ADAPTING AD HOC NETWORK CONCEPTS

to LAND MOBILE RADIO SYSTEMS

by

Duncan Scott Sharp
B.Sc.E.E., University of Alberta, 1972

PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF
THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ENGINEERING
in the School
of
Engineering Science

SIMON FRASER UNIVERSITY
December 2002

Approval

Name: Duncan Sharp

Degree: Master of Engineering

Title of project: Adapting Ad Hoc Network Concepts to Land Mobile Radio Systems

Examining Committee:

Dr. Paul Ho
Chair

Dr. Ljiljana Trajkovic
Senior Supervisor

Dr. Joseph Peters
Supervisor

Date Approved: December 9, 2002

ABSTRACT

Ad hoc networks are self-organizing networks of user terminals that form without the need for prior infrastructure. In theory, an ad hoc network could deliver adaptable, robust, and rapidly deployable communication services to meet the needs of public safety related agencies for emergency response and disaster recovery operations. In this project, I investigated the potential to develop a next generation land mobile radio system for public safety communications using ad hoc network architectures and concepts. I applied a four step methodology: (i) identify the communication requirements of public safety agencies in terms of the types of services, traffic characteristics, and quality of service; (ii) explore current technology and research relating to mobile ad hoc networks; (iii) conceptualize a design for a hypothetical next generation network by selecting approaches from the literature that should provide good results against the needs of public safety; and (iv) assess the potential performance of this hypothetical design. Among the many factors considered, the following four had a major influence on the design: (i) the dominant communication need is half duplex multicast voice; (ii) in most instances users have access to a vehicle; (iii) location information is becoming economically available through the Global Positioning System; and (iv) satellite-based mobile communications is available. The hypothetical network I propose is hierarchical with single hop "cluster nets" that are interconnected by a dominating-set based "backbone net". A satellite network tier simplifies routing across large geographic distances and provides a backbone of last resort for sparse networks. For the cluster net media access control, I applied the well known Packet Reservation Multiple Access protocol. The delay performance of this approach was investigated by applying genuine traffic traces to a software model of the cluster net. Before a complete terrestrial-based backbone net can be developed, further work is required; particularly in the area of multi hop routing. A central conclusion is that, although there are major challenges (e.g., spectrum, network self configuration algorithms, routing protocols, standards, and security), enough critical elements are available that prototypes and simple first generation systems can be built.

Table of Contents

List of Tables

List of Figures

Chapter 1
Introduction

Geographically distributed teams involved in real-time operations typically coordinate their work using mobile radio. Messages are often broadcast (multicast) among the team, and when a situation is changing rapidly and real-time reactions between team members becomes critical, voice calling dominates. Typically, these work groups are involved in public safety operations (e.g., police, fire, rescue, and ambulance agencies), utility operations (e.g., electric, telecommunications, and pipeline networks), and other team operations (e.g., construction sites, special event coverage, emergency response, and disaster recovery). A characteristic of many of these applications is that operations tend to be sporadic, with high activity for a period of time in an area, then no activity for an extended period.

Ad hoc networks are self organizing networks of user terminals. The user terminals form the infrastructure. Because these networks are self organizing, they have the potential to be adaptable, robust, and fault tolerant. Because no prior infrastructure is needed, a network can be rapidly deployed. These characteristics are an excellent fit for pubic safety, emergency response, and disaster recovery communication applications.

Currently, the communication needs of public safety related agencies are met by land mobile radio systems. This technology has changed slowly over the past three decades and tends to lag commercial cellular telephone technology by at least a decade. In addition to cellular and ad hoc network technology, other potential enabling technology for future land mobile radio systems includes satellite based mobile systems (e.g., MSAT, Iridium, and Globalstar), 3rd generation cellular developments, adaptive antennas, software defined radios, the Internet, mobile, and pervasive computing, Global Positioning System (GPS) receivers, and high-powered, compact Personal Digital Assistants (PDAs). In particular, the availability of low cost GPS receivers has stimulated its use for navigation support and to support location-aware applications.

This report consolidates the findings of my investigation into the potential for next-generation mobile radio systems to exploit recent advances in mobile ad hoc network technology. Specifically, the objective of the project was to identify evolving technologies, algorithms, protocols, and architectures for mobile ad hoc networks that may be adopted by next generation land mobile radio systems – and, in particular, by mobile radio systems designed for public safety communication applications.

The report follows the general sequence of the project. Chapters 2 and 3 present a needs analysis for public safety mobile communications covering operational requirements and traffic characteristics, respectively. Chapter 4 hypothesizes a system design based on current approaches from the literature and covers radio channel, media access, routing, and network management aspects. Chapter 5 is a preliminary feasibility assessment of the hypothetical system design. Chapter 6 concludes the report with a synopsis of the findings. Appendix A lists selected terminology used in this report. Appendix B and C are primers on ad hoc networks and land mobile radio, respectively.

Interest in radio based ad hoc networks extends back to the 1970's. Development languished somewhat in the 1980's; partly because we lacked the low cost processors and memory needed to run algorithms for ad hoc routing. Since about 1995 interest has been rekindled. The reference list in this report is a small sample of the vast and rapidly growing literature. Throughout the project, I drew heavily on this extensive literature base. References to specific source material are made throughout the text. Although I hesitate to highlight one reference, the "Handbook of Wireless Networks and Mobile Computing", edited by I. Stojmenovic, was an inspiration and a broad source of recent and highly relevant information from many contributors to the field [Stoj02].

This project contributes to the subject by focusing on a specific application (public safety communications) and drawing together a synopsis of the current state-of-the-art as it applies to this application. The synthesis of an example design and the assessment of its potential performance illustrates what is achievable today as well as where more work is needed. In addition, during the course of the project, some potentially useful work on characterizing multicast voice traffic was undertaken. The report could also provide a starting point for research topics or product development studies relating to ad hoc networks.

Chapter 2
Operational Requirements for Public Safety Related Agencies

This chapter and Chapter 3 (on traffic) provide an analysis and characterization of the needs of public safety related agencies for mobile radio communications. These two chapters establish the requirements for the system design described in Chapter 4.

2.1 Missions, Agencies, and Groups of Users

Various organizations are deployed, generally by government, to protect the public against threats to life, safety, and property damage. These agencies include police, fire, rescue, and ambulance. Effective and efficient response to public safety incidents requires the coordination of teams. Without adequate communications, the initial response to a situation, and the ongoing response as the situation changes, is neither effective nor efficient.

Within the overall role of public safety, each agency has a specialized mission to accomplish, usually across a defined jurisdiction or geographic area. This mission, and the context within which it is accomplished, gives rise to different operational support and communication needs. Notwithstanding these differences, in the event of a large scale event (e.g., an earthquake or wild fire that affects a wide area), communication is essential within teams, and between teams. These teams may be in the same or different agencies, or with external groups including utilities (e.g., electric power, gas and telecommunications) and the military. The importance of this communication inter operability for joint operations and mutual aid cannot be overemphasized.

For the purposes of this project, the "users" can be defined as comprising agencies, each with one or more user groups, populated by individual users that are equipped with user terminals. Slightly more formal definitions appear in Table 2.1.

Table 2.1
Operational elements and basic terminology.

Term	Description
Agency	A single administration comprising one or more user groups or teams.
Groups	Teams or work groups (usually corresponding to multicast or talk groups) whose membership may grow or shrink as users come and go and/or subdivide or merge depending on activities.
Talk group	A common voice channel where the above groups can coordinate their operations (functionally similar to a party-line or chat-line).
Users	Individual users that generate and absorb information (i.e., traffic sources and sinks); each user has one or more user terminals.
User terminal	A communication device that includes as a minimum a radio transmitter and receiver (transceiver) and some user interface (e.g., handset/headset, keypad, and displayed).

2.2 General Operational Requirements and Context

In all instances, work forces are mobile. Work is spatially distributed including indoors and outdoors, and may be dangerous from natural or man-made causes. Response must be rapid and in all types of weather. Although users are vehicle borne, they often work outside the vehicle. The work will generally involve handling tools and/or weapons. The form factor of user communication terminals must consider this, and terminals must also be intuitive and easy to use as users may be distracted by the tasks at hand as well as by risk and stress. Not all work will be urgent; there will also be periods, often long periods, when mainly routine work is undertaken.

Although users may be geographically dispersed, command-and-control will usually be centralized. This centralization may be to one or to a hierarchy of points – e.g., a dispatch room or a local command post and/or emergency operations center. Information on the location and disposition of users and equipment is often vital to optimize the deployment of available resources.

Communications security is important. Obviously, protection against unauthorized usage and disruption of communications, including denial of service attacks, is essential. In many instances, security of information to prevent eavesdropping is required (e.g., through encryption).

2.3 Communication Services and Applications

Communication services can be divided into voice and data. In general, voice communication takes priority over data. The types of voice and data communications are identified and briefly described in Table 2.2 and Table 2.3, respectively.

Table 2.2
Typical voice service types.

Application	Description
Group Call	A call on a user selected talk group (channel) that is heard by all users that have that talk group selected.
Agency Call	A call to all users in an Agency – i.e., all talk groups operated by the specific agency.
Zone Call	A call amongst all agency or system users in a defined geographic area.
Inter-agency Call	A group call in a talk group or channel that has been designated for joint operations or mutual aid between agencies.
Individual Call	A call between any two users, or between a user and another network (e.g., the public switched telephone network).
Emergency Call	A call, invoked by pressing the "emergency" call key that results in a channel opening immediately to a monitoring point (e.g., dispatch), the identity (and, if possible, location) of the user is forwarded, and the radio unit microphone opens for a preset time.
Request to Talk	A key on suitably equipped or configured radio units that, when invoked, signals the dispatcher of a request. Once invoked, the request is queued until answered.

Table 2.3
Typical mobile data service types and applications.

Application	Description
Status Messaging	A remote (field) user forwards one of a set of predetermined ("canned") messages to a dispatcher or host application (often relating to user status – e.g., enroute, on-site)
Instant Messaging	Short text exchanges between users that can be held in near real time.
Short Message Service	A short text message that is prepared then sent between users; similar to e-mail but limited to a maximum message size (e.g., 140 characters).
E-mail	The exchange of e-mail messages, including attachments, between users in the field and/or corporate and/or public e-mail systems.
Paging	One way messaging – often to alert or inform a specific user or group of users (e.g., advise of an alarm condition).
Computer Aided Dispatch	A type of host application that ranks, formats and delivers information required by field personnel to perform specific tasks (e.g., respond to incidents, pick up loads).
Fleet Management	A set of applications that involve monitoring, and in some instances, remotely controlling vehicle functions.

Information Query and Retrieval	A host application that may include data base access, web browsing, and operating procedure look-up.
Reporting	Host applications that allow field users to remotely update databases or reports and file information.
Automatic Vehicle Location	Used for asset tracking, navigation support, personnel safety and other applications requiring or benefiting from location awareness.
Mobile Office	Word processing, spreadsheets, remote file transfers, Internet access, and other office type applications.
Image	The transfer of medium and high resolution images; often from a central point to a remote (field) user.
Video	Slow scan or full motion video; often from the remote (field) user to a central dispatch or command point.

Important features for voice and/or data communication include the following:

- Provision of priority levels (e.g., head of queue and/or ruthless pre-emption)
- The ability of the dispatch or command point to preempt a call
- Security – including authentication to prevent unauthorized service access and encryption to prevent unauthorized message interception
- Caller identification (caller ID)
- Recording of communications for debriefing, training and legal purposes
- Inter operability (communication between agencies)
- Voice performance in high noise environments (this is also a terminal issue)
- Ability to guarantee delivery and to know about late delivery (knowing who received a message and when, and who did not receive it)
- In general, status messages are short and high priority and should be combined with position reports
- Provision for remote radio signal strength check capability

All voice applications (Table 2.2) are in widespread use over land mobile radio systems. Many of the data applications (Table 2.3) are in general use over private and public data radio networks, with the exception of video which is inhibited by the lack of spectrum and equipment. Note that group calling is the dominant mode of communication. These calls are half duplex as the call originator (source) activates a push to talk key for the duration of the transmission (talk spurt or call).

2.4 User Terminals

The ultimate success of a communication system is absolutely dependent on how the user interfaces with the service. This interface is the user terminal. Generally, all user terminals should have the following characteristics:

- Rugged and capable of reliable service under heavy-duty industrial and public safety operations
- A minimum number of easy to understand and operate controls
- Easy to read and understand displays (including in low light and bright sunlight conditions)
- Integral radio signal strength indication (RSSI)
- Compact and lightweight

Although ideally in an ad hoc network, every terminal has equal capabilities, this is not an essential condition, nor even desirable, in the context of public safety operations. As a minimum, I expect there will be five types of user terminals, as follows:

- Personal Radios: User terminals carried (or worn) by personnel and forming the primary field user interface.
- Vehicle Radios: Vehicle mounted user terminals that are normally interfaced through the Personal Radio but also need provision for direct user interface.
- Transportable Radios: User terminals that are self contained and intended for unattended operation at locations to supplement connectivity.
- Fixed Radio: User terminals that are intended for fixed location installation to provide access to the network (e.g., at dispatch and operations centers), to interface with other networks (e.g., gateways), and in some instances, to supplement coverage.
- Control-point Radios: User terminals that have added capabilities to support resource management functions covering both field resources (personnel, vehicles, equipment) and network resources (i.e., for dispatch/command purposes and for communication services management). This would be a split or two component device that allows mounting the radio unit separately from the interface (console) unit.

The Personal Radios require long battery life – at least 12 hours, and preferably days, under normal duty cycles. In standby mode, weeks between charging is desirable. Charging procedures and equipment should be simple and foolproof. The Personal Radio needs to be provided with a range of accessories to facilitate different operational modes (e.g., for motorcycle and snow-mobile users). Vehicle Radios should be modular, easy to install and de-install, and relatively energy efficient. Control-point Radios will feature a computer workstation-like interface to support various operational and network management applications (e.g., mapping and dispatch per Table 2.3 above and per Section 2.6 below).

Finally, user terminals should be available in tiers with different levels of packaging and robustness to enable various user group needs to be economically met, including, for example: (i) low tier for lower cost, less demanding user groups, (ii) high tier for heavy-duty use by user groups with bigger budgets, and (iii) intrinsically safe units for use in hazardous environments.

2.5 Quality of Service

In general terms, communication systems must provide timely, accurate and dependable voice and data message delivery. Table 2.4 provides a reference service level specification. These metrics are from the users perspective.

Table 2.4

Representative service level specification.

These metrics apply during expected operational busy periods and within the operations area.

Service level specifications should also stipulate how each metric is measured.

#	Metric	Definition	Performance
1	Network Availability	Probability of having connectivity to workgroup and/or incident or agency control (dispatch) point.	At least 97% per [TIA96] and preferably 99% of locations
2	Service Availability	Probability that the network has sufficient capacity to carry offered traffic.	Network capacity shortfalls are manifested as set-up delay. See metric 4.
3	User Terminal Availability	Probability at any given time that a user terminal with a charged battery is capable of operating to specification.	At least 99.95% (e.g., 3 years mean time between failure and 12 hours mean time to replace).
4	Session Set-up Delay	Probability that set up time will NOT exceed 1 second.	At least 99.9% for priority traffic. At least 97% for non priority.

5	Message End-to-end Delay	Probability that end-to-end delay from source user terminal to destination user terminal will NOT exceed (a) 600 ms for voice and high priority data messages;	At least 99.9% for high priority talk groups and data messages.
		(b) 5 sec for other messages.	At least 97% for other.
6	Voice Fidelity	Probability of having to repeat a message due to poor fidelity (i.e., a distorted and/or incomplete message).	At least 97% and preferably 99% of locations (relates to voice codec and packet loss ratio).
7	Position Accuracy	Probability that a location fix is more than some amount in error or is stale.	Requirements are application specific; to be determined.

It should be noted that packet loss ratio is also an important metric. However, from the user's perspective, packet loss manifests itself as (i) delay for non real-time applications where the transport protocol assures delivery (i.e., through a repeat request procedure), and (ii) fidelity or data integrity for real-time applications (e.g., for voice or high priority, time-sensitive data).

2.6 Network Management

Agencies and/or user groups must be able to configure the services within network limits and within the bounds of fairness to other users. Configuration changes could include, for example, the following:

- Talk group configuration including priorities and membership rules
- User terminal configuration (e.g., interface preferences)
- Remote enable and disabled user terminals
- Remote (over-the-air) reprogramming

The system must have provisions for security and to detect malicious, fraudulent or unauthorized use. At some level, there should be access to various system support and network management functions including:

- Alarm and trouble reports
- Remote diagnostics
- Performance monitoring
- Configuration monitoring, statistics and reports
- Traffic monitoring, statistics and reports

Ideally, and perhaps essentially, user equipment will be available from multiple manufacturers. The equipment must be warranted and supported by the supplier including user documentation, training and repair/replacement capability.

Chapter 3
Traffic Characteristics for Public Safety Networks

This chapter summarizes the methodology and results of analysis undertaken during the project to characterize public safety mobile radio voice traffic. This was necessary because information on this traffic is not well represented in recent research literature. Earlier work is available [Hess81], [Cohe83], [Heer92], [Hess93], and [Ston96], but it is based on non-trunked (conventional access) radio and trunked radio systems operating in message trunking mode (the channel is held for the duration of a message exchange session and not released after each talk spurt).

3.1 Voice Traffic Source

I obtained real traffic data from E-Comm, a large public safety mobile radio network operator. E-Comm serves public safety agencies in the lower mainland of British Columbia with an "EDACS" system from M/A-Com (formally Comnet-Ericsson).

A trunked mobile radio system, such as E-Comm's, consists of a central system controller, repeater sites, fixed user sites (e.g., dispatch locations), and user radios. User radios include both portables and mobiles (i.e., hand-held and vehicle mounted radio units, respectively). The system elements are shown in Figure 3.1.

Each repeater site has a control channel and several working or traffic channels – typically up to 20 or 30 channels. Obviously the number of channels required depends on the traffic load and desired quality of service. Each repeater site typically covers a radius from one to over 25 kilometers, depending on traffic as well as terrain and propagation issues. To provide wider area coverage, additional repeater sites are used.

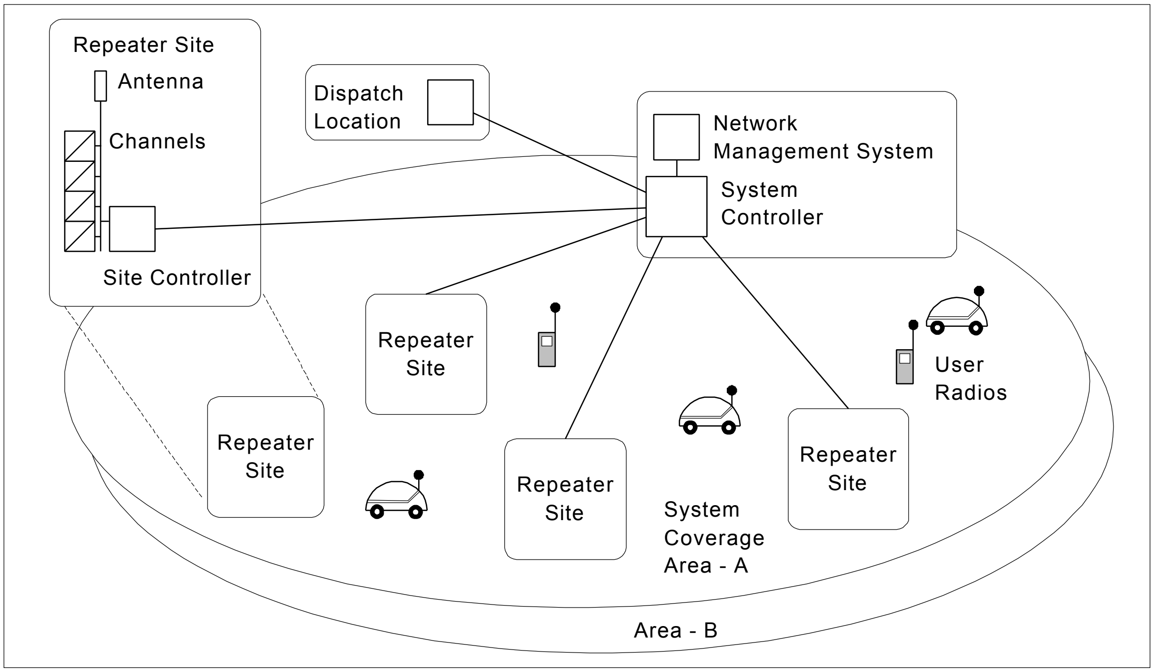

Figure 3.1
Trunked mobile radio system concept diagram.

Similar to other large scale trunked radio systems, the E-Comm system is divided into several coverage subsystems. Although all subsystems form a part of the overall system and are managed by a central controller, each subsystem forms a separate pool of channel resources for traffic purposes.

Users wishing to communicate with others in their work group must have their radios set to the appropriate "talk group". Then, when the user presses the push to talk (PTT) key on their radio, a signal is sent over a control channel to the central controller. The central controller looks up all other units that have selected the same talk group and it verifies that there are channels available on all required repeater sites. The central controller then assigns a channel on each repeater site that is needed for the call and signals every radio in the talk group to tune to the appropriate repeater frequency for the call.

Assigning and setting up a channel takes typically 500 milliseconds, after which the caller can begin talking. The call is received by all radios on the talk group that are within signal coverage. When the PTT key is released, the central controller releases the channel at each repeater site. If no channels are available, then the PTT request is put in a queue and served on a first-in-first-out

(FIFO) basis within each priority level. In some systems, including E-Comm's, recent "talkers" may be given priority over new calls to insure low delay to existing call sequences during congestion[1].

The system supports multiple agencies (currently E-Comm serves approximately 14 separate agencies, some of which operate across multiple geographic jurisdictions). Each user agency has one or more talk groups. A specific group may range from a few units up to several hundred. Large agencies usually have several talk groups; for example, different talk groups may be assigned for different functions, such as: dispatch, queries and information retrieval, chat, and for specific incidents and operations.

The system will support multiple priorities (up to eight). The highest priority is reserved for emergencies (which are invoked by pressing an emergency key). Among public safety groups, priorities may be assigned to give urgent and important dispatch and command talk group traffic precedence over other traffic. The E-Comm system is also capable of being partitioned so that specific channels are reserved for designated user groups. This practice is discouraged, however, as breaking the channel resources into multiple traffic pools lowers overall efficiency.

The system controller that assigns channels and manages the system, is a software programmed real-time central processing unit. The controller maintains a call activity detail file that is a chronological list of all call related system events. The system also provides a summary report that integrates and summarizes the call activity events for specific systems (coverage areas), or users, or talk groups. This data is initially written to a hard disk, then later transferred to magnetic tape for backup and storage. The data fields in these files that are most relevant to this project are summarized in Table 3.1

[1] Recent talkers can also be ensured access to the channel by using a "hang-time" – i.e., not releasing the repeater channels at the end of each PTT but holding the channel for some predetermined time in case someone else in the talk group wants to transmit. Obviously, the use of a hang-time is less efficient in terms of traffic handling.

Table 3.1
Relevant traffic data fields from the call activity detail and summary report files.

Call Activity Detail File	
Time	date time stamp (0.01 second resolution)
Type	call type, such as group, emergency, etc.
State	a state transition such as assign or drop a channel
Caller	the identification of the radio unit originating the call
Callee	the talk group the call is destined to
Duration	call holding time in seconds (0.01 second resolution)

Summary Report File (60 minute integration time)	
Start at	date time stamp of period start
Active talk groups	number of talk groups active during the period
PTTs	number of call attempts
Granted calls	number of calls carried
Granted duration	call holding time sum total
Queued calls	number of call attempts queued
Queue duration	queue delay time sum total
Max queue depth	maximum queue depth
Max call delay	maximum call queue delay

The call activity detail data was used to extract traffic traces comprising the call arrival times and call duration in seconds. This processing also removed all confidential information (e.g., relating to, and identifying, specific user groups).

The system files were ported from the network system manager console to text files on disks (for security reasons the network management system does not connect to any external networks). The text files were manipulated using conversion and sort functionality built into Excel (the widely used spreadsheet program from Microsoft). Files exceeding 65,000 records were partitioned in a text editor and then reassembled after processing. The extracted traces were compiled as spreadsheet tables for analysis (including porting into S-plus which smoothly interworks with Excel).

These traces provided detailed insight into the traffic processes. Deciding which system and periods to extract was guided by the desire to avoid any traffic shaping by the system – e.g., as would occur during periods of congestion when the system is queuing (thus smoothing the peaks). Two complete days of traffic were examined from the Vancouver system. Table 3.2 summarizes key characteristics.

Table 3.2
Summary of Vancouver system user count and channel capacity.

Parameter	Amount
Estimated Number of Users	
Police	1527
Fire	223
Ambulance	167
Transit	244
Total	2161
Number of Traffic Channels	12

The summary reports were used to identify any daily and weekly traffic patterns for the Vancouver system over a seven week period in the September, October, and November 2001 time frame. The number of agencies and users on the system did not change during the study period. Therefore no long term trends were expected (nor detected).

3.2 Traffic Analysis

In order to focus on traffic generation and to control the scope of the project, only traffic within a single coverage system was considered. This removes the issue of mobility patterns and each coverage area can be treated as a single traffic pool.

By removing mobility, the traffic is reduced to a function of two random processes – a call arrival process and a call size (duration) process. For voice calling, the Erlang is a convenient traffic unit, and in our case an Erlang simply represents the number of simultaneous voice calls in progress. For example, if nine channels are in use at a given time, nine Erlangs of traffic are being carried. Because calls arrive and depart randomly, the traffic is also random.

Traffic is measured across some integration period (e.g., for mobile voice 60 minutes is common). This measurement is a statistic and to be statistically valid the process must be sufficiently stationary during the integration period.

Traffic (A) in Erlangs is the sum of all call durations (Tc) across some integration period (Tp) divided by the period (Tp and Tc are in the same units so traffic in Erlangs is dimensionless). Taking the average call duration ($avgTc$), then if we have some number of calls (Nc) during the period (Tp), traffic is given by:

$A = (Nc \times avgTc) / Tp$

The average call inter-arrival time (*avgTi*) is given by:

$avgTi = Tp / Nc$

With a bit of rearranging,

$A = avgTc / avgTi$

Note that inverting *Tc* and *Ti*, gives the departure and arrival rates, respectively; leaving us with the well known expression that traffic is the ratio of the arrival rate over the departure rate.

The traffic was analyzed in a three step process, as follows:

- Step 1: Analysis of the longer-term traffic patterns looking for weekly and daily cycles.
- Step 2: Analysis of call holding times from detailed traffic traces.
- Step 3: Analysis of call inter-arrival times from detailed traffic traces.

Traces were extracted using simple logic statements and re-compiled using sort functions built into Excel. The validity of the extraction process (used to obtain traces from the detailed data) was confirmed by comparison against the system summary reports. Statistics were generally determined using Excel spreadsheets with results verified by running the data through S-Plus.

3.3 Overall Voice Traffic Characteristics

Communication traffic usually exhibits daily, weekly, and seasonal patterns that reflect the underlying activities of the users. Hourly traffic summaries of the Vancouver system for a 51 day period , starting on 2001 09 24 were analyzed. Figure 3.2 shows the hourly traffic for the whole period. Each week is demarked by the relative traffic lulls on Sundays. Figure 3.3 and 3.4 show summaries for two days – i.e., Wednesday and Friday.

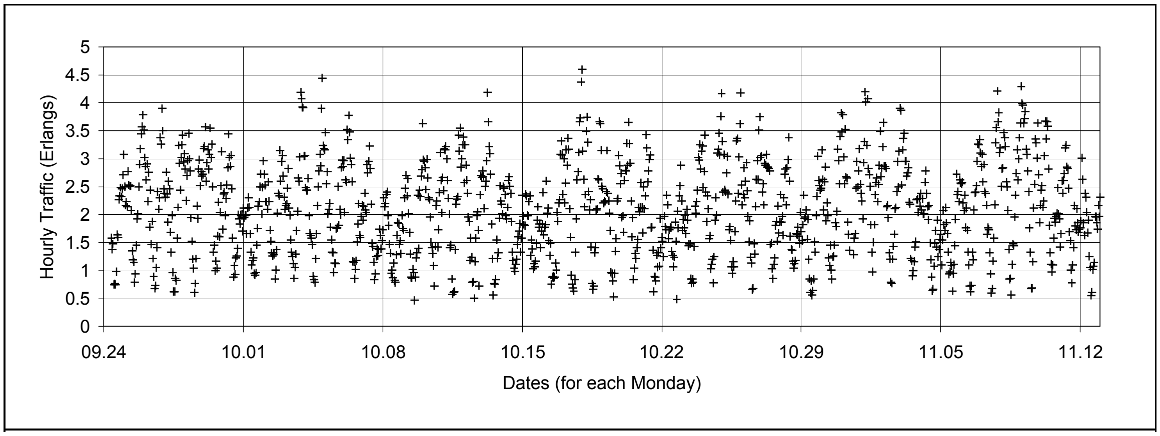

Figure 3.2
Average hourly traffic.
Average hourly traffic in Erlangs (y-axis) through the 51 day study period (x-axis), starting on September 24 (shown as 09.24). Note that the vertical lines divide the overall period into weeks.

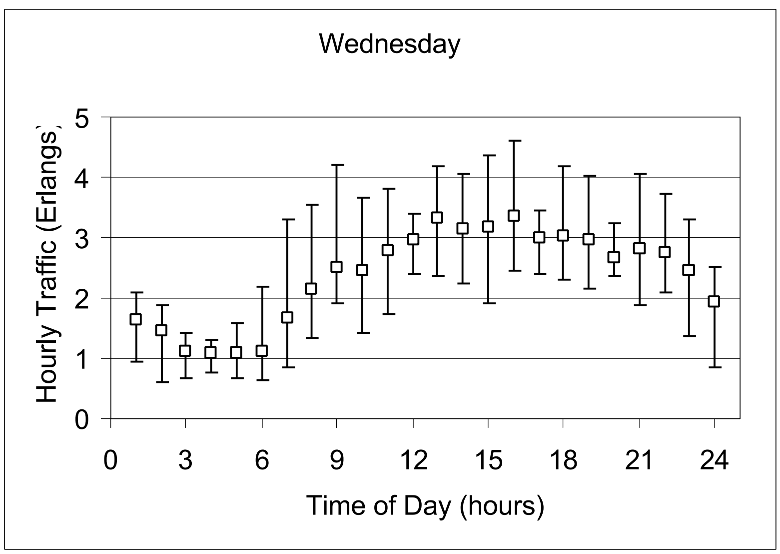

Figure 3.3
Hourly traffic averages for all Wednesday during the 51 day study period (from Figure 3.2). The square marker is the average traffic during the hour and the bars depict maximums and minimums for the period.

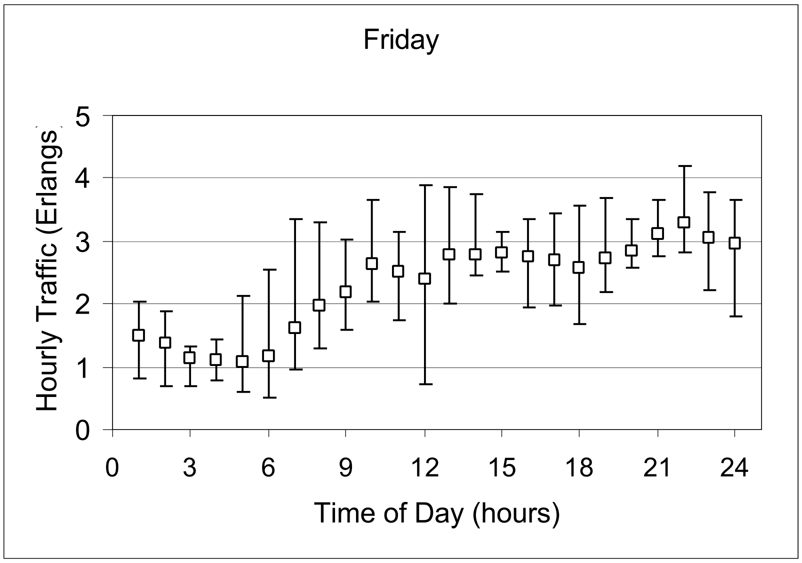

Figure 3.4
Hourly traffic averages for all Fridays during the 51 day study period (from Figure 3.2). The square marker is the average traffic during the hour and the bars depict maximums and minimums for the period.

Examining the graphs, we can see relatively high traffic corresponding to expected periods of high user activity – e.g., Friday night (as the police are busy with the party scene). We also note, from the wide swings between the minimums and maximums, that the actual day and hour of the busiest period is variable and difficult to predict.

In Figure 3.5, the daily averages, average peak hours, and highest peak hours are plotted for each day of the week. From this we see that Wednesday and Thursday are the highest traffic days. The data for Figure 3.5 and other related results are shown in Table 3.3. Of some interest is the dispersion metric given here as the ratio of the standard deviation to the average. This is sometimes used as a quick measure of the degree of variability or goodness of the average as an indicator of behavior. Although the sample size (7-8 days) is too small for statistical validity, dispersion ranges from about 0.2 to 0.35 were observed, with the peak days having the highest dispersion.

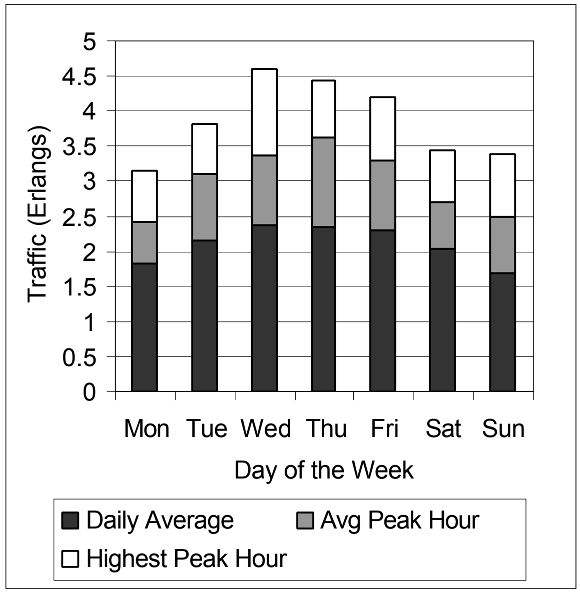

Figure 3.5
Summary of daily traffic for all days of the week in the 51 day study period (from Figure 3.2).

Table 3.3
Average daily, average peak hour, and highest peak hour system traffic (Erlangs).

Parameter		Mon	Tue	Wed	Thu	Fri	Sat	Sun
Daily Average	(A)	1.81	2.15	2.36	2.34	2.30	2.03	1.68
StDev of Hourly Averages	(S)	0.50	0.71	0.78	0.80	0.72	0.50	0.35
Average Peak Hour	(P)	2.42	3.12	3.37	3.63	3.30	2.70	2.50
Dispersion as	(S/A)	0.28	0.33	0.33	0.34	0.31	0.25	0.21
Peak to Average	(P/A)	1.34	1.45	1.42	1.55	1.44	1.33	1.48
No. of S to Peak *	(P/S)	3.04	3.50	3.65	3.94	3.69	3.37	4.01
Maximum Peak	(Pm)	3.15	3.82	4.60	4.44	4.18	3.44	3.38
Max Peak to Average	(Pm/A)	1.74	1.77	1.95	1.89	1.82	1.69	2.01
* from P/S = A + (P - A) / S								

Table 3.4
Per unit traffic summary (from Table 3.3).

	Per Unit Traffic in milli-Erlangs						
Parameter	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Average Daily Traffic per Unit	1.482	1.762	1.932	1.917	1.879	1.664	1.378
Average BH Traffic per Unit	1.983	2.548	2.752	2.966	2.699	2.208	2.043

Table 3.4 converts the total traffic to the average traffic per user unit. The number of units is an approximation based on the estimated number of radios registered to the system for the coverage area – it is not the number of units that are active at any given time. The average daily traffic per unit on the busy days is nearly 2 milli-Erlangs and, from Figure 3.3, the maximum peak to average is approximately 2 on those days, indicating busy hour traffic of about 4 milli-Erlangs per unit. If we consider that during busy periods up to two-thirds of all units may be active, then on-duty units will be generating about 6 milli-Erlangs on average. This is consistent with various other estimates including the 6 to 7 milli-Erlangs from Hess [Hess93].

Figure 3.6 and 3.7 are frequency and cumulative distribution plots of the hourly traffic from the data set (i.e., Figure 3.2). For reference, Figure 3.7 includes a theoretical Gaussian plot generated using a probability plot that assumed a Gaussian distribution (the fitted line correlation was 0.995). Although the fit is not perfect, it is of interest because for low probabilities (e.g., below 10%) the Gaussian curve appears to be a reasonable bound.

There is some validity in considering the talk group to be the traffic source (instead of individual user radio units). As talk groups become heavily loaded, users may delay a call waiting for a gap in the traffic. Consequently limiting the number of talk groups and driving up talk group loading, is a way of shifting delay from the system to the talk group [Hoa91]. Thus the distribution of the number of talk groups is a potential parameter of interest. Figure 3.8 and 3.9 are frequency and cumulative distribution plots of the number of talk groups active during each hour of the data set. From these figures we note that, in the Vancouver system, the number of active talk groups range from 10 to 54 with 80% being between 20 and 35; and the average, at about 26.

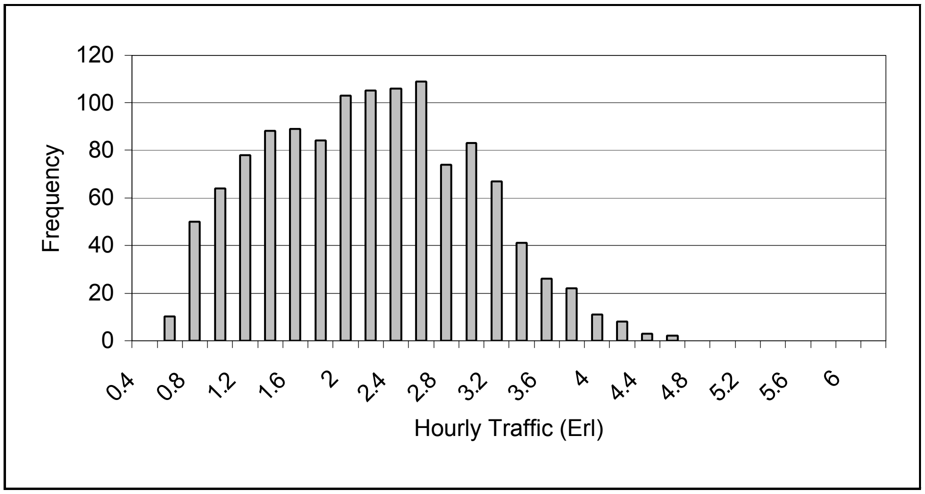

Figure 3.6
Frequency chart (histogram) of the hourly traffic in Erlangs across the 51 day study period (from Figure 3.2).

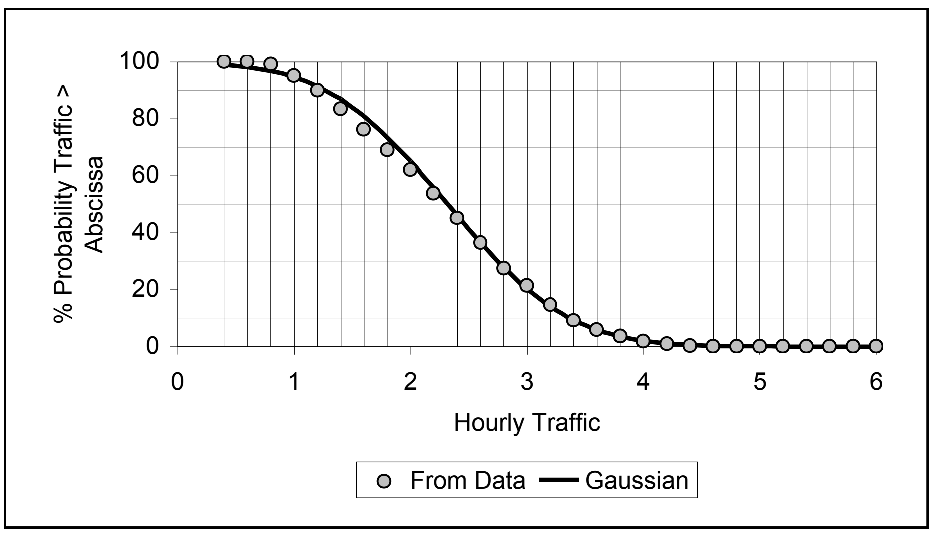

Figure 3.7
Cumulative distribution of hourly traffic across the 51 day study period (from Figure 3.6).

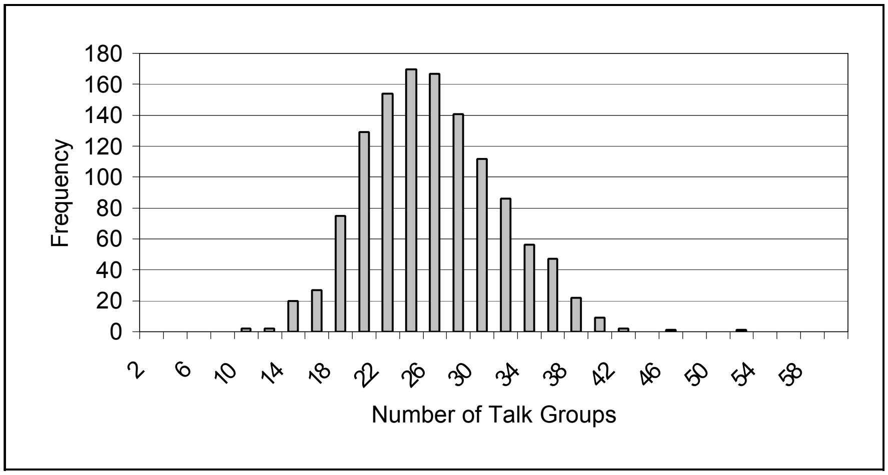

Figure 3.8
Frequency chart (histogram) of the number of talk groups (averaged by hour across the 51 day study period from Figure 3.2).

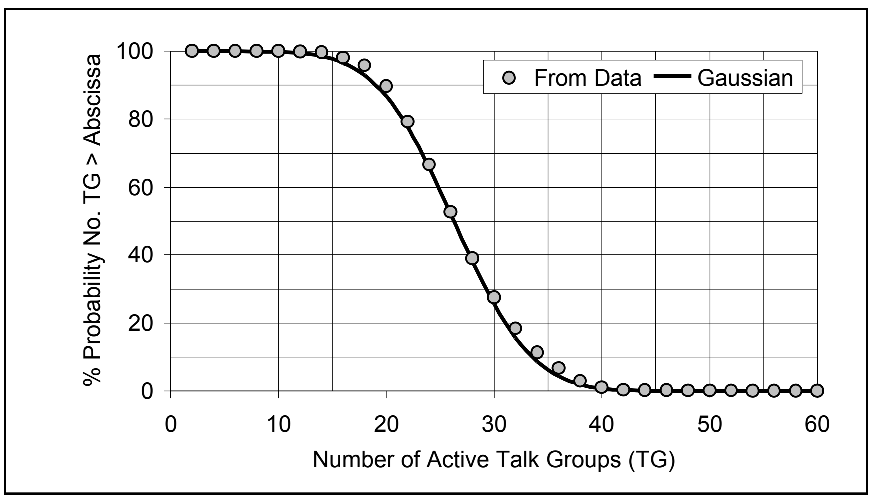

Figure 3.9
Cumulative distribution of the number of talk groups (from Figure 3.8).

3.4 Detailed Analysis of Call Holding Times

The distribution of call holding times (the departure process) was analyzed by looking at lag plots, frequency charts (histograms), cumulative distributions, and autocorrelation. The analysis included examination across time periods from one hour to several days of system traffic (all talk groups), selected talk group sets, individual talk groups, and individual users. Figures 3.10 through 3.12 are representative of the results. No significant anomalies from these figures were detected in the holding times of the other voice traffic traces examined.

The shotgun blast, or lack of regularity or pattern, shown on the lag plot (Figure 3.10) is a good indication of an underlying random process (below 0.1 seconds there is evidence of the 0.01 second resolution of the measurement process). The distributions in Figure 3.11 exhibit the classic exponential distribution seen in other communication traffic processes (notably voice telephone call holding times). In particular, note the straight line shape of the cumulative distribution after about three seconds (the probability axis is log scaled) – a good indication of an exponential distribution. Finally, the low values and irregularity of the autocorrelation plot in Figure 3.12 indicates an absence of any significant time dependencies in the distribution. We can conclude that the voice call holding times follow a reasonably well-behaved exponential distribution and, in fact, this process appears relatively time invariant (stationary) across all time periods.

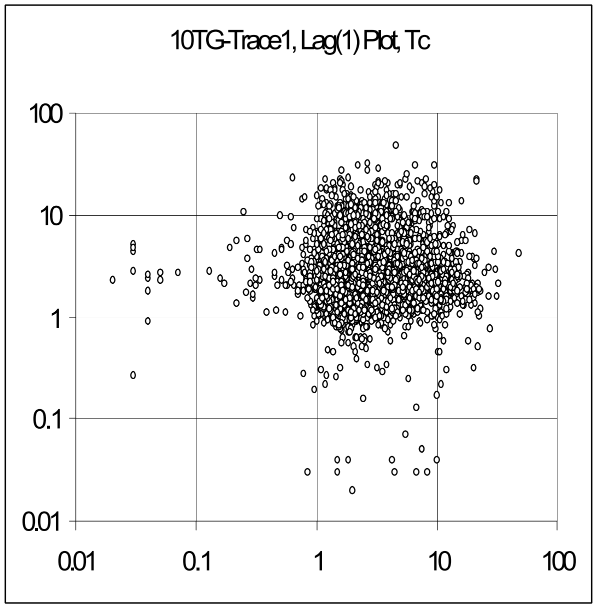

Figure 3.10
Lag plot, call holding time (Tc) in seconds, 10 talk groups, genuine traffic traces, 1 day.

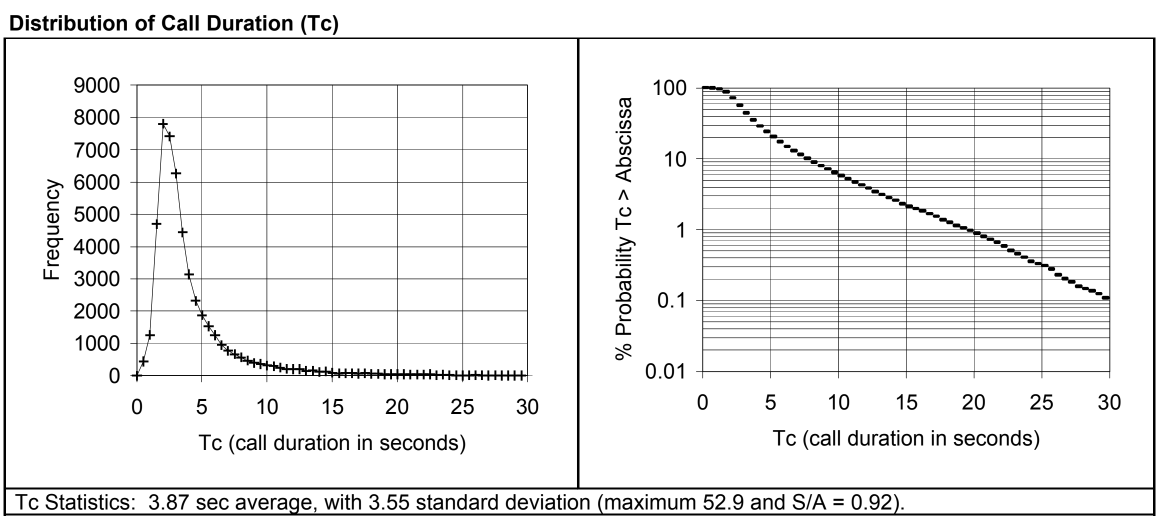

Figure 3.11
Distribution plots, call holding time (Tc), all Vancouver system, genuine traffic traces, 1 day.

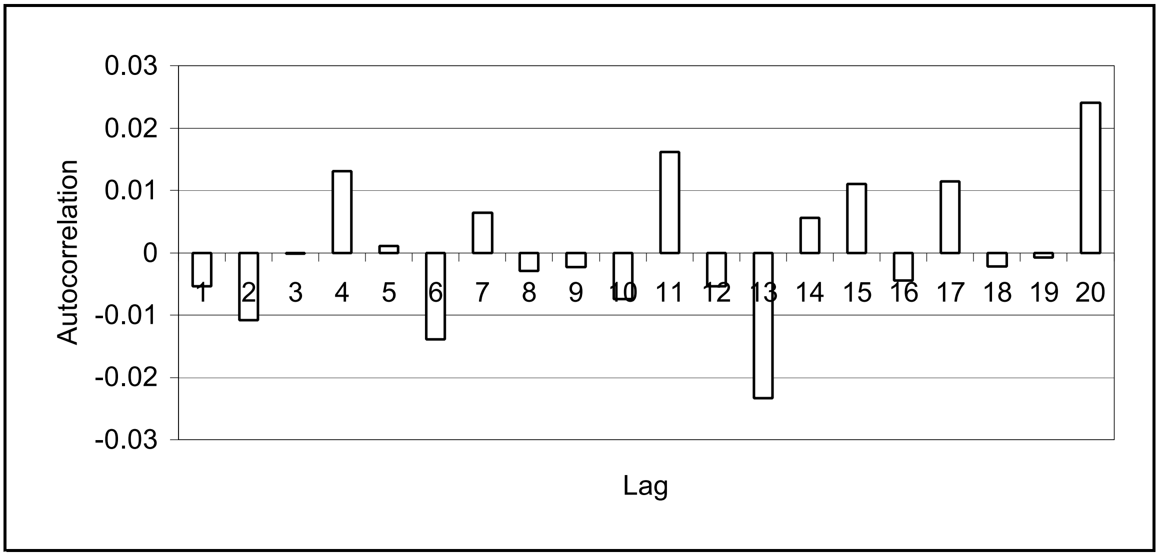

Figure 3.12
Autocorrelation plot, call holding time (Tc), 10 talk groups, genuine traffic traces, 10 hours.

3.5 Detailed Analysis of Call Inter-arrival Times

The distribution of call inter-arrival times (the arrival process) was analyzed by looking at lag plots, frequency charts, cumulative distributions, autocorrelation, and variance-time plots.

Figures 3.13 through 3.15 are lag plots of a single talk group to demonstrate that individually most (but not all) talk groups have session behavior – i.e., a series of calls form a sort of conversation. Specifically, the lag plot in Figure 3.13 lacks the shotgun blast signature of a single underlying random process. Figure 3.14 and 3.15 show that when we separate sessions, the call inter-arrival time (within a session) looks like a random process and session inter-arrival times also look likes a random process. Obviously, as talk groups are aggregated onto a communication system or traffic pool, the inter-session arrival times are lost.

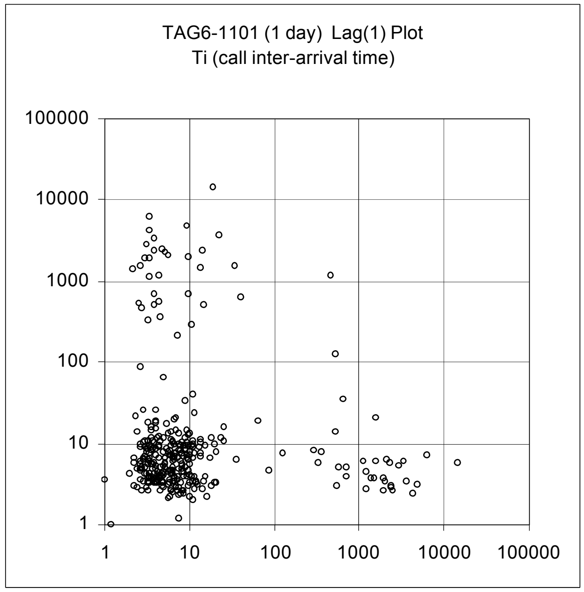

Figure 3.13
Lag plot, call inter-arrival time (Ti) in seconds, chat talk group, genuine traffic trace, 1 day.

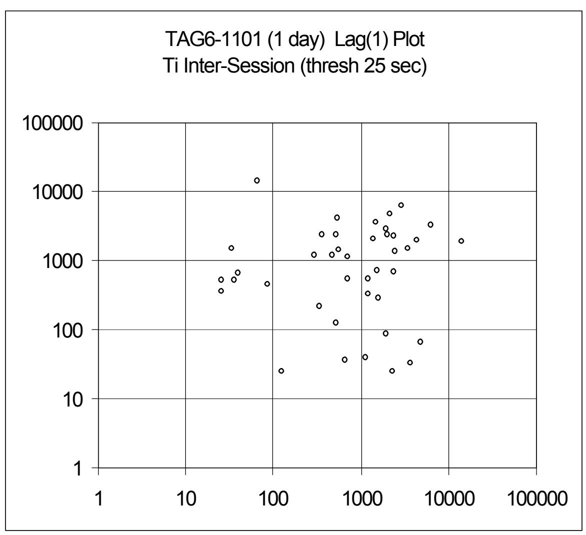

Figure 3.14
Lag plot, "inter-session" call inter-arrival time (Ti) in seconds, chat talk group, genuine traffic trace, 1 day.

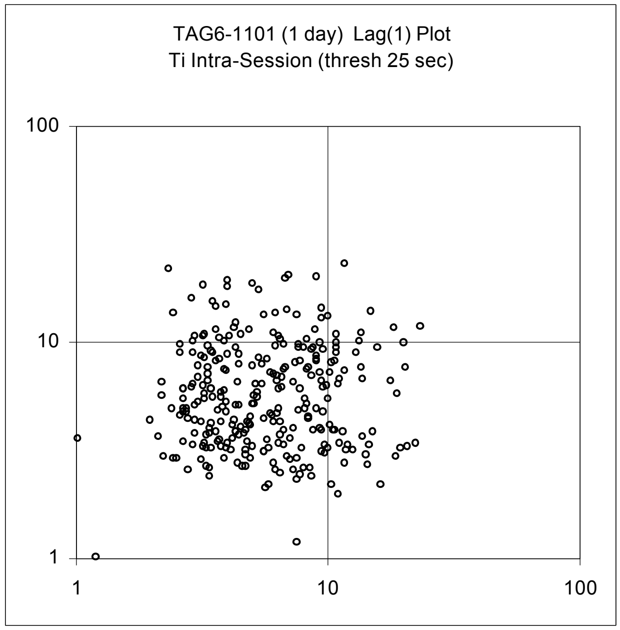

Figure 3.15
Lag plot, "intra-session" call inter-arrival time (Ti) in seconds, chat talk group, genuine traffic trace, 1 day.

Figures 3.16 through 3.19 are representative results of the analysis for system traffic (aggregated talk groups). The analysis here uses a one hour period; because, from Section 3.3, any longer period risks being non stationary. The lag plot in Figure 3.16 is not too informative (note the 0.01 measurement resolution artifacts below 0.1 seconds). The shape of the distribution plots in Figure 3.17 signal that we may not have an exponential distribution – which means that the process is not exhibiting memoryless behavior. This signal is confirmed by Figure 3.18, where the autocorrelation plot, although decaying, does not decay quickly. The variance-time plot in Figure 3.19 provides further evidence of a long-range dependency (i.e., H, the Hurst parameter, lies between 0.5 and 1) [Lela94]. I have some concern that the apparent long-range dependency may be indicating the process is not stationary across the one hour study period. Future analysis should consider attempting to de-trend the data to remove the time dependent variations and/or look at shorter study periods (such as 10 or 15 minutes). From Section 3.3, another issue related to the traffic period is how to determine an appropriate busy period to use when dimensioning the network. This important topic for network designers requires further study.

In any event, an important project conclusion is that, because, the inter-arrival time has evidence of long range dependencies, performance evaluation should use, or be supplemented by, experiments or simulations with genuine traffic – and not rely solely on analysis [Jian01].

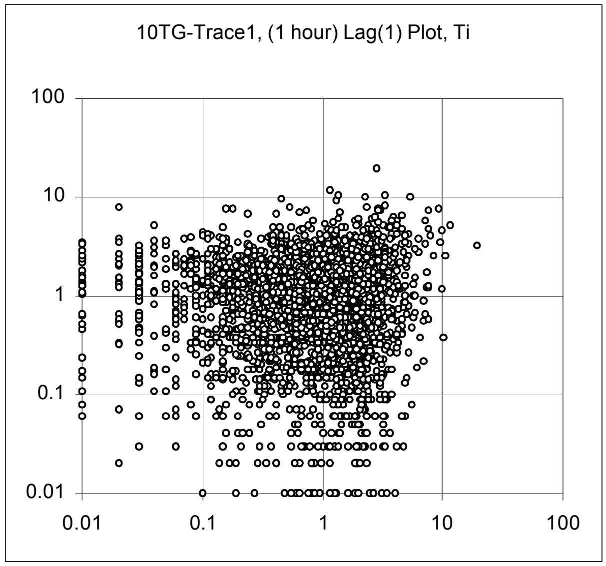

Figure 3.16
Lag plot, call inter-arrival (Tc) in seconds, 10 talk groups, genuine traffic traces, 1 hour.

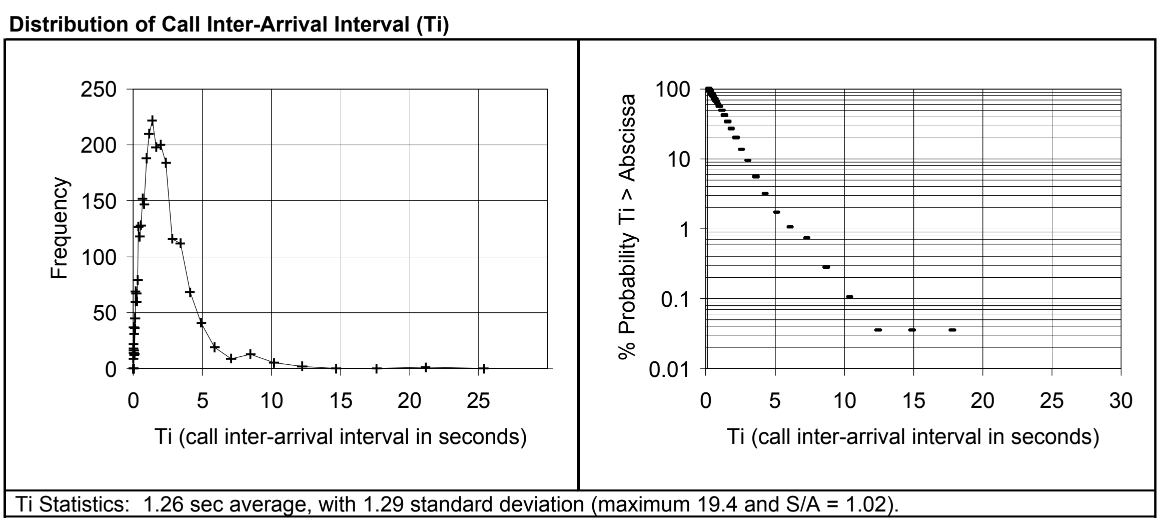

Figure 3.17
Distribution plots, call inter-arrival time (Ti), 10 talk groups, genuine traffic traces, 1 hour.

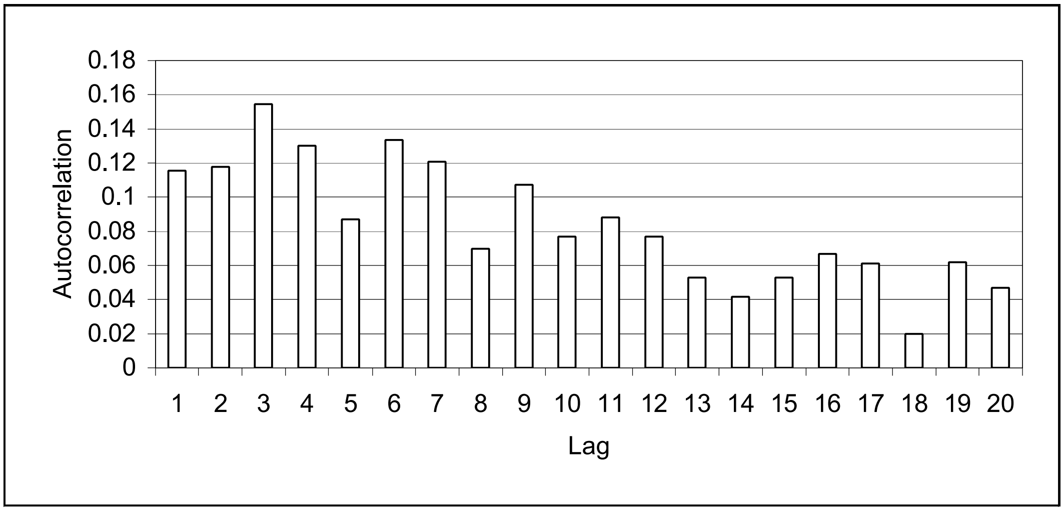

Figure 3.18
Autocorrelation plot, call inter-arrival time (Ti), 10 talk groups, genuine traffic traces, 1 hour.

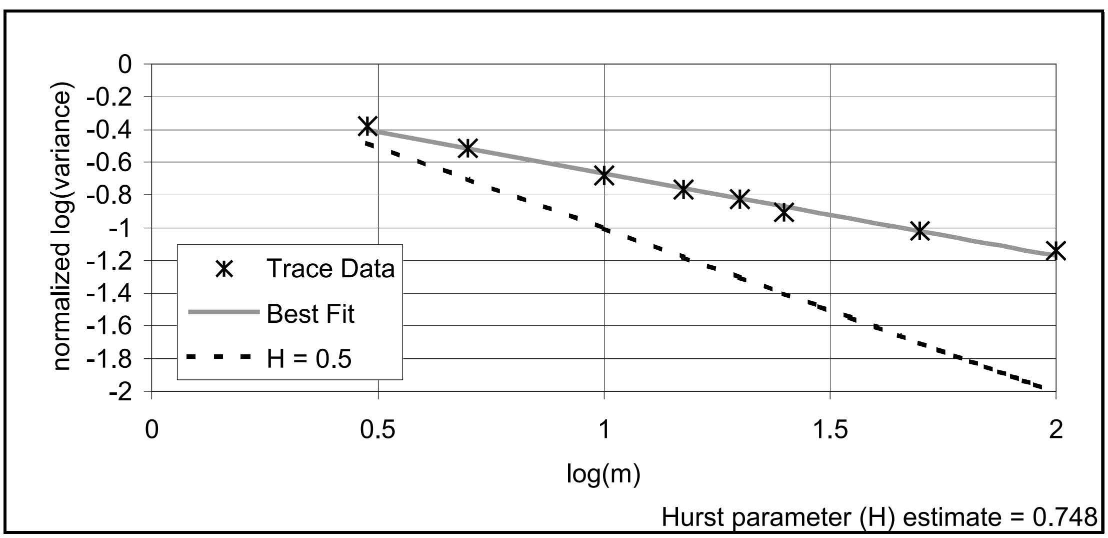

Figure 3.19
Variance-time plot, call inter-arrival time (Ti), 10 talk groups, genuine traffic traces, 1 hour.

3.6 Geographic Distribution and Mobility

Mobility patterns and the geographic distribution of traffic were outside the scope of this project. Characterizing these distributions must be completed in order to undertake complete design studies for wide area systems – i.e., for design of a multi hop ad hoc backbone network that is suited to the public safety communications context (as discussed in Section 4.6). Therefore, this is identified as a critical area of future work.

3.7 Data Traffic

Data traffic characterization was also outside the scope of this project. There is information on data traffic available in the literature. For example, D. Staehle et al. provide source traffic models for wireless applications in [Stae00]. Although perhaps dated, estimates of data traffic volume for public safety is available from Stone in [Ston96]. In addition, there are genuine traces for packet data available in the public domain. It would, however, be useful to obtain genuine traces from an operational public safety system (for those applications that are deployed); particularly as noted in Chapter 2, there are a variety of data applications with different bit rate, priority, and quality of service requirements. Data traffic traces that include automatic vehicle location (AVL) information would be of great interest and support the geographic distribution study work noted in Section 3.6. Therefore, data traffic characterization is identified as future work.

Chapter 4
Conceptual Design of a Hypothetical Ad Hoc Network

This chapter describes a hypothetical system design concept that I created to illustrate a possible future mobile radio system. This design addresses the communication needs of public safety related agencies as described in Chapters 2 and 3, and adopts elements of an ad hoc network architecture. The elements of the design are taken from the literature; there combination into an overall system design is original. In the design process, I attempted to make reasonable choices between conflicting requirements (e.g., simplicity versus efficiency) and to select schemes that are mutually compatible. Obviously, with more research, more in-depth trade-off studies, and prototyping, superior designs will be possible. And, of course, this is to the extent that multi hop ad hoc networks and good performance are realizable – a hypothesis that remains to be proven.

4.1 System Elements, Basic Organization, and Topology

As shown in Figure 4.1, users are organized into groups. Usually a group is formed based on some common task or purpose and while membership may overlap (one user may belong to more than one group at a time), usually membership in one group is dominant at any given time. The users are also associated with an agency that provides overall management and administration of the users. Each agency may have one or more groups with membership flexible and changing across time. The number of groups and properties (e.g., priority level of a talk group) is also flexible but less changeable (i.e., changes only infrequently). The system supports multiple agencies. Agencies who may need to communicate should be on the system (e.g., for joint operations and mutual aid). In addition, other agencies should be encouraged to participate in order to increase the node density and therefore increase the probability that a connected network graph can be established and maintained. In particular, agencies with non coincident busy periods can increase overall network utilization (increasing system efficiency). The benefit of increasing participation is discernible from Figure 4.2 (e.g., the group C members are relatively spread out and will rely on relay by group A and B members to maintain connectivity).

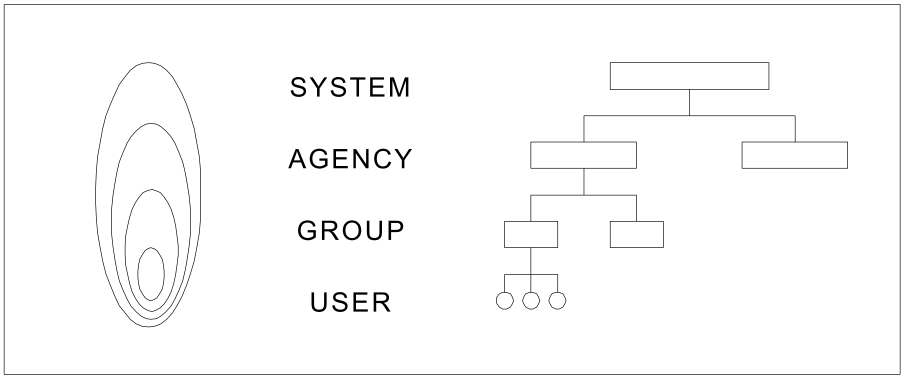

Figure 4.1
Global view of the system elements and how they form into a logical hierarchy.
Groups are teams or work groups, and communication within the group is within a "talk group" (i.e., multicast groups).

Figure 4.2 shows an example spatial distribution of users and also shows that some groups will tend to be clustered (into one or, in some instances, multiple clusters) and often the group will be monitored, and perhaps controlled, from a remote point (dispatch or command center).

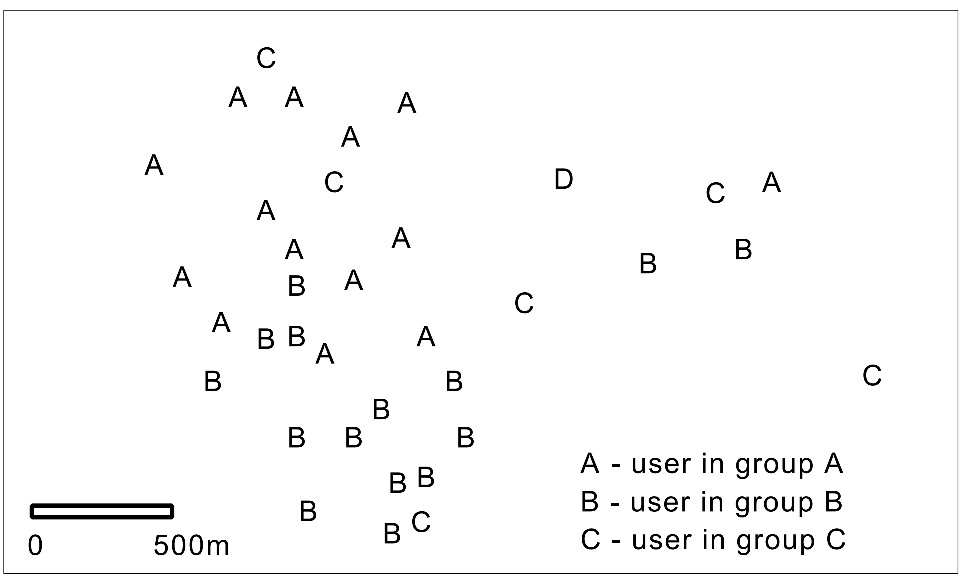

Figure 4.2
Example of user and group spatial distribution.

The system elements are shown in Figure 4.3. The basic physical element is the user terminal which consists of a user interface and the necessary computer and radio hardware. Each user will have one or more user terminals and user terminals may be personal (portable), vehicle mounted, transportable or fixed. Some user terminals may be unattended and interface with servers or act as repeaters (to extend coverage) or gateways (that link to other networks, including fixed networks).

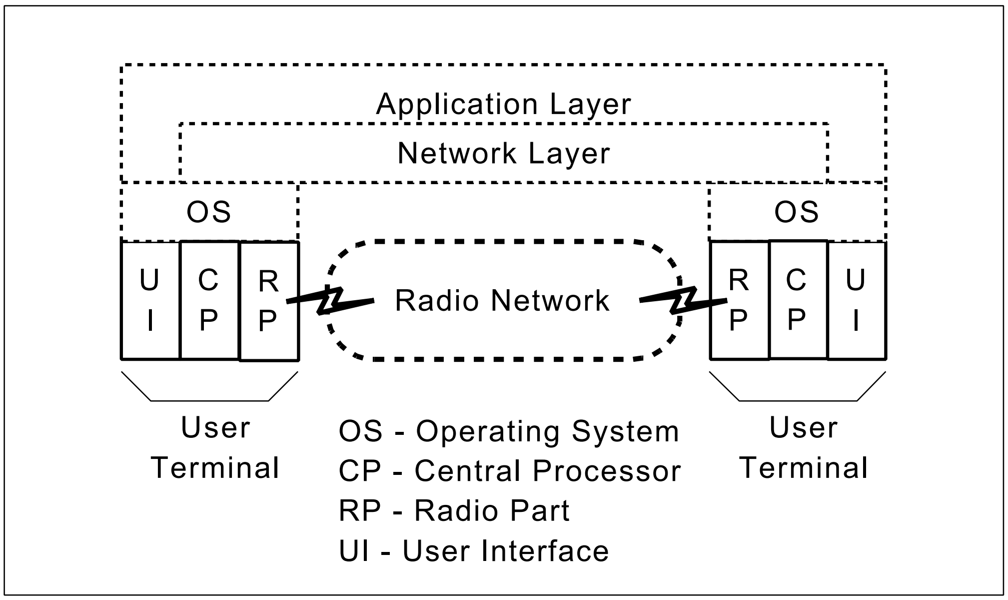

Figure 4.3
Overview of major system elements.
Note that, user terminals may be unattended and function as servers, repeaters (to extend coverage), and gateways (to other networks).

The availability of resources, especially power, for user terminals will vary depending on whether the terminal (radio) is personal, vehicular, transportable or fixed. This provides a natural hierarchy that can be used to scale the network in terms of the number of nodes as well as the distances to be spanned. Figure 4.4 shows a three layer network hierarchy comprising an access layer and two backbone layers. The access network layer deals with localized clusters; for example, linking pedestrian to pedestrian and/or pedestrian to vehicle. The terrestrial backbone network generally relies on vehicle to vehicle or vehicle to transportable/fixed terminals. The satellite network provides a backbone of last resort and is available to all users with sufficient sky view.

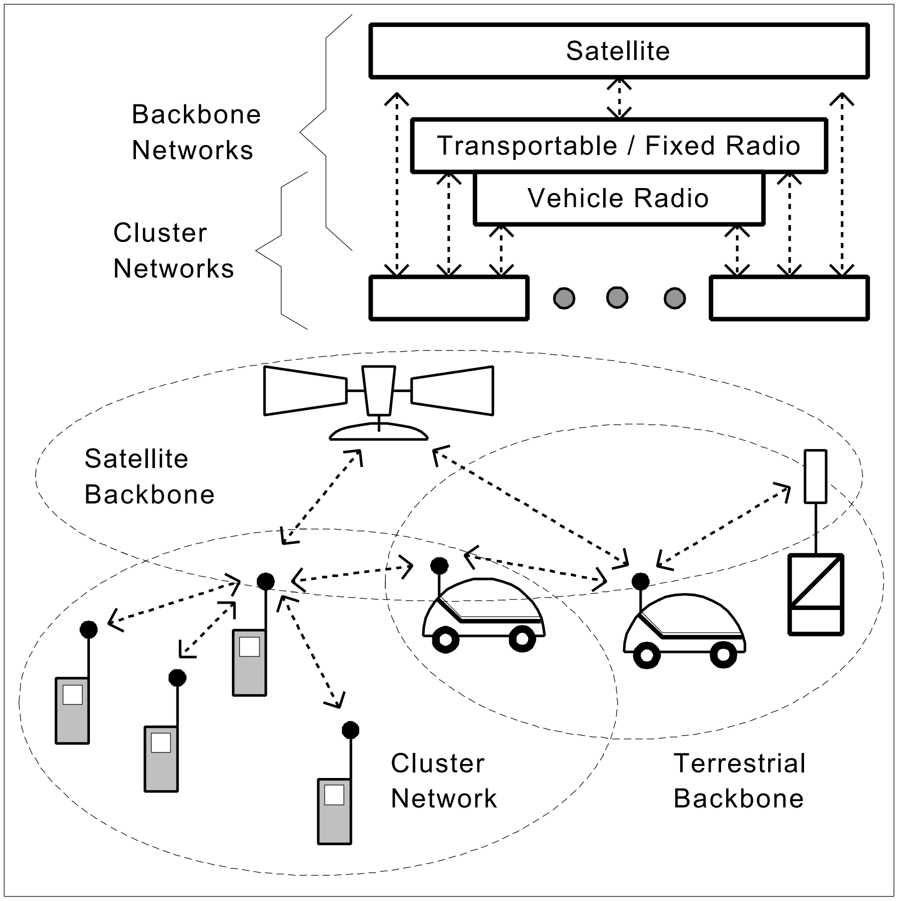

Figure 4.4
Network hierarchy for scalability in size (number of nodes) and span (physical distance across).

A packet based network is proposed because it enables a fully distributed and self organizing architecture that is capable of forming a stable network with no pre existing infrastructure and in the presence of topology changes. Packet size and framing must consider the priority needs of the voice traffic and is discussed later.

As explained later, in Section 4.6, a dominating-set based terrestrial topology is proposed [Stoj02] pp. 425. This approach features clusters with member nodes one hop from the cluster head (or gateway) node. Each cluster can form a simple and efficient sub net (cluster net) with some common functions deferred to the cluster head. The dominating-set based approach also simplifies the routing process as routing then only applies to the smaller subnet that is formed by the connected dominating-set. The dominating-set of cluster heads and relay nodes provide an inter cluster backbone.

4.2 Radio Frequency Spectrum

A fundamental criteria and constraint is the availability of spectrum. The frequency band will set the propagation characteristics and, to a certain extent, the cost of the RF components. The amount of spectrum and the spectral efficiency (bits per Hertz) of current state of the art radios, in terms of coding and modulation schemes, will set the overall network capacity limits.

The spectrum must be designated for shared use by self organizing networks that observe suitable spectrum sharing etiquette. The use of the 2.4 GHz and 5 GHz licence-exempt bands are examples of such bands and etiquette. These license-exempt bands are generally unsuitable for public safety applications, since they may be congested by non public safety usage.

The recent allocation by the FCC of 50 MHz of spectrum at 4,900 MHz [FCC02] is of interest (4940-4990 MHz), as is the 24 MHz in the 700 MHz band (2 x 12 MHz sub bands of which 2.6 MHz is designated for inter operability and 6 MHz is reserved). I have assumed that 30 MHz of the 50 MHz in the 4,900 MHz band can be split into two each 15 MHz sub bands – one sub band for the clusters and one sub band for the inter-cluster backbone. This allocation is consistent with current FCC proposals that suggest the following three-way split: (i) personnel/vehicular area networks, (ii) wireless LANs for on-scene communications, and (iii) fixed hot spot access points [FCC02]. In the 700 MHz band, the 6 MHz of reserved spectrum is of interest. I have assumed for the purposes of this project that it can be used.

Ultra-wide band (UWB) radios[2] have been suggested for various applications including a precision location radio for voice communication and locating firefighters in buildings [Siwi02] and dense, short-range ad hoc networks (e.g., sensor networks). Various experimental links have been tested in the 1-2 GHz bands but the effects of interference into existing systems is still under investigation. This may be a technology to watch; however, at the present stage of development it is far from clear what future role it can play for longer range and sparser networks, as needed by public safety groups.

[2] The FCC has tentatively defined ultra-wide band systems as those with bandwidth (i) over 1.5 GHz or (ii) more than a 25% of the center frequency.

Finally, note that to meet present needs, public safety agencies currently use spectrum at 150 MHz and 800 MHz (and to a limited extent, at 400 MHz). Existing and follow-on infrastructure based systems will continue to use these bands. Thu,s the development and deployment of ad hoc networks systems for public safety would be adjuncts and supplement these systems.

4.3 User Terminals

The basic user terminal will be a portable unit carried by the individual – referred to here as a Personal Radio. Each user terminal will necessarily comprise a data processing unit with memory, a radio part with antenna, and provisions for power and to be carried and used by an individual user. A single unit hand-held version is required as well as versions that will operate hands free (except that a push to talk and talk group selector will usually be required). The hands free unit should probably be a wearable design with (i) a lightweight antenna and RF elements in the head gear (helmet or cap) that includes (ii) an ear piece and a microphone; and these connect to (iii) a battery pack with a display and control panel (that may be demountable so the battery is attached to the belt but the control and display element is worn on the wrist). The hand-held version would need to be rugged, compact, and include belt clips, lanyards, headset jack, and other common accessories. The user interface must be simple, and for voice, needs only (i) an on/off and volume control, (ii) talk group/channel selector, (iii) a push to talk button, and (iv) a display. The data input device may be based on evolving cell phone or PDA data interfaces (including, for example, the keyboard on the RIM Blackberry).

The vehicle mounted unit, or Vehicle Radio, will normally interface wirelessly with the user via the user's Personal Radio, described above. A separate display and corded handset/microphone should also be available. The display and data input component, in particular, can be larger. The vehicle antenna and RF elements should be roof mounted. Conceptually the vehicle mounted user terminal can be expected to be more data capable and feature mobile workstation capability (in contrast to the portable unit which will be more voice oriented). The vehicle mounted user terminal will also be expected to provide relay and gateway functionality to support local cluster connectivity and to support connectivity to distant points, clusters, and other networks (e.g., the public switched telephone network).

To ensure connectivity and/or improved performance, there may be instances where transportable terminals are desirable. These would be unattended and, although they will provide the same communication functionality as a user terminal (i.e., fully self configuring), they would require only a simple user interface. At one extreme, the transportable terminals would have only a port for linking it to a user terminal interface that is carried by a user (this port could be radio based). Transportable terminals should be easily secured to fixed infrastructure (to avoid theft) and must be relatively self contained in terms of power generation (perhaps combining solar and/or small generator set cycle-charge plants).

The above user terminals include GPS receivers to support user navigation and agency resource management, and to support network functions including location assisted routing schemes and, as applicable, synchronization. The use of multi mode, multi band and multi transceiver radios, made possible by software defined radio technology, will greatly facilitate the development of flexible and low-cost (multi use) user terminals. Two transmitters and receivers are used to simultaneously communicate in the cluster access and backbone (transport) networks (as shown in Figure 4.4). When not designated as a backbone node, the transceiver would be shut down to conserve power. In total there should be 16 receivers. Note that GPS receivers typically have 8 to 12 receive channel capability, and the ability to simultaneously receive separate voice and data channels is an advantage.

The user terminals will have the processing power and memory of a high end Pocket PC class PDA. The radio part will use the bands noted earlier for the terrestrial ad hoc backbone network and the 2 GHz band for satellite.

As a baseline for the terrestrial network, rate 1/2 channel coding is assumed. For voice packets, convolutional coding and bit interleaving depth corresponding to the packet size is assumed. This is a simple, low delay, and relatively robust approach [Stee92]. For data packets, there may be benefits to using Reed-Solomon and concatenated codes for better error correction and detection properties. However channel coding overheads for data is a trade-off with retransmission overhead and delay, and requires careful consideration that is beyond the scope of this project.

For reference, a 4 level continuous phase frequency shift keying modulation scheme is adopted – for example, a version of minimum shift keying (MSK) or Gaussian MSK (GMSK). This relaxes the need for linear power amplification, requires a relatively simple receiver, and provides a reasonable compromise between power and bandwidth (i.e., spectral efficiency). For voice, bit error rate (BER) performance of 10^{-4} is sufficient considering our packet sizes (Table 4.2), as this should deliver a packet loss rate better (lower) than 5% [Nguy01].

This modulation scheme, under Rayleigh fading and considering a 3 dB implementation margin should deliver 10^{-4} BER performance with an average receive signal level of about -78 dBm in a 200 kHz channel bandwidth (per Section 4.5) [Stee92]. Considering a 15 dB coding gain at that BER [Stee92], a threshold receive signal level of -93 dBm should be achievable. The wavelength at 700 MHz is 42 cm and at 4,900 MHz, 6 cm. Thus diversity receive is feasible at 4,900 MHz (approximately 6 dB gain [Lee95], pp. 217) on Personal Radios but not at 700 MHz. At 700 MHz, 1 Watt transmitters (30 dBm) should be economically achievable and at 4,900 MHz, 0.5 Watt (27 dBm). These radio system parameters provide system gains on the order of 120 dB (see Section 4.9).

Although not user terminals in the strict sense, to ensure reliable wide area connectivity, a satellite network tier is seen as a vital element. The satellite tier could be a single geostationary satellite, a low earth orbit (LEO) or mid earth orbit (MEO) constellation, or a combination. The large coverage footprint and broadcast nature of satellite can be exploited for efficient multicast delivery as well as distance independent linking to remote points or clusters. A fundamental trade off is the amount of on-board processing versus ground based processing. This matter is not discussed further in this report.

4.4 Addressing Scheme

An addressing scheme is needed to support basic network formation, mobility management, security provisions, routing as well as user and network management. Each user terminal should have a permanent equipment identifier. Each user needs a unique name that is part of the user terminal address and should have a part that has global significance to identify member agencies (e.g., for a secure activation procedure). Upon activation, and periodically (time and/or distance change based), a location fix would be generated and become part of the user terminal address. In addition, a group or channel selector would establish the talk group connectivity requirements of the user. For example, one channel may be to an agency resource control center (e.g., a dispatch point or emergency operations center), one for local incident command, one or more for inter agency communications, one or more for special task force use, one or more for interconnect with other networks (e.g., public switched telephone network). The selected channel or talk group identifier will also form part of the user terminal address scheme.

Table 4.1
Addressing space build up.

#	Description	bits	Bytes	Total Space
1	User and agency identifier	28	3.5	268,435,456
2	Terminal / equipment identifier	32	4	4,294,967,296
3	Channel / group (e.g., talkgroup) identifier	26	3.25	67,108,864
4	Location address (GPS coordinates)	26	3.25	67,108,864
	Totals	112	14	

Table-4.1 proposes the size of each address and identifier. To minimize packet size, local clusters need not use full addressing. The cluster head would maintain a table mapping the full addresses to the local addresses. Local addresses would be used on the cluster net and full addresses on the backbone net. Cluster addresses should only require about 16 bits (e.g., 5 bits for 32 locally unique cluster addresses plus five bits for 32 unique local cluster member addresses and the remaining six bits for parity).

4.5 Media Access Control Protocol

A media access control (MAC) protocol is needed to provide efficient and fair sharing of the radio channel resources. Since we are considering a hierarchical network, a MAC layer is needed for both the single hop access cluster network as well as the satellite and multi hop backbone networks.

Since multicast voice is expected to be the dominant high priority traffic, my design process worked back from this criteria. For reference, the ITU-T G.729A voice codec is proposed (refer to Section 4.8). To minimize packet overhead, 6 voice frames per packet have been assumed, as shown in Table 4.2. As shown later (Chapter 5), this introduces significant encode/decode delay; however this is acceptable for push to talk channel delay (but, as also shown later, could cause problems for a full duplex voice channel).

Table 4.2
Voice packet build up.

#	Description	bits	Bytes
1	Speech frames, 6 each G.729 frames per packet	480	60
2	Full address (from Table 4.1) for source and destination	224	28
3	Local (cluster net) address for source and destination (see note)	16	2
4	Allowance for synchronization preamble	8	1
5	Allowance for header error control bits	8	1
6	Time stamp (used to drop packets that have been too long in transit)	16	2
7	Sequence number (used to reorder out of order delivery)	8	1
8	Traffic type or classification (used to deduce, for example, priority)	8	1
9	Spare / future use	16	2
10	Total with full address	768	96
11	Total with local address	560	70

Cluster Net MAC

The cluster access network (cluster net) refers to the network formed by the clusters that home on a member of the backbone network dominating-set. The push to talk channel is, by definition, a half duplex channel. Although push to talk is not a strict requirement as voice operated transmission and sophisticated conference bridging functions are available, there are issues relating to operational procedures and disciplines that make its use here a reasonable assumption. Given half duplex operation, time division duplexing (TDD) is a natural choice. A very simple approach is to have the cluster head echo (repeat) each packet in the next slot. This provides a broadcast to all nodes in the cluster (i.e., some nodes may be out of range of the source node) and provides a delivery acknowledgement to the source (for delivery at least as far as the cluster head). A slotted channel is assumed (the GPS signal can be used for slot synchronization).

The cluster nets are single hop ad hoc networks. Because the dominant set member is a cluster head, the various carrier sense multiple access with collision avoidance (CSMA/CA) schemes may be used (see Appendix B). However to minimize multiple access control overhead and complexity a relatively simple random access scheme is proposed as a first generation approach. The scheme is based on an adaptation of slotted Aloha [Abra73] that was proposed by Goodman [Good89] as the Packet Reservation Multiple Access (PRMA) protocol.

Recall that the cluster head echoes every transmission by a cluster member. This improves link reliability as some of the nodes will hear both the original and the echo transmission. Recall that each voice packet carries 60 milliseconds of coded speech (6 each G.729 10 millisecond frames). If we assume that six time slots fit in this time frame, then the time slot duration is 10 milliseconds. Each cluster member listens and keeps track of the last six slots, marking each slot as idle, in collision (unsuccessful contention) or busy; and if the slot is busy, whether it was busy with their traffic or some other node's traffic. For example, in Figure 4.5, node C is the cluster head and echoes the transmission from each member. Node A, B and D successfully sent packets in the past six slots. Note A will transmit in the current slot as it was successful six slots past. Nodes B and D are also holding reservations and will honor the reservation for A in the current slot by not transmitting (they will wait for their slots). Node E and F are waiting to transmit and unsuccessfully contended in the previous slot (i.e., slot -1). Nodes E and F will contend for each upcoming unreserved slot – i.e., previously idle or collided.

The number of slots in each frame is a complex trade off involving many factors including bandwidth, channel capacity, and performance. The 6 slots chosen here is loosely based on keeping the RF bandwidth on the order of 200 kHz yet providing sufficient channels to obtain some traffic pooling efficiency.

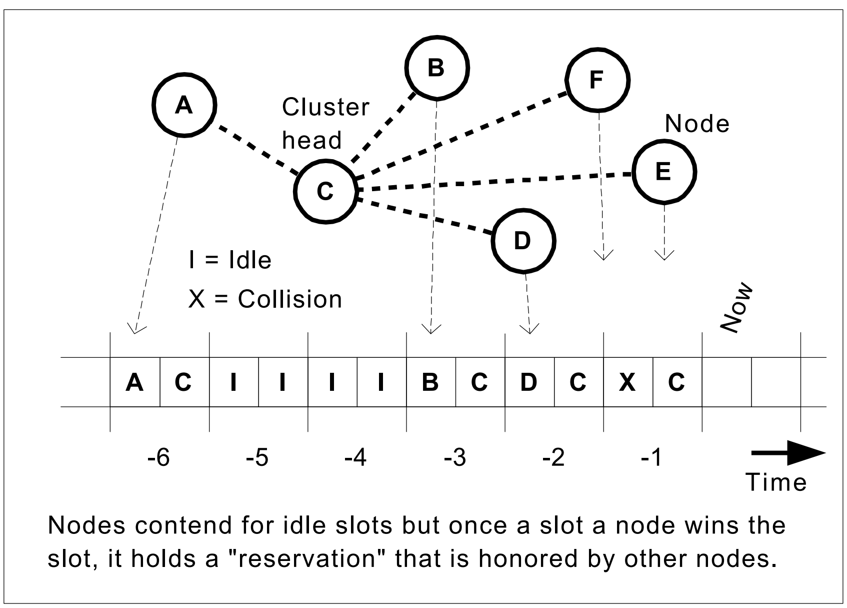

Figure 4.5
Time slot concept for the cluster access network (cluster net).

The probability of transmitting is calculated based on the total number of talk groups in the cluster (known by the cluster head and broadcast periodically) less the known reservations. The probability can also be adjusted by a priority factor that considers the number of idle slots and message (talk group) priority before making the decision to contend or not. Essentially the probability of a successful outcome is maximized when the probability of each node transmitting is set to the inverse of the number desiring to transmit. The above approach is an approximation that can be made by each node locally by listening to the channel.

Using the packet size per Table 4.2 and the six slot frame size noted above that must fit in each 0.03 second half-frame (the 0.06 second frame is time division duplexed), the channel bit rate can be calculated as 224 kb/s (from 6 slots per frame x 2 directions x 1 packet per slot x 560 bits per packet x 2 encoded bits per un-encoded bit x 1 frame / 0.06 sec). Adding guard bits for go return delay over at least a 2 km path, implies a channel rate on the order of 256 kb/s. And from the assumed modulation scheme (Section 4.3), and a roll off factor of 0.5, an RF channel bandwidth of about 200 kHz (including a nominal guard band) can be calculated.

Taking the 50 MHz allocated for cluster nets in the 4,900 MHz band and 200 kHz channel bandwidth, 75 RF channels can be formed. These can be arranged in 75 orthogonal hopping sequences that can be used for frequency reuse and interference mitigation. As a reference design, seven sets of 10 orthogonal sequences can be formed into a seven cell reuse plan (see Figure 4.6). Assuming a cell size of 1 km, this provides a reuse distance of 4.6 km [Rapp96]. Adjacent reuse groups will use uncorrelated hopping sequences [Stee92] for graceful degradation under interference. At 700 MHz, 30 each 200 kHz channels can be formed in the 6 MHz noted earlier. Taking 7 of the orthogonal hopping sequences would provide one each hopping sequence per cell, for a similar N = 7 reuse scheme at 700 MHz.

Another issue is to assign channel resources without centralized control. The virtual cell layout (VCL) architecture from [Cayi02] resolves the issue of allocating radio resources to facilitate frequency or code reuse. This architecture is based on a preset mapping of the frequency plan to a hexagonal grid that is overlaid on the service area. User terminals use knowledge of their geographic position and this grid to determine which frequency set or code to use (in our case it is a hopping code). As terminals move across the grid, they perform handoffs to conform with the preset plan.

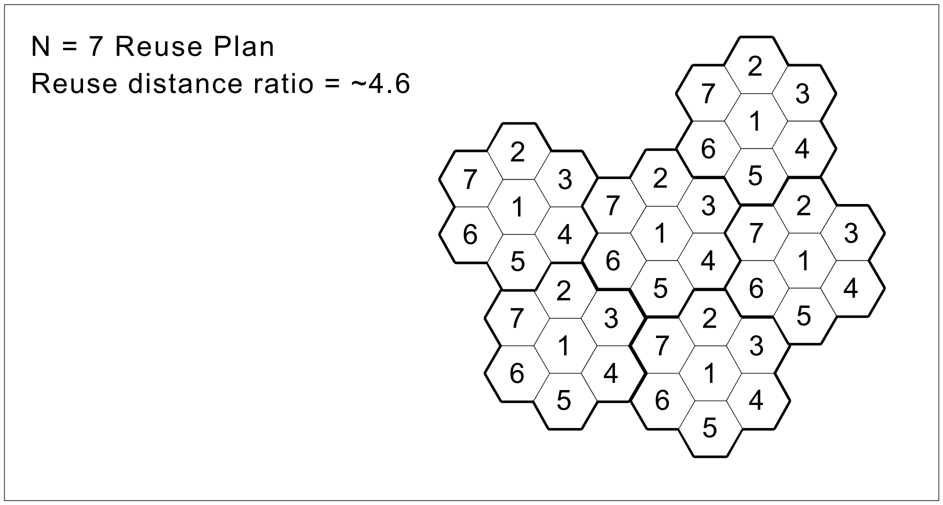

Figure 4.6
Hexagonal grid based reuse plan [Rapp96].

Backbone Net MAC

The backbone MAC design will be influenced to a certain extent by the backbone routing protocol. For example, sleep modes (for energy conservation) must be coordinated between the two layers. As noted in the following section, a multi hop routing scheme was not proposed in this project. Therefore the MAC protocol of the backbone is also left open.

4.6 Multi Hop Routing

As noted in Appendix B, routing in mobile ad hoc networks is a very challenging problem. Within the scope of this project, I have reviewed recent literature and identified promising directions. This section introduces these promising directions, identifies the outstanding issues and concludes that hybrid ad hoc solutions that use satellite is a pragmatic near term approach.

The routing design challenge

A routing protocol finds the destination (a mobility management issue) and finds a route to it (the routing issue). In our case, the destination and route discovery will be a set up phase that precedes the message forwarding session. Sending the packets associated with a multicast talk spurt (call) through the network constitutes the message forwarding session. The routing protocol must: (i) be loop free, (ii) consider sleep modes (to conserve energy), and (iii) consider security (to resist denial of service attacks). To scale in a mobile ad hoc network, where topology changes are the norm, hierarchical architectures will need to be supported, and at some level of the hierarchy, on-demand (reactive) routing procedures will be needed.

Assessing the performance of a routing protocol involves metrics. For example relevant metrics include: (i) route acquisition (destination and route discovery) or set up time, (ii) end-to-end throughput, (iii) end-to-end delay, (iv) delivery rate as the ratio of messages sent over messages received (guaranteed delivery means delivery rate = 1), and (v) efficiency (e.g., average data bits received over the average total data bits sent plus overhead bits). Other metrics that maybe useful in our application are (i) proportion of messages delivered out of sequence, and (ii) flooding rate (defined as the number of messages sent divided by the shortest path hop count). Performance measures must be made under realistic traffic and graph density (average degree) conditions. Several authors have observed that for efficiency and good performance, the design of a mobile ad hoc network needs to consider all layers from radio (physical) to application [Stoj02], and [Gold02]. In particular, the radio, MAC, routing, and application layers should be aware of each other (share information), coordinate their activities and adapt as conditions change.

Factors specific to our situation

Our requirements dictate that each user and user terminal will have a home agency. Each agency will need to maintain a resource control center that is responsible for managing their field assets, including the users and the ad hoc network. This center forms a logical home location register that can assist with mobility management.

Each talk group will need a group control station that maintains the membership list. Some talk groups will have agency or multi agency significance, and will therefore need to be maintained at the agency or joint emergency operations center levels, respectively. Other talk groups will be related to a specific task or incident and the talk group control station will be in the field. Finally the mobility and multicast management scheme must permit the ad hoc creation of talk groups but with constraints to ensure overall network integrity is not compromised.

Satellite resources are key assumptions. Firstly, since location information is needed by the agency resource control center, it is also available to the user terminal (which we have assumed is GPS based). Secondly, a satellite backbone layer is assumed available – in fact, it is essential for connectivity in sparse wide area networks. These are key capabilities and assets that the routing scheme can exploit.

Clusters, hierarchy, and dominating-sets

Forming clusters and providing network hierarchy is used to provide scalability. Without hierarchy, only very small networks are practical. Fortunately a certain amount of hierarchy is built into our set of user terminals – i.e., Personal Radios, Vehicle Radios, Transportable, and Fixed Radios, and the satellite layer provide a natural basis for devising network tiers. Unfortunately, the optimum of finding clusters and cluster heads to minimize topology reconfigurations is an NP-hard problem [Stoj02], pp. 338. A promising approach, and the one that I have suggested in this project, is to use dominating-set based clustering with distance-1 – i.e., only one hop neighbors of the cluster head are in the cluster. The primary advantage is it makes the cluster routing problem trivial – it is reduced to a MAC issue. The inter cluster backbone is formed by a connected dominating-set. This induced sub graph, being smaller, reduces the complexity of the routing problem.

Obviously, the starting point for dominating-set based routing schemes is to form a connected dominating-set. An optimum is a minimum connected dominating-set (fewest possible number of nodes). Unfortunately, finding a minimum dominating-set on an arbitrary graph is an NP-complete problem [Stoj02], pp. 428. Fortunately, J. Wu has developed a distributed marking scheme that forms a reasonably good dominating-set, with extensions to support hierarchical dominating-set and multicast routing [Stoj02], pp. 431. The marking based procedure uses distance-2 neighborhood information. The procedure provides good results on network graphs where the diameter is not too large and when used in conjunction with rules to reduce (prune) the induced dominating-set. The cost of marking at each node grows on the order of the average degree squared and the amount of message exchange grows linearly with degree. As noted later, average degrees beyond 15 provide no further connectivity benefit. The multicast extension is a hybrid of flooding and a multicast shortest tree scheme [Stoj02], pp. 445.

Dominating-set routing involves three steps: (1) if the source is not the cluster head, then send the packets to the cluster head; (2) the cluster head routes packets in the induced sub graph created from the connected dominating-set; and (3) if the destination is not a cluster head, then the last cluster head forwards the packets to the destination. Routing on the connected dominating-set induced sub graphs can be pro-active, reactive or a hybrid. As discussed next, a hybrid is likely where reactive techniques play a prominent role.

Backbone routing

It must be possible to form and maintain a viable network in the absence of infrastructure or connection to a control center. Thus fully distributed operation is essential. Ideally, for fully distributed operation each node decides packet forwarding using only the location of itself, its neighbors and the packet destination. This can be a greedy algorithm. A greedy algorithm is loosely defined as an algorithm that uses simple local behavior to achieve a desirable global objective. Such a scheme provides a highly scalable arrangement as each node does not need to have or maintain a global view. Also for scalability, the routing scheme will have hierarchy, with different local and long distance routing algorithms.

How big should the local neighborhood be? Neighborhood distance-2 information is needed to support the marking process that is used to form the dominating-set. So this is a good starting point. Thus within this neighborhood, pro-active table-based routing is essentially built-in (and note that the distance-1 flooding happens by default). However, frequent topology changes are expected due to mobility. This makes pro-active routing inappropriate by itself (to avoid overwhelming the network with routing update traffic). Thus outside the immediate neighborhood, reactive (on-demand) routing can be expected. This creates an issue at the transport and application layers as most existing transport protocols and applications were designed for pro-active routing schemes.

Since we have access to node geographic location information, the reactive routing scheme can exploit this information. Many routing algorithms have been proposed that use the notion of forward progress toward the destination [Stoj02], pp. 455. One recent variation referred to as Compass routing [Stoj02], uses the destination's geographic location at each node to select the next hop that is closest direction-wise (nearest compass bearing) to the destination. Although simple, if forward progress is blocked (e.g., in a graph with a convex hull), the algorithm fails. To overcome this limitation, various dual mode schemes have been advanced, wherein upon being blocked another mode is invoked until the block is circumvented.

Small world graph approaches have been advocated for route discovery in very large networks [Stoj02], pp. 341, [Blaz01], and [Huba01]. In [Blaz01], a small subset of nodes (called "friends" in Friend Assisted Path Discover) is used to understand the approximate shape of the geographic path from source to destination. From this, anchor points are set (Anchored Geodesic Packet Forwarding) that can be used with greedy routing algorithms, like Compass. Forming and maintaining these geographic anchors (e.g., routing signposts) may be facilitated by the satellite backbone layer. These anchors may also be candidates for mobile agent technology where anchors are identified and their routing tables updated by mobile agents.

Multicast

Although multicast protocols are available for fixed networks, they rely on some fixed infrastructure and are thus not suited to mobile ad hoc networks without significant adaptation. Specifically, most of these multicast algorithms are pro-active. Some reactive and multicast protocols have been developed for MANET. Two of these are reviewed in [Stoj02], pp. 502; Multicast Ad Hoc On-demand Distance Vector (MAODV) and On-Demand Multicast Routing Protocol (ODMRP). The latter, being mesh based, has higher overheads (which limits scalability) but better performance in high traffic and high mobility situations.

A potentially useful observation is that multicasting sets up multiple routes through the network which suggests the possibility of using multi path reception. Specifically, if the backbone network is sufficiently large and complex, multiple routes may be discovered. These can be exploited by applying a combination of diversity routing and diversity coding as proposed by Tsirigos and Haas [Tsir01]. This scheme divides each packet into N blocks, adds M blocks of redundancy and distributes the transmission among the available paths. This approach improves reliability and should mitigate delay caused by congestion, as it provides a measure of flow or load balancing.

Key findings on routing

In closing this section, we should note that an architecture for delay constrained quality of service support needs resource reservation and/or prioritized scheduling. This is largely an open issue for mobile ad hoc networks [Stoj02], pp. 339. Although an efficient routing algorithm for delay constrained multicast traffic in a multi hop mobile ad hoc network is an open problem, I believe that useful first generation ad hoc networks can be fielded with current technology. Routing on the backbone of of these early networks will greatly benefit from, or totally rely on, satellite occupying the high ground in the network hierarchy. This is not a pure ad hoc network, but what this hybrid lacks in pedigree, it will make up for in practical performance in real world networks.

4.7 Network Management

The International Organization for Standardization (ISO) defines five network management functional areas: (i) fault management, (ii) accounting management, (iii) performance management, (iv) configuration management, and (v) security management. Configuration management involves network resource allocation to meet connectivity and performance requirements and provides these functions during network initialization, reconfiguration, and shut down. This is a core function and it must be included in a fully distributed manner. For example, as a user terminal is turned on (or regains coverage after a loss of coverage event) it will go through an initialization procedure such as the following:

- Check location and status (where am I and what is my condition)
- Listen and page (hello, anyone out there) on an appropriate radio channel
- Support in the formation of the ad hoc network
- Find and report to the relevant agency resource control center
- Find and report to the appropriate (selected) work group (talk group).

It is also vitally important to monitor network health and know when connectivity and/or traffic load conditions threaten performance. This information needs to be made available, and as thresholds are crossed, the system should automatically lapse into graceful and stable failure modes. Obviously, recovery must be automatic.

The network management capability will coordinate power savings (sleep modes) and security functions (authentication, encryption, and resistance to denial of service attacks). To facilitate the distribution of these functions, minimize management overhead, and to provide resilience to network partitions, the network management functions are expected to be implemented in a hierarchy.

In terms of a network management protocol, the ad hoc network management protocol (ANMP) as proposed by [CheW99] may be suitable (the issue of support for IP is addressed in Section 4.8). ANMP is compatible with Simple Network Management Protocol (SNMP) version 3.

SNMP is well supported (popular), lightweight (low overhead). SNMP version 3 supports reporting by exception (to minimize management traffic) and includes a security model with authentication and encryption capability. ANMP features ad hoc specific extensions including support for network hierarchy (clustering and topology maintenance), power awareness, and secure multicast.

4.8 Application Design

The communication system must be designed to support multicast (work group) and unicast (point-to-point) voice and data traffic. The relative priorities and quality of service limits can be identified by the traffic originator (source node – for example, talk groups (voice messages) can be assigned priority levels and time to live limits, and data messages can have priority levels and delivery reliability settings preset by message type. All applications should be implemented with tolerance for variable levels of underlying network performance. Traffic must be classified by quality of service parameters (delay and/or throughput) and there must be an inherent and fully distributed ability to intelligently trade power, spectral efficiency, and/or other network resources for specific performance improvements (delay and/or throughput).

In this design, I have assumed voice inputs are coded using ITU-T G.729A. This codec is readily available and relatively robust under loss (packet loss ratios up to 5% are generally tolerable), and provides relatively low complexity and processing delay [Kost98] and [Nguy01]. The data capacity of the design concept is limited, thus applications must be lean. Any images must be compressed using relevant JPEG standards, with source and channel coding to improve performance during channel errors (for example, Huffman coding with resynchronization, and controlled redundancy added using advanced channel coding schemes, such as, turbo codes) [Hema01].

There are many existing and emerging data applications based on the Internet Protocol (IP). Clearly any next generation system must consider this fact. Unfortunately IP overhead is large and, with IPv6, growing. Header compression and other techniques will be needed, and this remains an open area for investigation.

4.9 Conceptual System Design Summary

The following table summarizes the salient design choices and characteristics of the proposed system design described in this chapter.

Table 4.3
Summary of the proposed ad hoc network system design concept.
In addition to the ad hoc operation modes, the radio should be capable of emulating legacy air interfaces to provide backward compatibility.

General User Terminal Characteristics	Frequency bands		700 MHz, 4900 MHz, 2000 MHz (satellite), GPS receive	
	Receivers		16 total (8 GPS only, 8 frequency / band agile)	
	Transmitters		2 total (agile to communication bands)	
	Battery save modes		yes (sleep cycles, coordinated with other layers)	
Voice Packet Characteristics	Voice coding		ITU-T, G.729A	
	Voice frames per packet		6	
	Voice duration per packet		60 ms	
	Voice encode/decode delay		65 ms	
	Packet size (before coding)	on Cluster net	560 bits (70 Bytes)	
		on backbone net	768 bits (96 Bytes)	
Cluster Net Routing	Topology		single hop cluster, ad hoc network	
	Routing protocol		part of MAC	
Cluster Net MAC Characteristics	Duplexing		time domain (TDD)	
	Channel access		frequency hopping, 75 orthogonal codes	
	Code reuse		7 cell reuse plan (N = 7)	
	Orthogonal code sequences per cell		10	
	Code allocation scheme		virtual cell layout	
	Channel framing		6 forward and 6 return slots	
	Time slot duration		0.5 ms	
	Multiple access scheduling protocol		PRMA	
	Typical one way delay (after set up)		10 ms	
Cluster Net Radio Characteristics	Code rate		1/2	
	Channel rate		256 kb/s	
	Channel bandwidth		200 kHz	
	Modulation		Continuous Phase Frequency Shift Keying, 4 symbol	
	Transmit power	4900 MHz	0.5 Watts (27 dBm)	
		700 MHz	1.0 Watts (30 dBm)	
	Average receive level for BER 10^-4		- 93 dBm (in Rayleigh fading, with coding gain)	
	Diversity receive gain, 4900 MHz		6 dB	
	Antenna gains		-1 dBi for Personal Radio	3 dBi for Transportable Radio
			1 dBi for Vehicle Radio	5 dBi for Fixed Radio
Backbone Net Characteristics	Topology		dominating-set based, multi hop ad hoc	
	Routing		hierarchical with local proactive and hybrid reactive (e.g., Compass and small world graph with anchors)	
	MAC		to be determined	
	Radio characteristics		to be determined	

Note that, although many of the design elements are well-established, several important ones are not. For example, the network routing and management (configuration, sleep modes, resource allocation) schemes as well as addressing and security arrangements are not specified above, incompletely specified, and/or untested in a simulated or actual context. Additional research, trade-off studies, and prototyping are needed to complete the design and to refine and optimize performance and robustness.

Chapter 5
Preliminary Feasibility of the Hypothetical Ad Hoc Network

Chapter 4 describes my proposed conceptual system design, built by selecting architectural and system elements that consider the operational context and requirements. In this chapter, I document my preliminary assessment of the feasibility of this design. Specifically, while the scope of the project (and the state of the art) precludes a definitive assessment, in this chapter, connectivity and fundamental capacity issues are examined. Although performance needs to consider many aspects, the focus here is delay and cost performance. Consistent with the scope of this project, the treatment is cursory and seeks only to establish the potential for adequate performance.

5.1 Performance Evaluation Scenarios

To provide a context and basis for performance evaluation, two scenarios are used. They represent two relatively different situations and attempt to test the design concept across a wide range. The first scenario, an urban fire incident, represents a high density situation. The second scenario, a remote highway accident, represents a very low density situation.

Urban fire scenario

In an urban area, there are 12 vehicles with a 60 person task force proceeding to a warehouse fire. This is after a major earthquake reeks havoc on commercial communication systems and knocks out several key public safety communication sites.

Altogether in the overall urban area, there are a total of about 400 active public safety and related agency vehicles (e.g., public utilities such as gas, electric, telecommunications, and transit) in a 10 x 10 kilometer area. Each vehicle is equipped with a vehicle mounted ad hoc network capable user terminal – Vehicle Radio. Associated with these vehicles, and carrying (or wearing) Personal Radios, are some 800 personnel. In addition, at each fire hall, police station, major telecommunication facility, and agency headquarters there is a Fixed Radio. Therefore, over 1,200 ad hoc network radio assets are available across the urban area.

At each agency headquarters, a resource control center (dispatch or command center) has access to location and status information for all of its field assets and personnel through the ad hoc network. In addition, the emergency operations center has summary agency information as well as detailed information on the ad hoc network including, for example, connectivity and battery resources.

Meanwhile enroute to the fire, all personnel (60) are on a common talk group and receiving a briefing on what is known of the situation. The vehicles are arriving from three different directions and communications is being supported by an ad hoc backbone operating at 700 MHz that connects vehicle to vehicle and vehicle to various, in range, Fixed Radios. While in transit, within the vehicles personnel receive the briefing on their Personal Radios that are linked to the Vehicle Radio at 4,900 MHz. Upon arrival at the scene, the vehicles are positioned and personnel deployed. One vehicle is the incident command post. Three teams are formed and begin conducting fire suppression operations. Each team is in contact with all their members and the command post. Usually transmissions go from the Personal Radios to a cluster head Vehicle Radio using the 4,900 MHz band. Cluster members may receive a transmission directly from the originator, when it is repeated by the cluster head vehicle (in the next time slot) or both. If some talk group members are in range of another vehicle (or vehicles), then in the second time slot, the packet is simultaneously transmitted to the other vehicle (or vehicles) on the backbone channel and retransmitted at 4,900 MHz (each vehicle communicates with its local cluster on a different hopping code). Vehicles link to one another, including the command vehicle, either directly or through one or more relay vehicles in the ad hoc backbone at 700 MHz or at 4,900 MHz (but in the backbone portion of the 4,900 MHz, if at 4,900 MHz).

In the event that a member of the team loses contact with the cluster head vehicle, but remains in contact with one or more neighbor Personnel Radios, this other Personal Radio configures as a backbone node in the backbone frequency band. This relay action significantly increases the battery drain. The user receives an indication this event has occurred and as or if the battery depletes to a threshold, the radio may signal a reconfiguration to select (if possible) another radio as relay.

Remote accident scenario

In a remote rural jurisdiction, a highway maintenance vehicle on a routine maintenance trip encounters a two vehicle accident with one vehicle carrying hazardous materials. Altogether, this jurisdiction covers some 4,000 km of roads and is sparsely populated with only 20 communities scattered across it. Including public safety and government agencies, there are 800 vehicles equipped with Vehicle Radios and some 1,000 personnel, each equipped with a Personal Radio. Virtually all personnel are associated with a vehicle, although in the course of their operations they work remotely from the vehicle. In addition each agency maintains a Fixed/Transportable Radio at their major facilities, camps, and works zones.

Meanwhile at the accident scene, the highway maintenance vehicle has stopped and the driver is approaching the scene on foot. As the seriousness becomes apparent, she selects the talk group (channel) for the joint emergency response communications talk group and sends an alert. This request travels from the Personal Radio to the Vehicle Radio in the 4,900 MHz band. The vehicle is out of range of any other terrestrial radio, but has been maintaining contact on the satellite channel. Thus the request goes to a centralized emergency call and dispatch center (for example, a public safety answering point), and the necessary units from police, ambulance, fire, and rescue vehicles are dispatched. A separate talk group is assigned to the incident.

A police cruiser is within 35 km. The nearest ambulance and rescue vehicle are in a community 80 km distant. These vehicles proceed to the scene. The ambulance and rescue vehicles leave within a minute of each other and travel about 2 kilometers apart. Enroute all vehicles monitor the talk group channel and get updates from the scene over the satellite channel. As the vehicles approach the scene, the vehicle radios establish direct links between them using 700 MHz, and as the distance reduces, finally 4,900 MHz. At the scene, all units are in range and the network reconfigures automatically with one vehicle selected as the cluster head. Communication at the scene is to the cluster head vehicle which acts both as a repeater to insure local communications (easy if the incident is as local as most vehicle accidents) and as a backbone node linking through satellite to the emergency call and dispatch center. As the ambulance departs, the driver and paramedic change to a talk group with a medical clinic in the nearest community, which will be there first stop. Communications to the vehicle is at 4,900 MHz and to the clinic is via satellite. At the clinic community there are two nurses one at the clinic and one a short distance away. Communications between the nurses and the clinic is at 4,900 MHz with the clinic Fixed Radio acting as cluster head. As the ambulance closes on the clinic, the satellite link drops out and a direct the Vehicle Radio to clinic Fixed Radio link is established using the 700 MHz backbone frequency.

5.2 Connectivity Performance

Connectivity is the most critical performance issue as it absolutely defines system coverage and plays a substantial role in system capacity (covered in the next section). To assess the potential performance of the proposed system, the above two operational scenarios are considered.

Urban fire scenario

In the urban fire scenario, the probability of on-site connectivity can be determined from Figure 5.1 and Table 5.1. Specifically, if we assume a fire operational area that is 0.8 km on a side and, considering the conditions of a large warehouse, the Personal Radios should have a range of about 50 m to 200 m at 4,900 MHz (0.05-0.2 km); therefore, if randomly located, between 80 and 1,000 nodes would be needed to assure connectivity (from Figure 5.1 and a normalized radio range of 0.0625 to 0.25, which is calculated from 0.05 / 0.8 and 0.2 / 0.8, respectively). Careful placement of the vehicles could alleviate the issue. It is also likely that in such a large fire area, there would be clusters of activity. These clusters would be able to maintain a connectivity over a 700 MHz backbone if we assume at 700 MHz the backbone nodes have a range of 300 m, as now less than 40 radios assure connectivity (the normalized radio range is now 0.375, from 0.3 / 0.8).

Considering the urban area at large, with approximately 400 Vehicle Radios and 800 Personal Radios available in an area 10 km on a side, if we can achieve an average range of 1 km then, at 500 radios we are generally assured of connectivity (the normalized ratio is 0.1). The 1 km average is reasonable when we consider that fire halls and other facilities will have a fixed radio with a backbone range of some 0.5 to 3 km. These nodes should mitigate the formation of large coverage holes.

Table 5.1
Approximate radio ranges at 700 MHz and 4,900 MHz. Adapted from approach in [Rapp96]. Notes: (i) Based on 95% probability protection against log-normal shadow fading. (ii) Best case is free space transmission; may be limited by TDD guard time. (iii) Note that, satellite provides distance insensitive coverage, but requires adequate sky view.

User Terminal Type	Band (MHz)	In/Into Building	Dense Urban	Urban	Sub Urban	Rural	Best Case
Personal Radio	700	0.10	0.17	0.55	2.1	5.6	24.2
	4900	0.05	0.09	0.25	0.7	1.5	4.9
Vehicle Radio	700	0.11	0.18	0.62	2.5	6.7	30.5
	4900	0.05	0.09	0.28	0.8	1.8	6.1
Transportable Radio	700	0.12	0.20	0.70	2.9	8.1	38.4
	4900	0.06	0.10	0.31	1.0	2.2	7.7
Fixed Radio	700	0.13	0.22	0.78	3.4	9.7	48.4
	4900	0.07	0.12	0.35	1.1	2.7	9.7

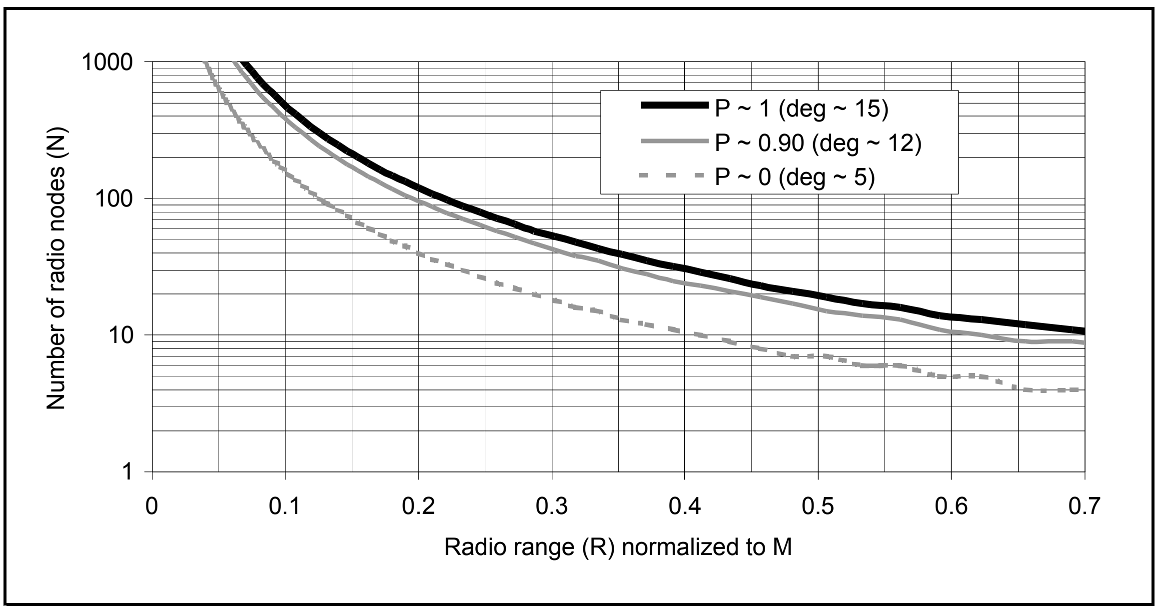

Figure 5.1
Probability of network connectivity .
Probability of network connectivity, P, given N nodes randomly distributed in an M x M area as a function of the node radio range R (normalized to M) [Lou02] and [Gold02]. The average degree (number of neighbor nodes) is shown in the graph legend.

Remote accident scenario

On the scene at our remote accident, connectivity is assured as distances are short, at most hundreds of meters. Enroute communication and links to the emergency operation center are handled by satellite. At the community, one nurse is away from the clinic. If the nurse is in a building sufficiently far from the clinic, say five or 10 km, the clinic Fixed Radio will not have the range. Being indoors, the nurse's Personal Radio cannot reach the satellite. However, in this instance, the nurse’s car has a Vehicle Radio on board. The nurse first communicates to the vehicle at 4,900 MHz, then via a 700 MHz backbone link to the clinic or, if beyond the 700 MHz band radio range, then via satellite.

In the sparse region scenario, the satellite provides the backbone of last resort and the availability of personal connectivity will depend on either having adequate direct sky view or being within range of a vehicle with adequate sky view. In the majority of practical situations, these conditions are expected to be met with a high probability (e.g., over 99% of the time).

5.3 Capacity Performance

System capacity for the remote accident scenario is not an issue (an implicit assumption is the satellite service is adequately provisioned).

At the urban fire scene, we have a talk group for each of 3 teams plus a common incident command channel. These 4 talk groups will all be high priority. The talk groups could potentially generate 0.5 Erlangs each for a total of about 2 Erlangs. It is likely that three or more clusters would form to meet the needs of the incident. If we assumed 2 clusters are needed per team, then there would be 6 clusters. Since the operational area is 0.8 km by 0.8 km, it may fit within one virtual cell. As noted in Chapter 4, at 4,900 MHz there are 10 hopping sequences available and each sequence can form a cluster with 6 channel capacity. Therefore, we conclude that even with 6 clusters forming, we still have 4 hopping sequences left over – even if all traffic is carried in every cluster (worst case situation where at least one member of each talk group is in each cluster). The backbone network, although not characterized, should not be a challenge in this scenario. This may not be true in the overall urban area – so let us go there next.

To assess the capacity for the total urban area, let's look at the cluster network capacity first. Considering a virtual cell layout with 1 km radius, means each grid has an area of approximately three square kilometers. In each grid, we have capacity for 60 simultaneous voice channels (from 10 hopping sequences per cell with 6 voice channels each). Therefore there is cluster net capacity to support up to 60 high priority talk groups in the urban area. This assumes a worst-case scenario where each talk group is distributed such that every cell has traffic from everyone of these talk groups. From Chapter 3, recall that the average number of talk groups for the Vancouver cell is approximately 26. Therefore, we can conclude that we have sufficient cluster net capacity.

How much backbone capacity is needed? In the 10 km by 10 km area, there would be approximately 33 cluster cells (from 100 km^2 / 3 km^2 per cell). Therefore, assuming each cluster may run to 60% utilization, then, worst-case, up to about 1200 Erlangs of voice traffic could be generated (from 33 cells x 10 cluster nets/cell x 6 channels/cluster x 60% utilization). Assuming 8 kb/s voice per Erlang and a 100% overhead factor on the backbone, a bottleneck link would need to have up to 20 Mb/s capacity (from 1,200 Erlangs x 8 kb/s x 2 for overheads). This is a significant but not unmanageable figure. In reality such a single choke point is unlikely. A more credible assessment requires characterization of the mobility patterns and the geographic distribution of traffic.

Although obvious, it should be restated that this design is voice oriented and relatively lightweight. It would not support heavy data traffic applications from extensive mobile office, image transfer or video applications (the last three data applications on Table 2.2).

5.4 Delay Performance

Table 5.2 builds up a delay budget for channel (talk group) set up latency and for the end-to-end voice stream latency. Delays at the user terminals and in the cluster nets are based on the hypothetical design in Chapter 4. The terrestrial backbone delays are rough estimates and should be considered targets. Multiple access delays for the cluster nets are shown in rows 3 and 18. This delay is a random number and depends on traffic and talk group priority. The values are considered reasonable budget allocations. They are based on a sample of results from a software model that I built to emulate the cluster net design – refer to Figure 5.2.

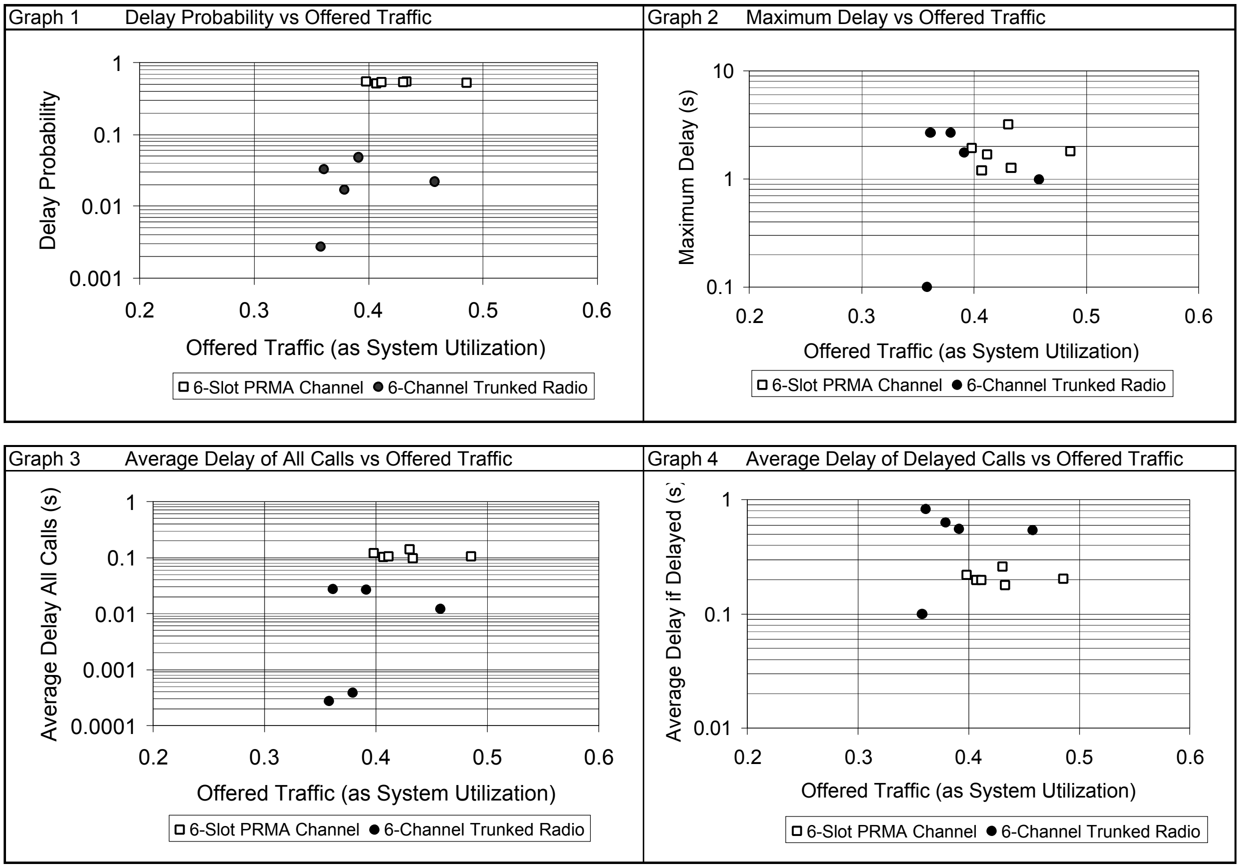

Figure 5.2

Comparison of the channel set up delay performance, for 10 talk groups.

Based on a 6 slot PRMA model built up in an Excel spreadsheet using the parameters from Chapter 4 and run by a Visual Basic program module within the spreadsheet. A genuine traffic trace comprising 10 talk groups was fed into the model. Genuine traffic was used because, as noted in Chapter 3, the underlying traffic may have long range dependency. The trace was taken from a one hour busy period and the performance results extracted from the model after every 10 minutes of run time. For comparison, this figure also shows the results of a 6 channel trunked radio system with the same genuine traffic trace. The trunked radio system is a 6 server queue model (and assumes negligible control channel contention delay); also built up in an Excel spreadsheet and run by a Visual Basic program module. The PRMA data points are shifted to the right because of packet overhead. This analysis is an example only and more extensive and rigorous treatment is left as future work.

Table 5.2

Voice call delay build up for priority traffic.

Based on the hypothetical design from Chapter 4. Figures are representative for priority traffic (no queuing delay).

#	Description	Delay (ms)		
			Voice end-to-end	
		Channel Set-up	Half Duplex	Full Duplex
Source User Terminal				
1	Voice source encode and assemble packet	3	65	65
2	Interleave, channel encode, baseband process	2	2	2
Cluster Net Inbound to Clusterhead				
3	MAC, PRMA contention delay allowance	200	-	-
4	Transmission time	4.4	4.4	4.4
5	MAC access frame delay (average)	3	3	3
6	Propagation delay (based on 10 km at 3.4 usec / km)	0.03	0.03	0.03
Terrestrial Backbone Net (delays per node traversed)				
7	Backbone route discovery	100	-	-
8	MAC scheduling delay	100	-	-
9	Transmission time	10	10	10
10	MAC access frame delay (average)			
11	Propagation delay			
12	Terrestrial link propagation delay (150 km)			
Satellite Backbone Net				
13	Satellite set-up delay excluding propagation time	200	-	-
14	LEO satellite link propagation delay (orbit at 1,000 km)	6.8	6.8	6.8
15	MEO satellite link propagation delay (orbit at 10,000 km)	68	68	68
16	GEO satellite link propagation delay (orbit at 36,000 km)	245	245	245
17	MAC, transmission and processing delay allowance	15	15	15
Cluster Net Outbound from Clusterhead				
18	MAC, PRMA contention delay allowance	200	-	-
19	Transmission time	4.4	4.4	4.4
20	MAC access frame delay (average)	3	3	3
21	Propagation delay (based on 10 km at 3.4 usec / km)	0.03	0.03	0.03
Destination User Terminal				
22	Voice source decode and assemble audio stream	3	65	65
23	De-interleave, channel decode, baseband process	2	2	2
24	Total - within one cluster net	225	149	149
25	Total - 1 terrestrial backbone link	635	159	159
26	Total - 10 terrestrial backbone links	2,525	249	249
27	Total - 30 terrestrial backbone links	6,725	449	449
28	Total - LEO backbone	654	171	171
29	Total - MEO backbone	776	232	232
30	Total - GEO backbone	1,130	409	409
31	Target	1,000	600	150

At the bottom of Table 5.2, overall delays for various network configurations are summed. And in the very last row, the target values are presented (from Chapter 2). Comparing these figures, we conclude that there is:

- Good overall performance within a single cluster
- Adequate end-to-end delay performance for all configurations assuming half duplex (i.e. push to talk) operation
- Poor channel set up performance for more than a few terrestrial backbone links and over a GEO satellite

Consequently, if the terrestrial backbone is to extend beyond a few hops, it is important that the backbone design beat the delay targets in Table 5.2. Although not readily evident from Figure 5.2 and Table 5.2, it should be noted that priority treatment will be necessary to meet the 99.9% quality of service requirement for high priority traffic per Table 2.4 (metric 4).

For reference on session set-up delay, modern high end trunked radio systems that are not too geographically dispersed can set up in less than one second (500 to 800 ms are typical figures). The MPT1327 standard, arguably the most widely deployed trunked radio system in the world, requires about 1.5 seconds to set up a channel in a multi site system. Low end trunked systems without a dedicated control channel and central controller can take 3 seconds to initiate a session. In systems where the set up time is so long, the channel is usually not released until after a multi call exchange (a session or conversation) has occurred. Typically this is accomplished by using a "hang-time" at the end of each call (talk spurt). While hang-time is effective to mitigate set up delay after the first call in a session, it is wasteful of channel resources.

5.5 Cost Estimates

All the hardware technology exists today to build the ad hoc network concept described above. The major missing links are software – i.e., writing code for network self configuration and routing. Based on hardware, and thus largely ignoring one-time development costs, an order of magnitude cost can be built up. Consider some typical retail prices, in US$: (i) dual band, dual mode satellite phone ~ $1,500-$3,000; (ii) high-end Pocket PC ~ $1,000; (iii) GPS receiver under ~ $150; (iv) integrated GPS and two way Family Radio Service (FRS) analog radio transceivers ~ $350; (v) wireless computer modems (e.g., IEEE 802.11b network interface cards) under ~ $100; and (vi) digital public safety grade mobile radio transceivers up to ~ $4,000.

These figures suggest that an ad hoc network user terminal should be feasible with a price target under U.S. $10,000. Of course, the market viability of this figure (i.e., would anyone buy it), depends on the cost of alternatives. Considering the cost of current public safety grade land mobile radio systems as the alternative, Table 5.3 presents a comparison. From the table, we conclude that, if a target of $10,000 per user terminal can be hit, we are in range.

Table 5.3
Cost estimate comparison.
All figures are in US$. Equivalent annual amount calculation is based on 10 year term and a 5% annual discount rate. Satellite usage is based on $30 per month per user terminal. The annual operating cost calculation is based on 8% of capital for infrastructure and 10% for user equipment.

Description	Capital Investment per User			Equivalent Annual Amount
	Infrastructure	User Equipment	Total	
Wide-area, public safety grade trunked radio				
High range	9,000	4,000	13,000	2,800
Typical	5,000	2,000	7,000	1,500
Low range	3,000	1,000	4,000	900
Ad hoc network (with satellite usage)		10,000	10,000	2,700

5.6 A Possible Phased Development Program

Considering the findings of this project, a long-term program could be initiated to develop a public safety grade mobile ad hoc network. The following phased approach is logical:

- Phase-1 Cluster nets: Develop the single hop cluster net capability that is centered on the Vehicle Radio. Note that this defines an air interface. The Vehicle Radio should be backbone radio equipped but the Phase-1 Personal Radio can be kept simple and need not be backbone capable.

- Phase-2 Satellite backbone net: Integrate the Phase-1 cluster nets with a satellite overlay. For simplicity, this may be on a GEO platform initially (it is duly noted that Mobile Satellite Ventures is currently scheduled to deploy a follow-on satellite series to supercede the current MSAT system over North America in the 2007 time frame). In the longer-term, a multi-tiered satellite layer may be feasible (e.g., using more than one constellation).

- Phase-3 Multi hop backbone: The ad hoc backbone can begin to be developed to the extent that a sound MAC protocol is identified. Using this MAC, the multi hop backbone can be evolved. If necessary, the backbone capability can be extended incrementally, starting with one or two inter cluster backbone links. Initially the multi hop backbone can focus on the Vehicle Radio.

- Phase-4 Full integration: This phase will provide full capabilities to the Personal Radio and fully integrate all elements. All failure modes should be addressed at this stage; for example, fall-back operations in the event that satellite systems are unavailable (including GPS location data).

Launching such a program would be non trivial. Among the challenges are:

- Funding for development
- Obtaining broad participation (for economies of scale and recall that performance depends on achieving sufficiently high user densities)
- Standards
- Multi vendor supply

In terms of process, a possible approach would be a government-led, multi-stage design competition (similar to military acquisitions such as for fighter aircraft). First round winners would be contracted (paid) to develop prototypes. The prototypes would compete in field trials. The winning design from the trials would be licensed to multiple vendors who would then compete on enhanced performance, product support, and price. Arguably the Association of Public Safety Communications Officials (APCO) could undertake to coordinate such an effort. In particular, it is noted that technology for Phase II of their Project 25 program is currently under consideration (APCO Project 25 Phase-1 was an initiative to develop standards-based, digital trunked radio systems for public safety).

Chapter 6
Conclusions

The operational requirements and context for public safety mobile communications features several key characteristics that can be exploited by an ad hoc network design. These include:

- The half duplex push-to-talk nature of the voice service, as it provides up to a second of channel set up time
- The prevalence of vehicles, as they have more tolerance for power consumption and better antenna mounting options (i.e., better that people)
- Location reporting capability using GPS, as it is economically available and being adopted now

Many of the key technologies needed for an ad hoc network design are available now – including all hardware elements. These include:

- Advanced radio technologies
- High-powered, small form factor computers
- In orbit satellites serving mobile communications now, and next generation are under active investigation

In this project, I developed a hypothetical example ad hoc based network design that focused on meeting the need to handle high-priority, real-time multicast voice traffic. The design proposes use of 4,900 MHz and 700 MHz for terrestrial links. The basic topology would be dominating-set based. The single hop cluster nets could use a form of the PRMA protocol. Likely the backbone routing protocol will be hierarchical in structure, use location assistance and be a hybrid scheme that uses both proactive and reactive algorithms. The use of a satellite tier is necessary for wide spread sparse networks and could prove useful to support location assisted terrestrial routing.

The estimated performance of the cluster net in the hypothetical design indicates the basic feasibility that next generation land mobile radio systems can begin adopting ad hoc network principles. With careful design, a robust and software evolvable user terminal platform should be possible.

Notwithstanding this conclusion, the following items summarize the major open issues that must be adequately addressed in pursuit of a mobile ad hoc network based communication system for public safety agencies:

- Sufficient spectrum
- Adaptation of applications and transport protocols – including IP compatibility
- Multi hop backbone – self configuring, routing protocol with support for real-time multicast voice
- Integration with a satellite tier – including use of joint satellite and terrestrial transmission approaches
- Integration with existing networks: (i) to exchange traffic for inter operability, (ii) to use as ad hoc backbone alternatives or supplements (for example, forming out-of-domain shortcuts or tunnels that cut across the fabric of the multi hop ad hoc network), and (iii) in particular, inter working with IP networks must be resolved
- Security – including resistance to denial of service attacks (for example, the black hole problem in AODV [Deng02])
- Network management – including consideration of graceful and stable failure modes

Addressing the routing related issues requires a better understanding of mobility patterns and the spatial distribution of traffic. Until the traffic characteristics are sufficiently understood, performance estimates should include the use of genuine traffic traces. The impact of vertical and horizontal propagation in high-rise buildings may require special study, as these connectivity patterns can affect traffic flow patterns. An appropriate busy period to use for design purposes also requires study.

Finally several authors have noted the need for integrated, cross-layer, design principles. Specifically, we need good ways to partition the overall design problem and structure the order of problem solving (which will necessarily be recursive) to efficiently find acceptable performance trade offs and compromises. Clearly much remains to be done.

Appendix A

Selected Terminology

Call	*A single, contiguous radio transmission; in transmission trunked radio systems, it is one talk-spurt.*
Cluster head	*The head of a cluster of nodes.*
Cluster net	*In this report, refers to the single hop ad hoc network cluster that forms around a dominating set member.*
FCC	*Federal Communications Commission in the US*
Fixed Radio	*An ad hoc network user terminal that is intended for deployment at a fixed location.*
GPS	*Global Positioning System*
IP	*Internet Protocol*
ITU-T	*International Telecommunications Union – Telecommunications Sector*
kb/s	*kilo bits per second*
kHz	*kilo Hertz*
LAN	*Local Area Network*
MAC	*Media Access Control*
Mb/s	*Mega bits per second*
MHz	*Mega Hertz*
MPT1327	*A trunked radio standard that originated in the UK (Ministry of Posts and Telecommunications)*
PDA	*Personal Digital Assistant*
Personal Radio	*An ad hoc network user terminal that is portable and carried by an individual user.*
PRMA	*Packet Reservation Multiple Access [Good89]*
PTT	*Push to Talk*
RF	*Radio Frequency*
Transportable Radio	*An ad hoc network user terminal that is unattended and autonomous, and easily re-deployable.*
VCL	*Virtual Cell Layout [Cayi02]*
Vehicle Radio	*An ad hoc network user terminal that is intended for mounting in a vehicle.*

Appendix B
Primer on Ad Hoc Networks

This appendix is based on a literature review and is partially drawn from a paper written and submitted by the author as part of the Computing Science course on the Theory of Communication Networks (CMPT-816-3, SFU) instructed by Dr. J. Peters in the 2000 Fall semester.

B.1 Overview

Ad hoc networks are self organizing networks where the user terminals form the infrastructure. The nodes are linked using unguided media – i.e., radio or light. There is no predetermined topology or control point and the nodes may randomly appear, disappear and move. Figure-B.1 illustrates the idea of topology formed by random connectivity as well as the notion of single hop and multi hop. It should be evident that the network protocols must be fully distributed.

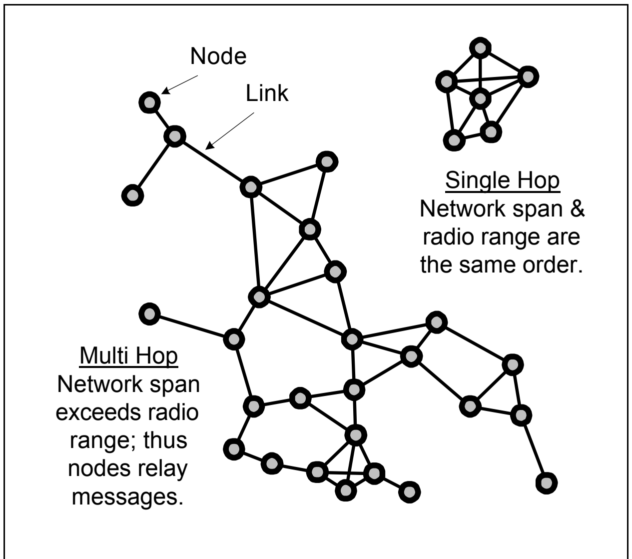

Figure-B.1 Topology from random connectivity.

Ad hoc networks can be isolated or interconnected to other networks. As shown in Figure-B.2, isolated means they are stand-alone, self contained networks that do not link to other networks. Interconnected means linked to one or more other networks as a "stub" – e.g., an access network to a fixed network such as the public Internet or telephone system.

The ability to self organize makes ad hoc networks adaptable, robust, and fault tolerant. And because no prior infrastructure is needed, a network can be rapidly deployed. These characteristics make ad hoc networks a good fit for classrooms, conference centers, office, and project environments, microsensor networks used to create smart spaces [Essa00], construction sites, emergency response, and disaster recovery communications, and, of course, the modern battlefield.

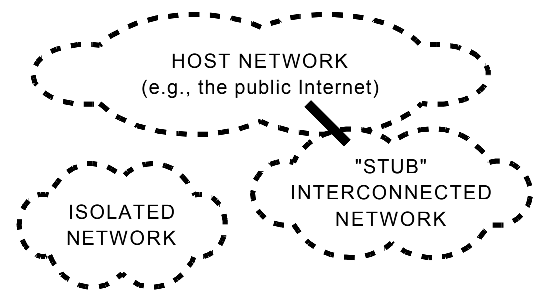

Figure-B.2 Isolated (stand-alone) & interconnected (stub) networks.

Mobile radio networks, including vehicular mounted and hand-held units, are classic examples of single hop ad hoc networks. A group of radios all tuned to the same channel communicate amongst themselves by transmitting to and listening on a common channel. Control of the network is resolved by simple etiquettes and protocols (over, over and out, etc.). These networks can be isolated (e.g., firefighters at an incident) or interconnected through a repeater site to a fixed network (e.g., a police team being dispatched).

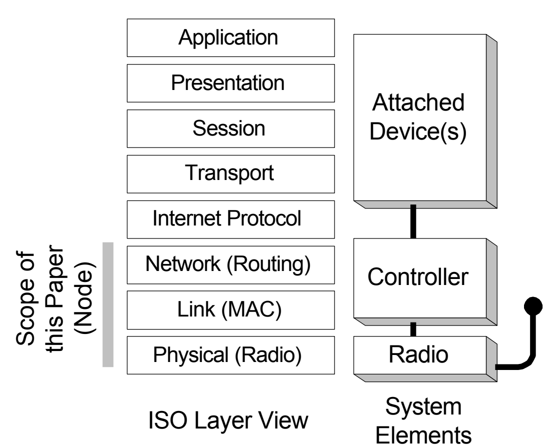

Figure-B.3 Anatomy of a node.

Generally, every node in a true ad hoc network must support the full suite of protocols. Figure-B.3 presents a layered view of a node along the lines of the ISO model. The upper layers (i.e., the attached devices) are generally assumed to generate both real-time traffic that is delay constrained (i.e., voice telephony and "multimedia") as well as best effort traffic (i.e., "data" or classic Internet). In this report, the focus is on the bottom three layers (i.e., the physical, link, and network). To foster economies-of-scale from mass production, the radio, and controller, which we generically call a node, should be identical (depending on the application, different terminal characteristics for the upper layers may be desirable for the attached devices).

The node requirements need to be defined in terms of the services to be supported and the environment in which the services will be delivered. This will lead to specifications for traffic levels and patterns, coverage (range), communications environment (e.g., topography, vegetation, noise, and multipath characteristics, etc.), performance (e.g., data rate, bit error rate, delay, etc.), management methodology[3], power (primary source and battery reserve), physical (size and weight constraints), and cost targets. From these requirements, the system designer will make decisions on frequency band, channel bandwidth, number of channels, frequency reuse, coding, interleaving, modulation, bit rate (and chip rate if spread spectrum), antenna type and configuration, MAC protocol, routing protocol, microprocessor type and speed, memory, power supply, and equipment packaging.

The following sections review each of the bottom three layers, and there is a brief concluding section on network management aspects.

B.2 Radio

The physical channel consists of a radio transmitter, the electromagnetic signal propagating through a noisy, fading, and dispersive medium, and the receiver. Figure-B.4 summarizes the various functions performed by the radio. The duplexing function is not shown – i.e., how the channel resource is shared for transmit and receive. The two basic methods are time division duplexing (TDD) and frequency division duplexing (FDD). Although both types are feasible for ad hoc networks, TDD is commonly used – it is simple, inherently spectrum efficient, and adapts to asymmetric traffic.

[3] The management methodology covers operational aspects such as performance monitoring, fault detection and reporting, and other administration, maintenance and support related aspects.

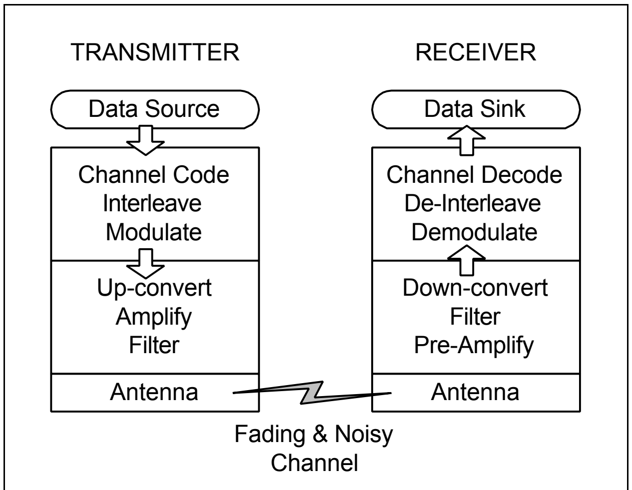

Figure-B.4 A simplified view of the functions performed by the radio subsystem.

The choice of a frequency band will profoundly affect and, in some cases, fix many other radio system parameters. For commercial ad hoc systems, the spectrum must be open to the public without licensing. This has drawn attention in the US market to the Industrial, Scientific, and Medical (ISM) bands, such as 902-928 MHz (26 MHz block), 2.4-2.4835 GHz (83.5 MHz block), and 5.725-5.875 (150 MHz block). Subsequently the FCC added 200 MHz to the spectrum at 5 GHz under the Unlicensed National Information Infrastructure (UNII) initiative. The UNII spectrum is in three 100-MHz blocks (5.15-5.25, 5.25-5.35, and 5.725-5.825 GHz). Each block has a different output power limitation governed by the equation: $A\ dBm + 10 \cdot log(B)$; where B is the -26 dB emission bandwidth in MHz and A is 4 dBm for the lower block, 11 dBm for the middle, and 17 dBm for the upper [Nee99].

The 2.4 GHz ISM band is of particular interest as many countries have opened at least part of this band for unlicensed applications. Consequently, it is no surprise that Bluetooth has selected this band for an open standard aimed at the world-wide market to replace cables between communication and computer devices [Haar00]. Of course others are also attracted to this band including the IEEE 802.11 wireless LAN standard and the European Telecommunication Standards Institute (ETSI) High Performance LAN (HiperLAN) [Nee99]. Needless to say, interference is a very real concern. To avoid high error rates (and/or reduced capacity), the effective range of each standard attempting to coexist in an area will be reduced. For example, while 802.11 under ideal conditions (i.e., line of sight and no interference) has a range of up to 100-300m and in typical office environments 30-50m, this will shrink to 10-20m in the presence of active Bluetooth devices [Zyre99]. And Bluetooth devices, which have a range of 10-30m under ideal conditions will need to be within 2m to maintain good performance in the presence of active IEEE 802.11 devices [Haar99].

The original IEEE 802.11 standard supported 1 and 2 Mbps data rates with 3 different physical layers – i.e., an infrared, a frequency hopping spread-spectrum (FHSS), and a direct sequence spread-spectrum (DSSS) based channel. Recently the DSSS standard has been extended[4] to support 5.5 and 11 Mbps rates (in the 2.4 GHz band). And a new 54 Mbps standard is being developed for the 5.2 GHz band using orthogonal frequency-division multiplexing (OFDM). These high data rate schemes are rate adaptive and fall back to lower modulation levels and code rates when propagation loss, dispersion or interference reduces the carrier to noise plus interference ratio [Nee99]. And, of course, for high mobility networks, there are lower data rates with simple binary modulation schemes such as Gaussian mean shift keying (GMSK) that is used in the General Packet Radio Service[5] standard.

Design trade-offs are complex and can affect other layers – for example, coding and modulation schemes can be combined using trellis modulation with Viterbi decoding; but the amount of forward error correction (FEC) applied should be coordinated with any automatic repeat request (ARQ) and other error control schemes used at higher layers.

Spectrum is the scarce resource and many uses of ad hoc networks are mobile and portable, and therefore need low power consumption. Thus spectrum efficiency and low power are key design goals. Future systems will use ever higher frequency bands; for example, 20-40-60 GHz are candidates. The higher frequencies have higher capacity and shorter range –a good trade-off in high user-density applications.

There is a lot of exciting development at the radio layer. Advances in digital signal processing are making low cost multi-mode and multi-band radios possible – with the ultimate goal, being a fully software defined radio [Reic99]. Another promising future technology is ultra-wideband radio which uses a form of pulse position modulation with spreading ratios that smear the signal across several GHz of spectrum [Shol93].

[4] All 4 DSSS data rates occupy compatible radio frequency bandwidths and the modulation and pseudo noise (PN) sequences are also compatible, allowing the higher data rates to fall back as necessary.

[5] General Packet Radio Service (GPRS) is an evolutionary step toward 3rd generation cellular from the Global System for Mobile communications (GSM) standard.

B.3 Medium Access Control

The Medium Access Control (MAC) or link layer manages the channel enabling adjacent nodes to communicate. Essentially the MAC protocol schedules transmissions by allocating channel resources (i.e., radio frequency bandwidth) such that flow control, fairness, and any quality of service requirements are met. The MAC protocol also has responsibility for coordinating FEC and ARQ error control techniques to meet link performance requirements; managing power (e.g., standby modes and sleep cycles); implementing a packet forwarding mechanism; and managing security (handling authentication and encryption).

In an ad hoc network, the MAC protocol must be distributed and designed to handle random traffic requests along with the random movement, appearance, and disappearance of adjacent nodes. To arbitrate requests for channel resources, contention based random multiple access broadcast channel schemes such as Aloha or carrier sense multiple access (CSMA) are generally in order. For situations where propagation delay is small compared to the transmission time, CSMA schemes are a good choice [Lein87].

Kleinrock provides a convincing analysis for this and also shows that unless the "hidden terminal" problem is dealt with, channel capacity suffers drastically [Klei76]. Figure-5 illustrates and explains the hidden terminal problem. In situations where hidden terminals may exist (as in an ad hoc network), a collision avoidance scheme is needed. The classic solution is to have receivers transmit a busy tone on a separate busy tone channel that nodes listen to before transmitting [Toba75]. This adds complexity and is spectrally inefficient, so various protocols based on a request to send (RTS) and clear to send (CTS) set up dialog have been proposed.

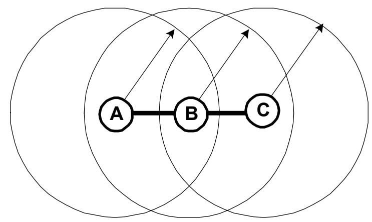

Figure-B.5 Hidden Terminal Problem: B is within range of A & C, but A & C are outside each others range. Thus, for example, if A is transmitting to B, then C will not be able to sense A's carrier and, if C transmits, its signal will interfere with A's transmission to B [Wu00].

RTS/CTS reduces the probability of a hidden terminal collision because the sender RTS includes a transmission length vector which is repeated by the receiver in its CTS. Listeners, hearing this exchange, avoid transmitting during the reserved period. However, problems occur under heavy load when the RTS packets begin colliding. There are different ways of implementing RTS/CTS based protocols – essentially these deal with the trade off between protocol overheads and the number of "hidden" terminals eliminated.

Interesting advanced MAC schemes include intelligent combining of the RTS/CTS and busy tone approaches with power control [Wu00] – this demonstrates not only improved channel utilization but also, through the use of power control, frequency reuse, and power consumption benefits are also realizable.

Quality of service support for delay constrained traffic, as may be imagined, requires some innovation over a CSMA based MAC. Several methods use the RTS/CTS dialog to reserve channel resources – for example, the Multiple Access Collision Avoidance / Packet Reservation (MACA/PR) scheme [Lin97]. These methods require each node to maintain information on channel state, but in return for this overhead, we can reserve contention free windows and guarantee bounds on link delay.

The black burst method proposed by Sobrinho and Krishnakumar [Sobr99] provides a means for real-time traffic to take priority over best effort data traffic without the need to maintain channel state. The compromise here, and for other prioritization schemes, is that delay bounds are statistical (not strictly guaranteed). Arguably, a more descriptive term for this scheme would have been the "squeaky wheel" protocol. The essence of the protocol is that nodes with real-time traffic, while waiting to access the channel increment their black burst timer (i.e., how long they will squeak). In the contention interval, all nodes with real-time traffic transmit their accumulated black burst (i.e., they squeak) and immediately after they listen to see if they had the longest black burst – which would indicate priority access (i.e., if the channel is silent then that node squeaked the longest, thus it has been waiting the longest and deserves priority).

The IEEE 802.11 wireless LAN standard is an interesting example [IEEE97]. The design assumes single hop connectivity (see Figure-B.1 above) between nodes in a basic service area (BSA) – specifically nodes must be close enough for single hop connectivity[6]. A group of nodes can form an ad hoc network (called IBSS for independent basic service set). The basic MAC protocol is called the Distributed Coordination Function (DCF) which uses slotted CSMA with RTS/CTS for collision avoidance. The IEEE 802.11 standard also handles packet acknowledgment at the MAC layer. After each data frame is sent, the receiver returns an acknowledgment (ACK) after waiting a "Short Inter Frame Space" (SIFS), whereas contention for new traffic must wait a longer "DCF Inter Frame Space" (DIFS) – effectively this gives ACK traffic priority (as shown in Figure-B.6). The DCF works as a fully distributed, autonomous MAC protocol unless it is in the presence of an access point (AP). An AP is an IEEE 802.11 base station which is part of the fixed infrastructure. If an AP is present, control becomes centralized at the AP (peer to peer communication ceases and client-server takes over).

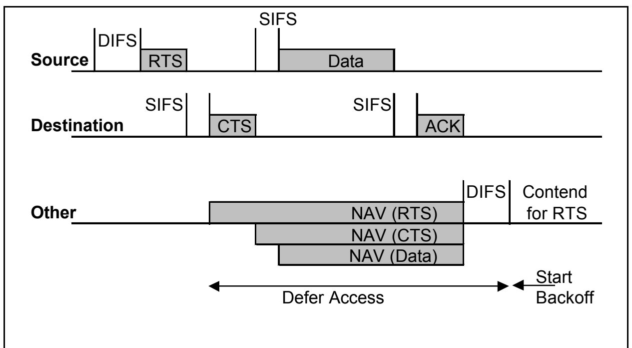

Figure-B.6 IEEE 802.11 DCF transmission with RTS/CTS (NAV is the Network Allocation Vector) [IEEE97], [Crow97].

To accommodate delay constrained traffic (e.g., real-time voice), there is an optional extension to DCF called the Point Coordination Function (PCF). The PCF is shown in Figure-B.7. PCF sets up a time division duplex (TDD) channel with contention free periods that are suitable for connection oriented services requiring delay guarantees. The PCF mode only works under the

[6] Note that single hop connectivity does not guarantee a fully connected network (complete graph) because propagation impairments, such as interference from frequency reuse, can make some terminals appear "hidden" from others.

control of an AP. Refer to [Crow97] for performance results from simulation of the IEEE 802.11 MAC.

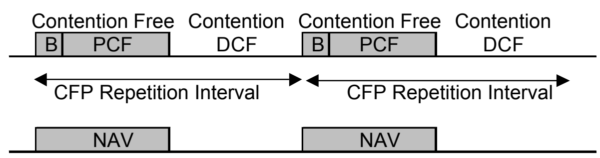

Figure-B.7 The PCF and DCF functions coexist by sharing the Contention Free Period (CFP) Repetition Interval [IEEE97], [Crow97].

Instead of a reservation scheme, like that used in IEEE 802.11, the European HiperLAN standard uses prioritization to statistically bound delay [Cors00]. This scheme is conceptually similar to the black burst approach described above, but wields an impressive name (and acronym), Elimination Yield – Non-Preemptive Priority Multiple Access (EY-NPMA).

To conclude this section, we highlight the need for an ad hoc wireless MAC to deal with the hidden terminal problem and note that, support for quality of service, will add further complexity. Also, as with all random access schemes, the MAC must ensure stability – i.e., avoid excess traffic from precipitating a sustained network failure. Refer to [Cors99] for a tutorial survey of wireless MAC protocols covering both ad hoc and centralized radio networks.

B.4 Routing Protocols

The routing protocol (network layer) manages connectivity for message delivery across the network. Obviously, this only becomes an issue for multi hop networks (see Figure-1) where node to node relaying is needed[7]. The routing protocol needs mechanisms for discovery (who are my neighbors and who is in this network[8]), exchange of routing and route control information, route computation (efficient algorithms to find routes that make efficient use of network resources), and addressing. And, for autonomous operation, the routing protocol must be distributed. The challenge is to deliver these functions with scalability (to large network sizes) while supporting delay constrained and multicast traffic in a bandwidth and power constrained

[7] IEEE 802.11, HiperLAN and Bluetooth standards define single hop networks; thus only the physical and link layers are specified.

[8] A converse of this is the routing protocol must allow the network to authenticate new nodes as eligible members.

environment where the topology is dynamic, physical security is limited, and congestion is often the norm [Cors99]. As we shall see in this section, this is a significant challenge.

Topology is, of course, a fundamental issue and was an area of early research. Kleinrock and Silvester reported on the relationship between routing and topology in [Klei87], which also surveys other relevant work. In a random structure, they report that the optimum number of neighbors (degree) needed to maximize throughput is an average of six to eight[9]. The optimum degree can be used to select transmit power. And this can be used to address the question of "giant stepping" versus baby stepping – is one big high-powered step better than a multitude of lower powered steps across the network. Another perspective on this, again by Kleinrock [Klei87], is to optimize the radio range based on minimizing the delay in an M/M/1 queue (hence the optimization parameters include arrival rates and the mean message transmission time – which is a function of capacity and mean message size).

Routing protocols can be classified as table-driven and on-demand. Table-driven are pro-active algorithms where each node globally pre-computes and maintains a routing table. On-demand are reactive algorithms that broadcast a routing query and routes are computed as they are needed. The dynamic nature of radio based ad hoc networks makes routing algorithms from wired networks unsuitable. In the next few paragraphs, we briefly review each of these routing protocol types from the perspective of ad hoc networks; then examine hybrid approaches that attempt to deal with controlling routing complexity for large scale networks. This section concludes with some of the open routing issues.

Table-driven routing involves (i) a discovery stage where each node identifies its neighbor (and may also gather other link state parameters such as capacity or quality[10]; (ii) each node then exchanges this information with all other nodes (essentially a network gossip); (iii) with global network state information at hand, each node builds some form of shortest path routing table; and, (iv) of course, the network state information must be maintained current, so some form of update routines are needed. Link-state routing protocols for wired networks, such as open shortest path first (OSPF) have high routing control traffic overhead (especially when network topology changes are frequent), and thus are unsuitable in a bandwidth constrained radio network. A distance vector protocol based on the classic Bellman-Ford routing algorithm by Perkins and Bhagwat [Perk94] called Dynamic Destination-Sequenced Distance-Vector routing protocol

---

[9] The degree of six to eight contrasts with three for a regular structure; this can be understood from the perspective that since neighbors come and go randomly, a higher average connectivity is required in the random structure to avoid risk of too little connectivity some of the time.

[10] Link quality metrics may be be signal level, bit error rate or signal to noise plus interference ratios, etc.

(DSDV) generates less control traffic but can be slow to converge with topology changes [John96]. Table-driven protocols, since routes are pre-computed, have the desirable feature that they have a low and generally fixed routing latency. To overcome the difficulty of these algorithms under conditions of highly dynamic topology, on-demand schemes are attractive.

On-demand routing, conceptually, involves the source node broadcasting a routing query that accumulates routing information as it searches outward. When the destination node receives the query, a routing reply packet back tracks to the source with the routing information – e.g., Dynamic Source Routing (DSR). Variations on this theme include the Ad hoc On-demand Distance Vector (AODV) protocol which distributes the discovered route as next hop routing information stored at each node [Perk99].

These protocols are simple and, since the need to maintain state is minimized or even eliminated, they scale well even in networks with highly dynamic topologies. The compromise, however, is high and variable path set up delay and, depending on how the protocol deals with the route request query, inefficiencies from flooding to find routes may occur.

From the above discussion, table-driven and on-demand routing protocols each have their strengths and weaknesses. Various hybrid schemes have been developed to capitalize on the advantages of each type. One such is the Zone Routing Protocol (ZRP) from Haas and Pearlman [Pear99].

The essence of ZRP is that a zone is defined for each node in terms of some number of nodes (for example, all nodes within "p" hops). Thus, as shown in Figure-B.8, the network graph is partitioned as a set of overlapping sub graphs. Routing within each zone can use some form of table-driven protocol. This is a particularly good choice if there is a significant amount of local traffic (intra zone) and/or if at the local level the network is relatively stable. Routing between zones (inter zone) can use some form of on-demand protocol. To improve the efficiency of the on-demand protocol, the query is "bordercast". However, because the zones overlap heavily, the protocol uses techniques to keep the query moving outward and avoid inadvertently flooding the network. Considering that we are channel capacity constrained and routing control traffic subtracts from useful throughput, the ZRP can be implemented to adaptively seek an optimum zone radius that minimizes routing control traffic. This is shown conceptually in Figure-B.9, which simply shows that as zones get bigger, the total intra zone routing control traffic increases and conversely as zones get smaller, the inter zone routing control traffic increases.

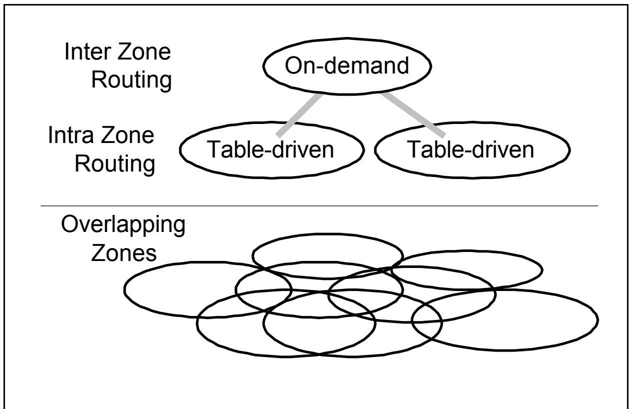

Figure-B.8 For the Zone Routing Protocol (ZRP), the network space is divided into overlapping zones [Pear99].

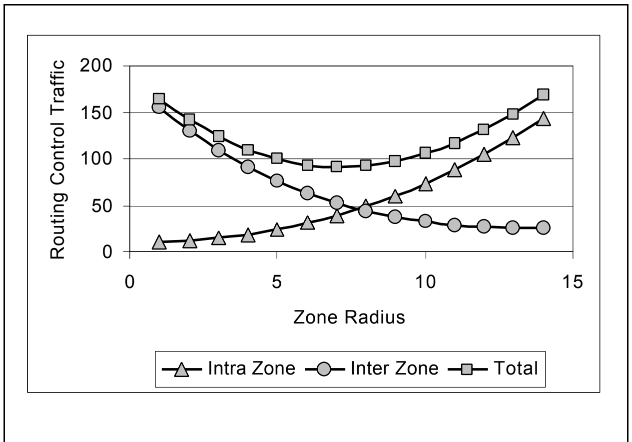

Figure-B.9 The concept for finding an optimal zone radius for ZRP [Pear 99].

Refer to [Pear99] for ZRP performance simulation results for networks with 200 to 1,000 nodes, zone radius of up to eight hops and node density (degree) up to nine.

The ZRP also illustrates the classic notion of hierarchy to deal with complexity as the network scales[11]. Other hybrid and hierarchical schemes have been proposed, including some with non overlapping zones [Joa99]. Refer to [Misr00] for a high level survey of routing protocols for ad hoc networks and [Lee99] for the results of simulations comparing table-driven and on-demand protocols in highly dynamic networks. In any event, scalable routing protocols for very large and/or very dynamic networks is still an active research area. Extending routing protocols to efficiently support quality of service and multicast, which are needed for handling real-time and multimedia traffic, are also open research problems.

[11] An early hierarchical ad hoc network was the U.S. Navy's High Frequency Intra Task Force Network [Ephr87].

Quality of service guarantees essentially require maintaining state in the network – an onerous task for a distributed routing protocol when the topology is dynamic and, thus, continuously rendering available state information incorrect. Delay constrained least-cost routing and bandwidth-constrained least-cost routing have been studied by Chen and Nahrestedt [CheS99]. The delay constrained problem is NP-complete and the bandwidth-constrained problem requires precise state information to solve in polynomial time. An example of a current proposal for quality of service routing in an ad hoc network is the Core Extraction Distributed Ad hoc Routing (CEDAR) protocol [Siva98].

The multicast problem is likewise difficult; specifically from [Sheu99] finding a multicast tree in a mobile network that minimizes the probability of multicast packet loss is NP-complete. This same paper proposes a heuristic algorithm called Degree Based Multicast Routing Algorithm (DBMRA).

In closing this section, it is useful to highlight some of the characteristics and performance metrics for MANET routing protocols as identified in [Cors98]:

- Network size (number of nodes)
- Network connectivity (average degree)
- Topological rate of change
- Link capacity
- Traffic patterns
- End-to-end throughput and delay
- Route acquisition time
- Average number of data bits transmitted versus data bit delivered
- Average number of control bits transmitted versus data bit delivered

Of particular importance is the need for robustness and for the performance to be acceptable and predictable across a wide range of expected traffic loads and operating conditions.

B.5 Network Management System

Our discussion of ad hoc networks would be incomplete without at least brief consideration for how, during actual operations, network availability and performance can be maintained and users supported. Consistent with ISO terminology, and using a bit of license from Minoli, the following the network management functions are identified [Mino91]:

- Fault management
- Accounting management

- Configuration and asset management
- Performance management
- Security management
- Network planning management – collecting, analyzing and applying usage and performance trends for capacity and coverage planning
- Programmability management - allows users to customize the network management system

The need to automate network management tools for ad hoc networks was recognized early as an area requiring attention [Shac87]. Refer to [CheW99] for a network management protocol developed specifically for ad hoc networks.

Although all of the above management functions are important, performance management holds a special place as it relies on, and is required by, many of the other functions. Performance management necessitates measurement and parameters measured should be from the user's perspective. The following are drawn from [Lein87]:

- System availability
- Delay (end-to-end)
- Throughput
- Coverage
- Accuracy (error rate)
- Security (breaches)

Network planning management is also of particular importance for a truly autonomous and self-organizing network. For example, frequency assignments should be dynamic and not require any intervention (IEEE 802.11 requires frequency selection). For some applications, for example, access networks, a default "meet and greet" frequency may be needed[12]. Related to frequency coordination, are interference avoidance mechanisms such as power control and beam forming antennas – all of which must be adaptive and operate autonomously.

B.6 Enabling Technologies

The following are seen as technologies and initiatives that are expected to contribute to the development and emergence of viable ad hoc systems:

- self organizing systems
- software defined radio
- miniaturization

[12] The meet and greet frequency would be the equivalent of Channel 9 on Citizens Band radio.

- battery technology
- adaptive (smart) antennas
- joint detection and multiple input / multiple output (MIMO) schemes
- user terminal evolution
- new frequency bands
- next generation satellite systems

Research on self organizing systems is progressing on a wide front, across many disciplines and fields of interest. Although work through the IETF MANET working group [IETF00] may be the most relevant today, the principles of self organization from other fields can be expected to be broadly applicable.

Software defined radio technology was noted earlier. The viability of software radio depends on progress in the areas of programming, digital signal processing as well as continuing advances in memory and processors (Moore's Law). Software defined radios are expected to (i) lower costs (one hardware platform will satisfy a wide range of applications), (ii) overcome air interface incompatibilities (radios can adapt to whatever service is available), and (iii) make radios dynamically adaptable and upgradeable (e.g., as improved protocols become available). Multi mode and multi band cell phones are progenitors of future software defined radios.

Continuing evolution of user terminals, including cell phones, personal digital assistants (PDAs), and pocket PCs, should open up new applications and lead to higher user densities. This combined with increasing onboard processing power and memory should help create an environment to support the emergence of ad hoc networks.

New frequency bands are being opened to wireless technology – especially for 3G cellular and broadband wireless local. Additional spectrum, or even a surplus, may be a good thing as the spectral efficiency of multi hop ad hoc networks may not be particularly high (MAC, routing and transport overheads for QoS and multi casting in multi hop ad hoc networks add up quickly).

Next generation satellite systems are listed because truly wide area ad hoc networks will likely have some form of hierarchy and the higher capacity and lower delay of the proposed next generation satellite systems may have a role to play. Teledesic is arguably the most visible of the next generation hopefuls. Current plans target commercial service in 2005 [www.teledesic.com] based on a 288 satellite constellation and operating in the Ka band (28.6-29.1 GHz up and 18.8-19.3 GHz down). Mobile Satellite Ventures (MSV), the operator of a mobile satellite (MSAT) system over North America with multicast (push to talk) service, is planning to deploy a next generation satellites in the 2007 timeframe.

B.7 Outstanding Issues

Although a lot has been accomplished toward realizing multi hop ad hoc networks and commercial market pioneers are appearing (particularly in fixed markets), much remains to be done. Specifically, major issues include:

- Routing protocols
- Frequency assignments and standards
- Security
- Interconnection
- Regulations
- Market model
- Network management

As discussed earlier, the routing protocol is the most difficult outstanding technical problem. Decisions at the routing protocol layer will likely require compromises at the MAC and radio layers; particularly to support quality of service and multicast. And, for example, taking advantage of smart antenna technology (e.g., to improve spectrum efficiency) requires adaptation at both the MAC and routing layers. Internet Protocol (IP) and IP derived or compatible routing protocols may enable flexibility and the formation of multiple logical overlay networks (multi graphs) to be formed [Cors99].

The standards issue is related to the regulatory issue. Standards are essential to form the mass markets needed to drive prices to commodity levels[13], and therefore, industry cooperation and vendor forums are pivotal.

Another aspect of standards is co-existence versus integration. For example, Bluetooth and IEEE 802.11 coexist in that the degradation is generally considered tolerable. Ideally standards would be integrated such that nodes conforming to different standards would recognize each other and adapt optimally. Hopefully, the IEEE 802.15 working group on personal area networks, will be able to integrate Bluetooth and wireless LANs in the next generation.

[13] Classic examples of standards creating large markets are fax machines and the Internet. And closer to our topic, note that wireless LANs have been around for a decade but until standards were agreed, wireless LANs remained a niche market.

Appendix C
Primer on Land Mobile Radio Systems

In broad terms and considering the thrust of the project, we can usefully partition mobile radio based systems into the following categories:

- Simplex mobile radio systems
- Conventional repeater systems
- Trunked repeater systems
- Cellular mobile telephone systems

Figure-C.1 summarizes key characteristics of these systems and each is briefly explained in the following paragraphs.

(a) Mobile Radio Systems			
Parameter	Simplex	Conventional	Trunked
Band (MHz)	150 / 450 / 800		
Range (km)	30-60 / 15-30 / 5-30		
Capacity (kbps)	1.2-9.6		

(b) Cellular Telephone Systems				
Parameter	2nd Generation		3rd Generation	
Band (MHz)	800	1900	1500-2000	
Range (km)	5-30	1-10	2.5	0.1
Capacity (kbps)	9.6-19.2		144	2048

Figure-C.1 Mobile (a) and cellular (b) system characteristics. Ranges are typical and ideal (maximum) distances. 3G adapted from [Kall01].

C.1 Simplex Mobile Radio Systems

Simplex mobile radio systems are simply a collection of radio transceiver units that are tuned to a common frequency. Any radio unit can communicate with the others in the group by simply transmitting on the common channel – forming a classic single hop, isolated, ad hoc network. Units can move about, join and leave the network without relying on any outside infrastructure. Often the radio units are frequency agile and capable of selecting from a variety of different frequencies. This allows large groups to break up into separate talk groups on different frequencies. Operation is simplex – i.e., at any point in time a radio can either transmit or receive, but not both at once. Alerting someone or the group of a message is accomplished by simply transmitting an announcement (voice signaling). Examples include Citizen Band (CB) radio, Family Radio Service (FRS), public safety agencies at an incident (fire, rescue), and workers at a construction site. The first two examples use unlicensed spectrum and the latter two normally use licensed spectrum (i.e. users have exclusive use which protects them from interference by others). Most systems are frequency modulated (FM) and compatibility for inter working is limited to frequency channel allocations.

C.2 Conventional repeater Systems

Conventional repeater mobile radio systems are an extension of the above where each radio transmits on one frequency and receives on another. As shown conceptually in Figure-C.2, this allows user radios to talk through a repeater radio station that receives on the user's transmit frequency and retransmits the call on the users' receive frequency. By putting the repeater on a hill, the system range can be extended – or several repeaters can be linked together to form an even wider area system. With two frequencies in use, the operation is described as half duplex[14]. Note that the ad hoc network component – the user radios – are now single hop ad hoc stub networks off the repeater site infrastructure. In other respects, operation is similar to the simplex system above – e.g., FM is the standard modulation scheme and each radio channel generally forms a talk group. Note that each talk group needs at least two frequencies. Over the years, paging tones and decoders have been added to enable other features, for example, to selectively signal one or a subset of the user radio units.

[14] For simplicity and to avoid the cost of filtering, the user radio units are usually antenna switched so they can either transmit or receive, but not both at once.

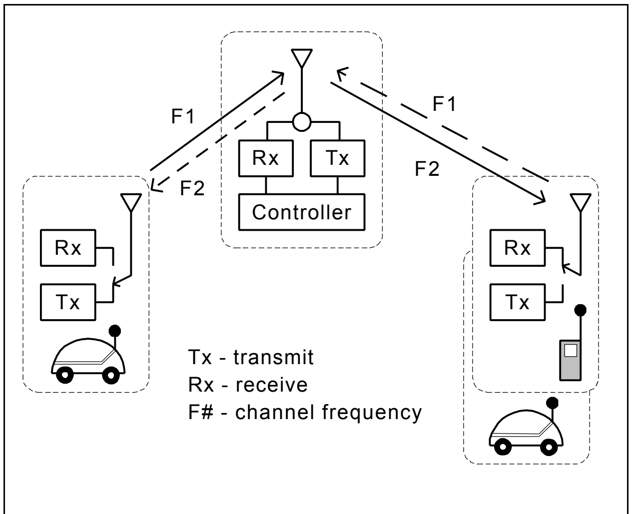

Figure-C.2 Mobile radio repeater concept.

C.3 Trunked Repeater Systems

Trunked repeater mobile radio systems take the above repeater approach one step farther, by letting a large number of talk groups dynamically share a smaller group of radio channels. This adds considerable complexity as we now need a means of taking requests and making channel assignments (a control subsystem). The benefit is improved spectral efficiency. Today, due to spectrum shortages, these types of systems are used in many large metropolitan areas. Control systems tend to be proprietary so radios built by one manufacturer will not work on other systems. Although there are various standards and standardization efforts, progress has been slow and limited[15]. The supply of systems is fragmented, although in North America most large trunked systems are supplied by Motorola and M/A-Com (formerly Com-Net Ericsson, which was formerly Ericsson which was formerly Ericsson/GE). A recent product from M/A-Com that is an interesting break from trunked radio tradition is an IP and packet based system developed as an extension to the Cellular Digital Packet Data (CDPD) protocol [MACo01].

[15] The LTR system from EF Johnson, which is an analog trunked system, has become a de facto standard through third-party licensing. There are European trunked radio standards including MPT 1327 for analog systems (which has so many incompatible variants that the purpose of this standard is defeated) and TETRA for digital systems. The TETRA system, although nearly a decade old, is slowly gaining ground and a variant may reach North America. Attempts by the Association of Public-safety Communication Officials (APCO), over a period of many years to establish a digital trunked radio standard for public safety agencies in North America, has led to mixed results with no accepted universal standard as yet.

C.4 Cellular Mobile Telephone Service

Cellular mobile telephone service has become a runaway success in the telecommunications world and, after nearly 25 years, it is poised to outstrip the telephone system that it was designed to augment. Taking a standards based approach, these systems led to vast commercial markets and consequently very low unit costs. Today, "second generation" digital cell phones, are being joined by higher data rate versions (2.5G) and over the next five years we may see the emergence of a yet higher data rate third generation (3G). Many of the recent enhancements to mobile radio systems have been trickle down from cell phone technology. Some of the second generation standards include extensions that provide talk group functionality similar to (or indistinguishable from) mobile radio systems. These extensions are available for the Global System for Mobile communications (GSM) as well as the North American digital standards. Finally, there is a proprietary cellphone and mobile radio system from Motorola that also blurs service boundaries. This is the iDEN product that has been deployed extensively in North America by Nextel in the U.S. and Clearnet in Canada (now Telus Mobility).

List of References

[Abra73] N. Abramson, *"The Aloha System" in Computer Communications Networks.* F. Kuo, Ed., Englewood Cliffs, NJ: Prentice-Hall, 1973.

[Blaz01] L. Blazevic, "Self organization in mobile ad hoc networks: the Terminodes approach," *IEEE Communications Magazine,* vol. 39, no. 6, pp. 166-174, June 2001.

[Cayi02] E. Cayirci and C. Ersoy, "Application of 3G PCS technologies to rapidly deployable mobile networks," *IEEE Network*, vol. 16, no. 5, pp. 20-27, September/October 2002.

[CheS99] S. Chen and K. Nahrestedt, "Distributed quality-of-service routing in ad hoc networks," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 8, pp. 1488-1505, August 1999.

[CheW99] W. Chen, N. Jain, and S. Singh "ANMP: Ad hoc Network Management Protocol," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 8, pp. 1506-1531, August 1999.

[Cohe83] P. Cohen et al., "Traffic analysis for different classes of land mobile comm systems," *IEEE Vehicular Technology Conference,* pp. 283-285, 1983.

[Cors99] S. Corson and J. Macker, "Mobile Ad hoc Network (MANET): Routing Protocol Performance Issues and Evaluation Considerations," *IETF RFC 2501*, [Online]. Available: http://www.ietf.org/rfc/rfc2501.txt, January 1999.

[Crow97] B. Crowe et al., "IEEE 802.11 wireless local area networks", *IEEE Communications Magazine,* vol. 35, no. 9, pp. 116-126, September 1997.

[Deng02] H. Deng, Wei Li and D.P. Agrawal, "Routing security in wireless ad hoc networks," *IEEE Communications Magazine*, vol. 40, no. 10, pp. 70-75, October 2002.

[Ephr87] A. Ephremides et al., "A design concept for reliable mobile radio networks with frequency hopping signaling," *Proceedings of the IEEE*, vol. 75, no. 1, pp. 56-73, January 1987.

[Essa00] I. Essa, "Ubiquitous sensing for smart and aware environments," *IEEE Personal Communications,* vol. 7, no. 5, pp. 47-49, October 2000.

[FCC02] Federal Communications Commission, WT Docket No. 00-32, Washington, D.C. 20554, Rule Making, adopted February 14, 2002; released February 27, 2002. [Online]. Available: http://www.imsasafety.org/wireless/docket0032.pdf.

[Gold02] A. Goldsmith and S. Wicker, “Design challenges for energy-constrained ad hoc wireless networks,” *IEEE Wireless Communications*, vol. 9, no. 4, pp. 8-27, August 2002.

[Good89] D. Goodman et al., “Packet Reservation Multiple Access for local wireless communications,” *IEEE Transactions on Communications,* vol. 37, no. 8, pp. 885-890, August 1989.

[Gumm00] A. Gummalla and J. Limb, “Wireless medium access control protocols,” *IEEE Communications Surveys,* [Online]. Available: http:comsoc.org/ pubs/surveys, 2nd Quarter 2000, pp. 2-15.

[Haar99] J. Haartsen and S. Zurbes, “Bluetooth voice and data performance in 802.11 DS WLAN environment,” *Bluetooth SIG Publication*, May 31, 1999.

[Haar99] J. Haartsen, “The Bluetooth radio system,” *IEEE Personal Communications*, vol. 7, no. 1, pp. 28-36, February 2000.

[Heer92] S. Heeralall, “Discussion of spectrum efficiency and the factors that affect it,” *IEEE CWC Conference Proceedings*, pp. 413-416, 1992.

[Hema01] S. Hemami, “Robust image communication over wireless channels,” *IEEE Communications Magazine*, vol. 39, no. 11, pp. 120-124, November 2001.

[Hess81] G. Hess and J. Cohn, “Communications load and delay in mobile trunked systems,” *IEEE Vehicular Technology Conference*, pp. 269-273, 1981.

[Hess93] G.C. Hess, *Land-Mobile Radio System Engineering.* USA: Artech House, 1993.

[Huba01] J. Hubaux et al., “Toward self-organized mobile ad hoc network terminals: the Terminodes project,” *IEEE Communications Magazine*, vol. 39, No.1, pp. 118-124, January 2001.

[IEEE97] IEEE Standard for Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications, IEEE Standard 802.11, 1997.

[IETF00] Internet Engineering Task Force Manet Working Group Charter [Online]. Available: (December 2002): http://www.ietf.org/html.charters/manet-charter.html.

[Jian01] M. Jiang et al., ``Impact of self-similarity on wireless network performance,'' in *Proceedings International Conference on Communications*, Helsinki, Finland, June 2001.

[Joa99] M. Joa-Ng and I. Lu, “A peer-to-peer zone-based two-level link state routing for mobile ad hoc networks,” *IEEE Journal on Selected Areas of Communications,* vol. 17, no. 8, pp. 1415-1425, August 1999.

[John96] D. Johnson and D. Maltz, "Protocols for adaptive wireless and mobile networking," *IEEE Personal Communications,* vol. 3, no. 1, pp. 34-42, February 1996.

[Kall01] J. Kalliokulju et al., "Radio access selection for multistandard terminals," *IEEE Communications Magazine,* vol. 39, no. 10, pp. 116-124, October 2001.

[Klei76] L. Kleinrock, *Queuing Systems, Volume 2: Computer Applications.* New York: John Wiley & Sons, 1976.

[Klei87] L. Kleinrock and J. Silvester, "Spatial reuse in multihop packet radio networks," *Proceedings of the IEEE,* vol. 75, no. 1, pp. 156-167, January 1987.

[Kost98] T. Kostas et al., "Real-time voice over packet-switched networks," *IEEE Network,* pp. 18-27, January/February 1998.

[Lee95] W. Lee, *Mobile Cellular Telecommunications.* New York: McGraw-Hill, 1995.

[Lee99] S. Lee et al., "A simulation study of table-driven and on-demand routing protocols for mobile ad hoc networks," *IEEE Network,* vol. 13, no. 4, pp. 48-54, July/August 1999.

[Lein87] B. Leiner et al., "Issues in packet radio network design," *Proceedings of the IEEE,* vol. 75, no. 1, pp. 6-20, January 1987.

[Lela94] W. Leland et al., "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, February 1994.

[Lin97] C. Lin and M. Gerla, "Asynchronous multimedia multihop wireless networks," in *Proceedings from IEEE Infocom 1997*, Kobe, Japan, pp. 118-125, 1997.

[Lou02] W. Lou and J. Wu, "On reducing broadcast redundancy in ad hoc wireless networks," *IEEE Transactions on Mobile Computing,* vol. 1, no. 2, pp. 111-122, April-June 2002.

[MACo01] M/A-Com OpenSky Web Site, www.opensky.com (visited 2001 10).

[Mino91] D. Minoli, *Telecommunications Technology Handbook.* Boston: Artech House, 1991.

[Misr00] P. Misra, "Routing protocols for ad hoc mobile wireless networks," [Online]. Available: http://www.cis.ohio-state/~jain/cis788-99/adhocrouting/index.html, July 2000.

[Nee99] R. van Nee et al., "New high-rate wireless LAN standards," *IEEE Communications Magazine*, vol. 37, no. 12, pp. 82-88, December 1999.

[Nguy01] T. Nguyen et al., "Voice over IP service and performance in satellite networks," *IEEE Communications Magazine,* vol. 39, no. 3, pp. 164-171, March 2001.

[Pear99] M. Pearlman and Z. Haas, "Determining the optimal configuration for zone routing protocol," *IEEE Journal on Selected Areas of Communications,* vol. 17, no. 8, pp. 1395-1414, August 1999.

[Perk94] C. Perkins and P. Bhagwat, "Highly dynamic destination-sequenced distance-vector (DSDV) routing for mobile computers," *Computer Communications Review,* pp.234-244, October 1994.

[Perk99] C. Perkins et al., "Ad Hoc On-demand Distance Vector Routing," IETF Draft, [Online]. Available: http://www.ietf.org/internet-drafts/draft-ietf-manet-aodv-04.txt, October 1999.

[Rapp96] T. Rappaport, *Wireless Communications Principles and Practice*. New Jersey: Prentice Hall, IEEE Press, 1996.

[Reic99] S. Reichhart et al., "The software radio development system," *IEEE Personal Communications,* vol. 6, no. 4, pp. 20-24, August 1999.

[Shac87] N. Shacham and J. Westcott, "Future Directions in packet radio architectures and protocols," *Proceedings of the IEEE,* vol. 75, no. 1, pp. 83-99, January 1987.

[Sheu99] S. Sheu et al., "Efficient multicast routing and backup algorithm in ad-hoc wireless networks," *IEICE Trans Fundamentals,* vol. E82-A, no. 7, pp. 1250-1260, July 1999.

[Shol93] R. Sholtz, "Multiple access with time-hopping impulse modulation," *IEEE MILCOM '93*, Boston, MA, October 11-14, 1993.

[Siva98] R. Sivakumar et al., "Core extraction distributed ad hoc routing (CEDAR) specification," *IETF Draft* [Online]. Available: http://timely.crhc.edu/Papers/idraft98.cedar.txt.

[Siwi02] K. Siwiak, "Ultra-wide band radio: a new PAN and positioning technology," *IEEE VTS News,* vol. 49, no. 1, pp. 4-9, February 2002.

[Sobr99] J. Sobrinho and A. Krishnakumar, "Quality-of-service in ad hoc carrier sense multiple access wireless networks," *IEEE Journal on Selected Areas in Communications,* vol. 17, no. 8, August 1999.

[Stae00] D. Staehle, K. Leibnitz, and P.Tran-Gia, "Source traffic modeling of wireless applications," *University of Wurzburg, Institute of Computer Science, Research Report Series,* Report No. 261, June 2000.

[Stee92] R. Steele (Editor), *Mobile Radio Communications*. New York: John Wiley & Sons, IEEE Press, 1992.

[Stoj02] I. Stojmenovic (Editor), *Handbook of Wireless Mobile Networks and Mobile Computing.* New York: John Wiley & Sons, 2002.

[Ston96] G. Stone, "Appendix D of SRSC final report, public safety wireless communications user traffic profiles and grade of service recommendations," *US Department of Justice*, 13 March 1996.

[TIA96] Telecommunications Industry Association, *A Report on Technology Independent Methodology for the Modeling, Simulation and Empirical Verification of Wireless Communications System Performance in Noise and Interference Limited Systems Operating on Frequencies Between 30 and 1500 MHz*, TIA TR8 Working Group 8.8, IEEE Vehicular Technology Society Propagation Committee, April 1996.

[Toba75] F. Tobagi and L. Kleinrock, "Packet switching in radio channels: Part II - the hidden terminal problem in carrier sense multiple access modes and the busy tone solution," *IEEE Transactions on Communications,* vol. 23, pp. 1417-1433, 1975.

[Tsir01] A. Tsirigos and Z. Haas, "Multipath routing in the presence of frequent topological changes," *IEEE Communications Magazine,* vol. 39, no. 11, pp. 132-138, November 2001.

[Wu00] S. Wu et al., "Intelligent medium access for mobile ad hoc networks with busy tones and power control," *IEEE Journal on Selected Areas of Communications,* vol. 18, no. 9, September 2000.

[Zyre99] J. Zyren, "Reliability of IEEE 802.11 hi rate DSSS WLANs in a high density Bluetooth environment," *Bluetooth '99*, June 8, 1999.