi cell that is close to the aggregation node is proportional to the reciprocal of its distance from the aggregation node. In order to maximize the capacity of a Random Aggregate Network, it suggests that the resource allocated to each Voronoi cell has the relationship as (A.2). Therefore, when the infrastructure is used as a shortcut for traffic that needs to traverse a long distance, infrastructure nodes would become traffic hotspots. In this sense, (A.2) provides a basis for designing scheduling algorithms to overcome the capacity bottleneck at the hotspots in close proximity to infrastructure nodes.

Appendix B

A Simple Distributed Channel Assignment Scheme

The theoretical results obtained in Chapter 2 show that the system capacity grows linearly with the system bandwidth under the assumption of perfect MAC and routing. However, in a system with practical MAC and routing protocols, the throughput gain with increased bandwidth is not linear due to the significant and non-linear overhead involved for the MAC layer and routing control as the system bandwidth increases (as discussed in Section 3.4). This motivates allocation of dedicated channels to wireless transmissions to avoid increased MAC layer overheads and scheduling inefficiency which arise in the single channel system.

Employing multiple orthogonal channels (or non-overlapping frequency bands) and multiple radio cards at nodes is an important way to reduce packet collisions and improve scalability. For example, the IEEE 802.11 physical layer specifications define multiple channels and allow the simultaneous and non-interfering use of some of these channels thus offering the opportunity to increase the effective capacity. The future IEEE 802.11s standard [6] is also expected to support multi-channel ad hoc network capability. A recent work in [80] has shown that in a static multi-channel wireless networks, if we assign channels (to links) and radios (to channels) appropriately, the maximum throughput can be achieved by fully utilizing all available channels.

Orthogonal Frequency Division Multiplexing (OFDM) could provide variable bandwidth and enough sub-carriers to a multi-hop wireless network such that packets are transmitted in a non-interfering manner.

More recently, linear programming (LP) has been used to compute the optimal throughput of a multi-hop wireless network, in which the multi-frequency and/or multi-radio scenarios can be included in the constraints [81–83]. However, this kind of approach requires a central controller entity which may not be practicable in certain real world scenarios. The dynamically changing topology as an input requires high computational effort to formulate. Therefore, heuristics are attractive for actual protocol design.

These observations motivate this work of designing a simple heuristic distributed channel assignment algorithm for multi-channel multi-radio wireless networks. First we briefly review the relevant graph coloring background.

B.1 Graph Coloring

Let us look at a multi-hop ad hoc wireless network consisting of radio devices. Correspondingly suppose we have a graph $G = (V, E)$ whose nodes represent radio devices and whose edges represent that the two end nodes are within the transmission range. Let $\Delta(G)$ be the maximum degree of G.

Vertex coloring of G is to assign a color to each vertex so that no two adjacent vertices have the same color. *Edge coloring* of G is to assign a color to each edge so that no two adjacent edges have the same color. For edge coloring, the smallest number of colors needed is the *edge-chromatic number* of G, denoted $\chi'(G)$. Vizing's Theorem states that every graph G satisfies $\Delta(G) \leq \chi'(G) \leq \Delta(G) + 1$ [50].

A highly utilized link is preferred to be assigned multiple channels. In this case, a multi-hop wireless network could be a *multigraph*, which is defined as a graph with possibly multiple edges between any two vertices. The *multiplicity* of a multigraph is the maximum number of edges joining two vertices, denoted $M(G)$. According to the edge coloring result of graph theory [50], it is

possible to color a multigraph with at most $\Delta(G) + M(G)$ colors, i.e. $\chi'(G) \leq \Delta(G) + M(G)$.

B.2 Problem Modeling

In our model, there are multiple orthogonal channels in the network. Each node is equipped with multiple radios (with half-duplex operation), which are operating at different channels to permit simultaneous transmissions. Each radio is tuned to a specified channel on a demand basis: when a source node wants to send packets to a destination, the routing function decides a path for each packet, and then the channel assignment algorithm allocate channels to the links along each path if not being assigned yet. It is possible that a channel is taken back if the corresponding link is idle for some period of time. Each radio operates on a single channel at a given time, and does not support channel switching on a per-packet basis. Multiple channels per link might be allowed in the model.

Suppose the number of orthogonal channels available is enough to allow non-interfering transmissions, and each radio device of the network is equipped with a sufficient number of radio cards by which it can communicate with other nodes using different channels. Therefore, there are no hardware constraints. Now we look at the interference-free constraints.

B.2.1 Interference Avoidance in Multi-Hop Wireless Networks

Based on the network connectivity and the interference avoidance requirements, a conflict-free channel assignment problem can be formulated as a graph coloring problem by equating channels with colors. We also want to determine the minimum number of channels, K, required for a conflict-free channel assignment.

In a multi-hop wireless network, the radios that are sufficiently close and cause interference to one another must be assigned different channels, while only these radios that are distant enough that there is no interference among them can share channels.

In particular, interference could be caused by transmissions of neighboring nodes (i.e. nodes that are one hop away); this is called direct interference. Interference or collisions can also be due to hidden terminals, i.e., due to transmissions of nodes which are two hops away; this secondary interference is also called hidden terminal interference [84, 85].

A successful channel assignment should avoid both direct and hidden terminal interference, i.e., a link cannot share the same channel with other links which are within two hops. Therefore, the channel assignment problem can be formulated as a two-hop edge coloring problem by equating channels with colors.

B.3 Distributed Channel Assignment Algorithm

We design a distributed algorithm for the two-hop edge coloring problem. As the graph coloring problem is NP-hard, this algorithm is based on a randomized, distributed edge coloring algorithm proposed in [86], which has been proved to be extremely simple while can compute nearly optimal coloring very quickly.

B.3.1 Background of One-Hop Edge Coloring Algorithm

In the randomized, distributed algorithm, initially each edge is given a list, or *palette*, of a certain number of colors. The computation proceeds in rounds. During each round, each uncolored edge, in parallel, first picks a tentative color at random from its palette. If no neighboring edges picks the same color, the color becomes final and the algorithm stops for that edge. Otherwise, the edge's palette is updated by removing the colors used by its neighbors, and a new attempt is performed in the next round.

During each round, a node performs the above computation and exchanges information with its neighbors in the graph. The complexity of the protocol is the number of rounds needed to get a

valid edge coloring of the graph.

It has been proved that this algorithm is very simple and distributed, yet guarantees that the number of colors used is a small multiplicative factor of the optimal value.

B.3.2 Distributed Channel Assignment Algorithm

If the two-hop edge coloring algorithm is directly applied to the network of physical connectivity, there may exist some idle links (links do not carry traffic) to which channels are still allocated. Since the computation complexity increases as the number of edges needed to be considered grows, it is preferred to assign channels to the links being used. For this purpose, we define the status of each link to be either "active" or "inactive": when a link is chosen by the routing function to transfer packets, it becomes "active"; when a link is idle for some period of time, it becomes "inactive". Note that in on-demand ad hoc routing protocols [12, 13], a route is removed from the routing table when it is not used for a period of time. Therefore, the link status defined here can be updated according to the routing information: a link becomes active when a route containing it is discovered, and becomes inactive when all routes containing it are removed.

Now we consider a graph $G = (V, E)$ based on the physical connectivity of the network. Suppose the maximum degree is $\Delta(G)$. We associate a palette of N ($N \geq K$, where K is the minimum number of channels required for a conflict-free channel assignment) possible colors to each edge. Initially, the edge palettes are identical. The two-hop edge coloring algorithm repeatedly executes the following steps for each active edge, until all active edges are colored.

- Step 1: Each uncolored edge $e = uv \in E$ chooses uniformly at random a tentative color from its associated palette.

- Step 2.a: Tentative colors of edge e are checked against the colors of direct neighboring edges (edges that are incident on either u or v) for possible color collision. If a collision occurs, the

edge rejects its tentative color.

- Step 2.b: Repeat Step 2 by checking against colors of edges that are two-hop neighbors (edges that are incident on either $N(u)$ or $N(v)$, where $N(u)$ represents the set of neighbors of node u) for possible color collision. If a collision occurs, the edge rejects its tentative color.

- Step 3: Edges that have chosen a valid color are marked colored, and their colors are removed from palettes of direct and two-hop neighboring edges.

The algorithm is distributed provided information is exchanged up to two-hop neighbors. Its complexity is the number of rounds needed to obtain a valid two-hop edge coloring.

If there is an edge becoming inactive, the algorithm executes the following:

- Step 4.a: Unmark the color(s) and put the color(s) in its associated palette.

- Step 4.b: The released colors are put in palettes of direct and two-hop neighboring edges.

Since removing a color from its edge would not create coloring conflicts with other edges, the algorithm will execute without being affected until all active edges in the graph are colored. Additionally, it can be shown that this algorithm supports assigning multiple channels to a link.

In order to implement the channel assignment algorithm in a distributed manner, the information generated by any node has to be propagated up to two hops away from the node, and may be carried by a dedicated control channel. Channel assignment works closely with route selection. If the path goes through a region where medium is busy, using this path would aggravate the congestion. Thus we may consider them jointly to achieve optimal performance.

B.3.3 Example

There are 30 nodes randomly distributed in an example network of dimension 1000mx1000m. Let the common transmission range be 250 meters. The obtained physical connectivity graph $G =$

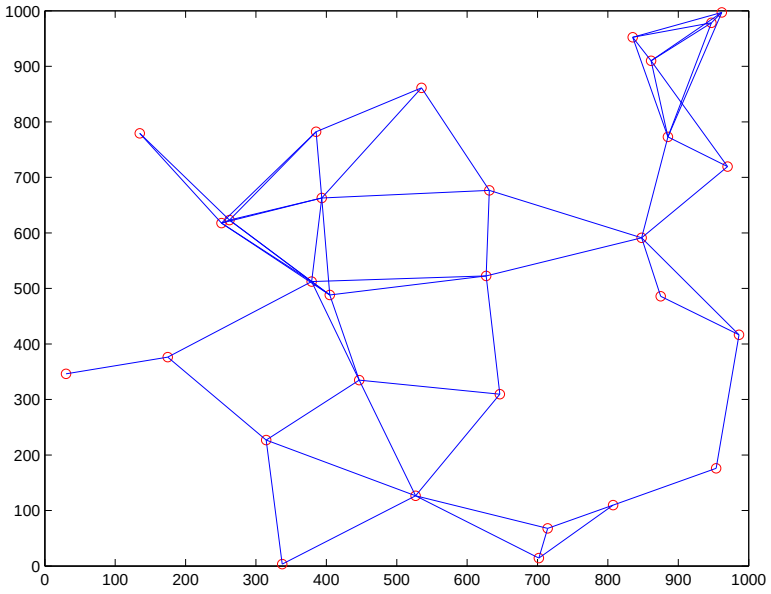


Figure B.1: Physical connectivity of a wireless network, G.

(V, E) is shown in Figure B.1. It is also known that the maximum degree is $\Delta(G) = 7$, and the edge number is $|E| = 61$.

Using our two-hop edge coloring algorithm, 22 colors, i.e., 22 channels are required for conflict-free transmissions. Note that only 8 channels are needed if we only consider direct interference avoidance.

Note that in this example, channel assignment is applied to the physical connectivity of all the available links. The total number of channels required would be less if the channels are assigned on a demand-basis, i.e., the channel is assigned to a link only when this link is chosen for data transfer.

References

[1] J. E. Wieselthier, D. J. Baker, and A. Ephremides, “Survey of problems in the design of an HF Intra-Task Force communication network,” Naval Research Laboratory (NRL), Washington, DC, Tech. Rep. 8501, 1981.

[2] R. E. Kahn, S. A. Gronemeyer, J. Burchfiel, and R. C. Kunzelman, “Advances in packet radio technology,” in *Proc. IEEE*, vol. 66, Nov. 1978, pp. 1468–1496.

[3] N. Shacham and J. Westcott, “Future directions in packet radio architectures and protocols,” in *Proc. IEEE*, vol. 75, no. 1, 1987, pp. 83–99.

[4] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specification*, IEEE Std. 802.11, 1999.

[5] *Internet Protocol, Version 6 (IPv6) Specification*, IETF Std., 1998. [Online]. Available: http://www.ietf.org/rfc/rfc2460.txt

[6] *IEEE 802.11s Tutorial: Overview of the Amendment for Wireless Local Area Mesh Networking*, IEEE Std., 2006. [Online]. Available: http://www.ieee802.org/802_tutorials/nov06/802.11s_Tutorial_r5.pdf

[7] Meshdynamics structured mesh. ACG and Meshdynamics. [Online]. Available: http://www.meshdynamics.com/

[8] Infrasructure free broadband communications. PacketHop, Inc. [Online]. Available: http://www.packethop.com/

[9] P. Gupta and P. R. Kumar, “The capacity of wireless networks,” *IEEE Trans. Inform. Theory*, vol. 46, pp. 388–404, Mar. 2000.

[10] C. E. Perkins, *Ad Hoc Networking*, C. E. Perkins, Ed. Addison-Wesley, 2001.

[11] D. J. Baker and A. Ephremides, “The architectural organization of a mobile radio network via a distributed algorithm,” *IEEE Trans. Commun.*, vol. 29, no. 11, pp. 1694–1701, Nov. 1981.

[12] D. B. Johnson and D. A. Maltz, “Dynamic source routing in ad hoc wireless networks,” in *Mobile Computing*, T. Imielinski and H. Korth, Eds. Kluwer Academic Publishers, 1996, ch. 5, pp. 153–181.

[13] C. E. Perkins and E. M. Royer, “Ad hoc on-demand distance vector routing,” in *Proc. 2nd IEEE Workshop on Mobile Computing Systems and Applications (WMCSA’99)*, Feb. 1999, pp. 90–100.

[14] C. R. Lin and M. Gerla, “Adaptive clustering for mobile wireless networks,” *IEEE J. Select. Areas Commun.*, vol. 15, no. 7, pp. 1265–1275, 1997.

[15] A. Iwata, C.-C. Chiang, G. Pei, M. Gerla, and T.-W. Chen, "Scalable routing strategies for ad-hoc wireless networks," *IEEE J. Select. Areas Commun.*, pp. 1369–1379, Aug. 1999.

[16] E. M. Belding-Royer, "Hierarchical routing in ad hoc mobile networks," *Wireless Communications and Mobile Computing*, pp. 515–532, 2002.

[17] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. IEEE INFOCOM'03*, vol. 2, Mar. 2003, pp. 1543–1552.

[18] U. C. Kozat and L. Tassiulas, "Throughput capacity of random ad hoc networks with infrastructure support," in *Proc. 9th Annual International Conference on Mobile Computing and Networking (MobiCom'03)*, Sept. 2003, pp. 55–65.

[19] A. Zemlianov and G. de Veciana, "Capacity of ad hoc wireless networks with infrastructure support," *IEEE J. Select. Areas Commun.*, vol. 23, no. 3, pp. 657–667, Mar. 2005.

[20] H. Holma and A. Toskala, Eds., *WCDMA for UMTS Radio Access for Third Generation Mobile Communications*. John Wiley & Sons Ltd, 2000.

[21] Z. Wu, S. Ganu, and D. Raychaudhuri, "IRMA: Integrated routing and mac scheduling in multihop wireless mesh networks," in *Proc. Second IEEE Workshop on Wireless Mesh Networks (WiMesh 2006)*, Sept. 2006.

[22] L. Raju, S. Ganu, B. Anepu, I. Seskar, and D. Raychaudhuri, "BEacon assisted discovery protocol (BEAD) for self-organizing hierarchical wireless ad-hoc networks," in *Proc. IEEE Global Telecommunication Conference (GLOBECOM 2004)*, Nov. 2004.

[23] S. Zhao, Z. Wu, A. Acharya, and D. Raychaudhuri, "PARMA: A PHY/MAC aware routing metric for ad-hoc wireless networks with multi-rate radios," in *Proc. IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2005)*, June 2005.

[24] *Air Interface for Fixed Broadband Wireless Access Systems*, IEEE 802 LAN/MAN Standards Committee Std. 802.16, 2001.

[25] *Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low Rate Wireless Personal Area Networks (LR-WPANs)*, IEEE 802 LAN/MAN Standards Committee Std. 802.15.4, 2003.

[26] S. Ganu, S. Zhao, L. Raju, B. Anepu, I. Seskar, and D. Raychaudhuri, "Architecture and prototyping of an 802.11-based self-organizing hierarchical ad-hoc wireless network (SOHAN)," in *Proc. IEEE International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC'04)*, Sept. 2004.

[27] D. Bertsekas and R. Gallager, *Data Networks*, 2nd ed. Prentice Hall, 1992.

[28] J. G. Proakis, *Digital Communications*, 4th ed. McGraw-Hill, 2001.

[29] A. Ephremides, "Ad hoc networks: Not an ad hoc field anymore," *Wireless Communications and Mobile Computing*, vol. 2, pp. 441–448, 2002.

[30] A. Kamerman and L. Monteban, "WaveLAN-II: A high-performance wireless LAN for the unlicensed band," *Bell Labs Technical Journal*, pp. 118–133, 1997.

[31] G. Holland, N. Vaidya, and P. Bahl, "A rate-adaptive MAC protocol for multi-hop wireless networks," in *Proc. ACM/IEEE MobiCom'01*, July 2001.

[32] A. S. Tanenbaum, *Computer Networks*, 3rd ed. Saddle River, NJ: Prentice Hall, Mar. 1996.

[33] C. E. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector routing (DSDV) for mobile computers," in *Proc. ACM SIGCOMM'94 Conference on Communications Architectures, Protocols and Applications*, 1994, pp. 234–244.

[34] P. Jacquet, P. Muhlethaler, and A. Qayyum, "Optimized link state routing (OLSR) protocol," Internet Draft, draft-ietf-manet-olsr-01.txt, Feb. 2000.

[35] Z. J. Haas and M. R. Pearlman, "A new routing protocol for the reconfigurable wireless networks," in *Proc. IEEE International Conference on Universal Personal Communications*, Oct. 1997.

[36] S.-J. Lee, M. Gerla, and C.-K. Joh, "A simulation study of table-driven and on-demand routing protocols for mobile ad hoc networks," in *Proc. IEEE International Conference on Communications (ICC'00)*, vol. 3, 2000, pp. 1702–1706.

[37] Y. Ko and N. H. Vaidya, "Location-aided routing (LAR) in mobile ad hoc networks," in *Proc. ACM/IEEE MobiCom'98*, Oct. 1998, pp. 66–75.

[38] D. Estrin, R. Govindan, J. Heidemann, and S. Kumar, "Next century challenges: Scalable coordination in sensor networks," in *Proc. ACM/IEEE MobiCom'99*, Aug. 1999, pp. 263–270.

[39] B. Karp and H. T. Kung, "GPSR: Greedy perimeter stateless routing for wireless networks," in *Proc. ACM/IEEE MobiCom'00*, Aug. 2000, pp. 243–254.

[40] N. Li, J. C. Hou, and L. Sha, "Design and analysis of an MST-based topology control algorithm," *IEEE Trans. Wireless Commun.*, vol. 4, no. 3, pp. 1195–1206, May 2005.

[41] F. Dai and J. Wu, "Control frame shaping in power controlled and directional MAC protocols," *Ad Hoc & Sensor Wireless Networks*, vol. 00, pp. 1–26, 2007.

[42] V. Kawadia and P. R. Kumar, "A cautionary perspective on cross layer design," *IEEE Wireless Commun. Mag.*, 2003.

[43] S. Toumpis, "Capacity bounds for three classes of wireless networks: Asymmetric, cluster, and hybrid," in *Proc. 5th ACM International Symposium on Mobile Ad Hoc Networking and Computing (MobiHoc)*, May 2004, pp. 133–144.

[44] A. Agarwal and P. R. Kumar, "Capacity bounds for ad hoc and hybrid wireless networks," *ACM SIGCOMM Computer Communications Review*, vol. 34, no. 3, pp. 71–81, July 2004.

[45] M. Franceschetti, O. Dousse, D. Tse, and P. Tiran, "Closing the gap in the capacity of random wireless networks," in *Proc. International Symposium on Information Theory (ISIT)*, June/July 2004, p. 439.

[46] M. Grossglauser and D. Tse, "Mobility increases the capacity of ad hoc wireless networks," *IEEE/ACM Trans. Networking*, vol. 10, no. 4, pp. 477–486, 2002.

[47] A. Özgür, O. Lévêque, and D. Tse, "How does the information capacity of ad hoc networks scale?" in *Proc. Forty-fourth Annual Allerton Conference on Communication, Control and Computing*, Sept. 2006, pp. 27–29.

[48] ——, "Hierarchical cooperation achieves linear capacity scaling in ad hoc networks," in *Proc. IEEE INFOCOM'07*, 2007.

[49] T. M. Cover and J. A. Thomas, *Elements of Information Theory*. Wiley-Interscience, 1991.

[50] R. Diestel, *Graph Theory*, 3rd ed. Springer, 2006.

[51] E. J. Duarte-Melo and M. Liu, "Data-gathering wireless sensor networks: Organization and capacity," *Computer Networks*, pp. 519–537, Nov. 2003.

[52] K. Fall and K. Varadhan. The *ns* manual (2002). The VINT Project, UC Berkeley, LBL, USC/ISI, and Xerox PARC. [Online]. Available: http://www.isi.edu/nsnam/ns/doc/ns_doc.pdf

[53] Aodv-uu: The aodv routing protocol implementation by uppsala university. Uppsala University. [Online]. Available: http://core.it.uu.se/AdHoc/AodvUUImpl

[54] C. E. Perkins, E. M. Royer, S. R. Das, and M. K. Marina, "Performance comparison of two on-demand routing protocols for ad hoc netwroks," *IEEE Personal Commun. Mag.*, pp. 16–28, Feb. 2001.

[55] L.-L. Xie and P. R. Kumar, "A network information theory for wireless communications: Scaling laws and optimal operation," *IEEE Trans. Inform. Theory*, vol. 50, pp. 748–767, May 2004.

[56] J. Li, C. Blake, D. S. J. De Couto, H. I. Lee, and R. Morris, "Capacity of ad hoc wireless networks," in *Proc. ACM/IEEE MobiCom'01*, July 2001, pp. 61–69.

[57] J. Mitola III, "Cognitive radio: An integrated agent architecture for software radio," Ph.D. dissertation, Royal Institute of Technology (KTH), Sweden, May 2000.

[58] A. R. S. Bahai, B. R. Saltzberg, and M. Ergen, *Multi-Carrier Digital Communications: Theory and Applications of OFDM*, 2nd ed. 233 Spring Street, New York: Springer Science+Business Media, Inc., 2004.

[59] V. D. Park and M. S. Corson, "A highly adaptive distributed routing algorithm for mobile wireless networks," in *Proc. IEEE INFOCOM'97*, Apr. 1997, pp. 1405–1413.

[60] J. Broch, D. A. Maltz, D. B. Johnson, Y.-C. Hu, and J. Jetcheva, "A performance comparison of multi-hop wireless ad hoc network routing protocols," in *Proc. ACM/IEEE MobiCom'98*, Oct. 1998, pp. 85–97.

[61] V. Ramasubramanian, Z. J. Haas, and E. G. Sirer, "Sharp: A hybrid adaptive routing protocol for mobile ad hoc networks," in *Proc. The Fourth ACM International Symposium on Mobile Ad Hoc Networking and Computing MobiHoc'03*, June 2003, pp. 303–314.

[62] A. B. McDonald and T. Znati, "A dual-hybrid adaptive routing strategy for wireless ad-hoc networks," in *Proc. IEEE Wireless Communications and Networking Conference 2000 (WCNC 2000)*, Sept. 2000.

[63] S. Giannoulis, C. Katsanos, S. Koubias, and G. Papadopoulos, "A hybrid adaptive routing protocol for ad hoc wireless networks," in *Proc. 2004 IEEE International Workshop on Factory Communication Systems*, 2004, pp. 287–290.

[64] X.-H. Lin, Y.-K. Kwok, and V. K. N. Lau, "On channel-adaptive routing in an IEEE 802.11b based ad hoc wireless network," in *Proc. IEEE Global Telecommunication Conference (GLOBECOM 2003)*, Dec. 2003.

[65] D. S. J. D. Couto, D. Aguayo, B. A. Chambers, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proc. ACM/IEEE MobiCom'03*, Sept. 2003.

[66] Y.-C. Hu and D. B. Johnson, "Caching strategies in on-demand routing protocols for wireless ad hoc networks," in *Proc. ACM/IEEE MobiCom'00*, Aug. 2000.

[67] *The Dynamic Source Routing Protocol for Mobile Ad Hoc Networks (DSR)*, IETF Drafts and RFCs Std. [Online]. Available: draft-ietf-manet-dsr-07.txt

[68] S. Zhao, I. Seskar, and D. Raychaudhuri, "Performance and scalability of self-organizing hierarchical ad hoc wireless networks," in *Proc. IEEE Wireless Communications and Networking Conference (WCNC'04)*, Mar. 2004.

[69] B. Awerbuch, D. Holmer, and H. Rubens, "High throughput route selection in multi-rate ad hoc wireless networks," in *Proc. 1st IFIP TC6 Working Conference on Wireless On-demand Network Systems (WONS 2004)*, Jan. 2004, pp. 251–268.

[70] D. S. J. De Couto, D. Aguayo, B. A. Chambers, and R. Morri, "Performance of multihop wireless networks: Shortest path is not enough," in *Proc. ACM 1st Workshop on Hot Topics in Network (HotNets-I)*, July 2002.

[71] D. S. J. De Couto, D. Aguayo, B. A. Chambers, and R. Morris, "A high-throughput path metric for multi-hop wireless routing," in *Proc. ACM/IEEE MobiCom'03*, Sept. 2003.

[72] R. Draves, J. Padhye, and B. Zill, "Routing in multi-radio, multi-hop wireless mesh networks," in *Proc. ACM/IEEE MobiCom'04*, Sept./Oct. 2004, pp. 114–128.

[73] Y. Yang, J. Wang, and R. Kravets, "Interference-aware load balancing for multihop wireless networks," University of Illinois at Urbana-Champaign, Tech. Rep. 361702, 2005.

[74] Y.-C. Hu and D. B. Johnson, "Exploiting congestion information in network and higher layer protocols in multihop wireless ad hoc networks," in *Proc. 24th International Conference on Distributed Computing Systems (ICDCS 2004)*, Mar. 2004, pp. 301–310.

[75] A. Goldsmith, *Wireless Communications*. Cambridge University Press, Aug. 2005.

[76] P. Gopalakrishnan, "Methods for predicting the throughput characteristics of rate-adaptive wireless LANs," M. Eng. thesis, Rutgers, The State University of New Jersey, 2004.

[77] E. M. Royer and C.-K. Toh, "A review of current routing protocols for ad hoc mobile wireless networks," *IEEE Personal Commun. Mag.*, pp. 46–55, Apr. 1999.

[78] S.-J. Lee, "Routing and multicasting strategies in wireless mobile ad hoc networks," Ph.D. dissertation, Univ. of California, Los Angeles, Los Angeles, 2000.

[79] B. R. Badrinath and P. Sudame, "Gathercast: The design and implementation of a programmable aggregation mechanism for the internet," in *Proc. IEEE International Conference on Computer Communications and Networks (ICCCN)*, Oct. 2000, pp. 206–213.

[80] P. Kyasanur and N. H. Vaidya, "Capacity of multi-channel wireless networks: Impact of number of channels and interfaces," in *Proc. ACM/IEEE MobiCom'05*, Aug. 2005, pp. 43–57.

[81] K. Jain, J. Padhye, V. N. Padmanabhan, and L. Qiu, "Impact of interference on multi-hop wireless network performance," in *Proc. ACM/IEEE MobiCom'03*, Sept. 2003, pp. 66–80.

[82] M. Kodialam and T. Nandagopal, "Characterizing the capacity region in multi-radio multi-channel wireless mesh networks," in *Proc. ACM/IEEE MobiCom'05*, Aug. 2005.

[83] M. Alicherry, R. Bhatia, and L. E. Li, "Joint channel assignment and routing for throughput optimization in multi-radio wireless mesh networks," in *Proc. ACM/IEEE MobiCom'05*, Aug. 2005.

[84] J. J. Garcia-Luna-Aceves and J. Raju, "Distributed assignment of codes for multihop packet-radio networks," in *Proc. IEEE MILCOM*, Nov. 1997, pp. 450–454.

[85] R. Battiti, A. A. Bertossi, and M. A. Bonuccelli, "Assigning codes in wireless networks: Bounds and scaling properties," *Wireless Networks*, vol. 5, pp. 195–209, 1999.

[86] M. V. Marathe, A. Panconesi, and L. D. R. JR, "An experimental study of a simple, distributed edge-coloring algorithm," *ACM Journal of Experimental Algorithms*, vol. 9, no. 1.3, pp. 1–22, 2004.

[87] S. Singh, M. Woo, and C. S. Raghavendra, "Power-aware routing in mobile ad hoc networks," in *Proc. ACM/IEEE MobiCom'98*, Oct. 1998, pp. 181–190.

[88] B. Liu, Z. Liu, and D. Towsley, "On the capacity of hybrid wireless networks," in *Proc. IEEE INFOCOM'03*, 2003.

[89] S. Zhao, K. Tepe, I. Seskar, and D. Raychaudhuri, "Routing protocols for self-organizing hierarchical ad-hoc wireless networks," in *Proc. 2003 IEEE Sarnoff Symposium*, Mar. 2003.

[90] *Bluetooth Core Specification v2.0*, Bluetooth Special Interest Group Std. 802.15. [Online]. Available: http://www.bluetooth.org

[91] L. L. Peterson and B. S. Davie, *Computer Networks, A Systems Approach*, 2nd ed. Morgan Kaufmann, 2000.

[92] D. H. Smith, S. Hurley, and S. U. Thiel, "Improving heuristics for the frequency assignment problem," *European Journal of Operations Research*, vol. 107, pp. 76–86, May 1998.

[93] E. Arikan, "Some complexity results about packet radio networks," *IEEE Trans. Inform. Theory*, vol. 30, no. 4, pp. 681–685, July 1984.

[94] A. A. Bertossi and M. A. Bonuccelli, "Code assignment for hidden terminal interference avoidance in multihop packet radio networks," *IEEE/ACM Trans. Networking*, vol. 3, no. 4, pp. 441–449, Aug. 1995.

[95] I. Cidon and M. Sidi, "Distributed assignment algorithms for multihop packet radio networks," *IEEE Trans. Comput.*, vol. 38, no. 10, pp. 1353–1361, Oct. 1989.

[96] J. Li, Z. J. Haas, and M. Sheng, "Capacity evaluation of multi-channel multi-hop ad hoc networks," in *Proc. International Conference on Personal Wireless Communications (ICPWC 2002)*, 2002.

[97] S.-L. Wu, C.-Y. Lin, Y.-C. Tseng, and J.-L. Sheu, "A new multi-channel MAC protocol with on-demand channel assignment for multi-hop mobile ad hoc networks," in *Proc. International Symposium on Parallel Architectures, Algorithms and Networks (I-SPAN 2000)*, Dec. 2000, pp. 232–237.

[98] C.-Y. Chang, P.-C. Huang, C.-T. Chang, and Y.-S. Chen, "Dynamic channel assignment and reassignment for exploiting channel reuse opportunities in ad hoc wireless networks," *IEICE Transactions on Communications*, vol. E86-B, no. 4, pp. 1234–1246, 2003.

[99] A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, "A multi-radio unification protocol for IEEE 802.11 wireless networks," in *Proc. International Conference on Broadband Networks (BroadNets)*, 2004.

[100] A. Nasipuri, J. Zhuang, and S. R. Das, "A multichannel CSMA MAC protocol for multihop wireless networks," in *Proc. IEEE WCNC*, 1999.

[101] P. Kyasanur and N. H. Vaidya, "Routing and interface assignment in multi-channel multi-interface wireless networks," in *Proc. IEEE WCNC*, 2005.

[102] R. Chandra, P. Bahl, and P. Bahl, "MultiNet: Connecting to multiple IEEE 802.11 networks using a single wireless card," in *Proc. IEEE INFOCOM*, Mar. 2004.

[103] (2005) Maxim 2.4GHz 802.11b Zero-IF transceivers. Maxim Integrated Products. [Online]. Available: http://pdfserv.maximic.com/en/ds/MAX2820-MAX2821.pdf

[104] P. Bahl, R. Chandra, and J. Dunagan, "SCCH: Slotted seeded channel hopping for capacity improvement in IEEE 802.11 ad-hoc wireless networks," in *Proc. ACM/IEEE MobiCom'04*, Sept./Oct. 2004.

[105] D. B. West, *Introduction to Graph Theory*, 2nd ed. Prentice Hall, 2001.

[106] (2004) Philadelphia proposes citywide wi-fi network. CORANTE Tech News. [Online]. Available: http://www.corante.com/bwia/archives/006073.html

[107] M. Gerla and J. T.-C. Tsai, "Multicluster, mobile, multimedia radio network," *ACM-Baltzer Journal of Wireless Networks*, pp. 255–265, 1995.

[108] T. Ojanperä and R. Prasad, Eds., *WCDMA: Towards IP Mobility and Mobile Internet*. Artech House, 2001.

[109] N. Gupta and P. R. Kumar, "A performance analysis of the 802.11 wireless LAN medium access control," *Communications in Information and Systems*, vol. 3, no. 4, pp. 279–304, Sept. 2004.

[110] S. Zhao and D. Raychaudhuri, "On the scalability of hierarchical hybrid wireless networks," in *Proc. Conference on Information Sciences and Systems (CISS 2006)*, Mar. 2006.

Curriculum Vita

Suli Zhao

1990-1994 B.S. in Electrical Engineering, Beijing University of Posts and Telecommunications, China

1994-1997 M.S. in Electrical Engineering, Beijing University of Posts and Telecommunications, China

2002-2007 Ph.D. in Electrical and Computer Engineering, Rutgers, The State University of New Jersey, New Jersey

1997-1998 R&D Engineer, CDMA R&D Center, Beijing, China

1998-2001 Research Engineer, Nokia Research Center/Beijing, China

2002-2007 Graduate Research Assistant, WINLAB, ECE Department, Rutgers University, New Jersey

1999 S. Zhao, "The applications of retransmission schemes in the radio interface of mobile communication systems," *Fifth Asia-Pacific Conference on Communications and Fourth Optoelectronics and Communications Conference (APCC/OECC'99)*, Beijing, China, Oct. 1999.

2001 Y. Wang, S. Zhao, and H. Guan, "Method for adaptively setting transmission parameters for a random access channel transmission uplink procedure in a wireless communication system," *U.S. Patent 7013146*, filed 2001.

2003 S. Zhao, K. Tepe, I. Seskar, and D. Raychaudhuri, "Routing protocols for self-organizing hierarchical ad-hoc wireless networks," *IEEE Sarnoff 2003 Symposium*, Trenton, NJ, Mar. 2003.

2004 S. Zhao, I. Seskar, and D. Raychaudhuri, "Performance and scalability of self-organizing hierarchical ad hoc wireless networks," *IEEE Wireless Communications and Networking Conference (WCNC 2004)*, Atlanta, GA, Mar. 2004.

S. Ganu, S. Zhao, L. Raju, B. Anepu, I. Seskar, and D. Raychaudhuri, "Architecture and prototyping of an 802.11-based self-organizing hierarchical ad-hoc wireless networks (SOHAN)," *IEEE 15th International Symposium on Personal, Indoor, and Mobile Radio Communications (PIMRC 2004)*, Barcelona, Spain, Sept. 2004.

2005 S. Zhao, Z. Wu, A. Acharya, and D. Raychaudhuri, "PARMA: A PHY/MAC aware routing metric for ad-hoc wireless networks with multi-rate radios," *IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM 2005)*, Taormina/Giardini Naxos, Italy, Jun. 2005.

S. Ganu, S. Zhao, L. Raju, B. Anepu, I. Seskar, and D. Raychaudhuri, "Architecture and prototyping of an 802.11 based self-organizing hierarchical ad-hoc wireless network (SOHAN)," *Emerging Location Aware Broadband Wireless Ad Hoc Networks*, Springer Publications, 2005.

2006 S. Zhao and D. Raychaudhuri, "On the scalability of hierarchical hybrid wireless networks," *Conference on Information Sciences and Systems (CISS)*, Princeton, NJ, Mar. 2006.

S. Zhao and D. Raychaudhuri, "Policy-based adaptive routing in mobile ad hoc wireless networks," *IEEE Sarnoff 2006 Symposium*, Princeton, NJ, Mar. 2006.

2007 S. Zhao and D. Raychaudhuri, "Multi-tier ad hoc mesh networks with radio forwarding nodes," *IEEE Global Communication Conference (GLOBECOM 2007)*, Washington, DC, Nov. 2007.

S. Zhao and D. Raychaudhuri, "Scalability and performance evaluation of hierarchical hybrid wireless networks," submitted to *IEEE/ACM Transactions on Networking*, 2007.