An Architecture for Mobile Radio Networks with Dynamically Changing Topology Using Virtual Subnets

Jacob Sharony

Hazeltine Corporation
Greenlawn, NY 11740
sharony@hazeltine.com

Abstract

An architecture adaptable to dynamic topology changes in multi-hop mobile radio networks is described. The architecture partitions a mobile network into logically independent subnetworks. Network nodes are members of physical and virtual subnets and may change their affiliation with these subnets due to their mobility. Each node is allocated an address based on its current subnet affiliation. We observe – especially in large networks with random topology – that partitioning of the network may result in significantly more balanced load than in one large multi-hop network, an attribute that can significantly improve the network's performance. The architecture is highly fault-tolerant, has a relatively simple location updating and tracking scheme, and by virtue of its load balancing feature, typically achieves a network with relatively high throughput and low delay. The addressing method, logical topology, mobility management and routing procedure are described, and network performance is evaluated.

1 Introduction

Mobile radio networks are expected to play an important role in future commercial and military applications, especially when a wired backbone network does not exist. These networks are suitable in situations where instant infrastructure is needed and no central system administration (like base stations in a cellular system) is available. Typical applications for this type of peer-to-peer networks include: mobile computing in remote areas, tactical communications, law enforcement operations and disaster recovery situations. A peer-to-peer mobile radio network consists of a collection of mobile packet radio nodes that create a network on demand without administrative support and may communicate with each other via intermediate nodes in a multi-hop mode, i.e., every node is a router. A critical issue in these networks is their ability to adapt well to dynamic topology changes caused by movement of nodes relative to other nodes in the network. Adaptation to topology changes requires changes both in channel assignment and routing.

Mobile radio networks have been around since the 1970's. Traditionally, these networks have aimed at providing classic data services such as file transfer. Recently, there has been a growing interest in rapidly deployable and dynamically reconfigurable wireless networks supporting multimedia traffic (voice, video and data). For these networks to support emerging multimedia services, carrying real-time data with stringent time delay constraints (voice, video), several important networking issues need to be resolved. Networks having high performance and reliability, e.g., high throughput, low delay and fault-tolerance, are desired.

Previous work on mobile radio networks with dynamically changing topology concentrated primarily on channel access and/or routing schemes in arbitrary physical topologies [2] – [7], [9], [11]. To improve network performance and reliability, several methods of topology control (by adjusting transmission ranges) were proposed [8], [10], [12], [13], [16]. More recently, a multi-cluster architecture for multi-hop mobile radio networks supporting multimedia traffic was proposed [6].

The motivation behind our approach is that *network partitioning* can improve critical functions such as media access, routing, mobility management, virtual circuit set-up, while reducing signaling/control overhead. It can be observed in this type of network that partitioning may result also in lower congestion compared to one large network.

This paper discusses an architecture based on a specific *logical* topology superimposed over a physical topology (determined by transmission coverage of network nodes); the architecture selects links to be activated (logical links) out of a pool of physical links. Our main concern is finding an efficient logical topology and a suitable routing procedure which result in high performance and reliability.

The paper describes an architecture suitable for mobile radio networks which is adaptable to dynamic topology changes due to node mobility. In this architecture, network nodes are grouped into two types of clusters (subnets): physical and virtual, and may dynamically change their affiliation with these subnets due to their mobility. Each node is allocated an address based on its current subnet affiliation as a result of its position relative to other nodes (i.e., its physical connectivity) and address availability. We consider networks that have several tens to several thousands mobile nodes. It is assumed that there exists a channel access protocol which resolves contentions and/or interference in the network (e.g., [1], [15]).

The rest of the paper is organized as follows. Section 2 describes the addressing method. Section 3 describes the network logical topology, explaining the formation of physical and virtual subnets, and discusses mobility management. Section 4 considers the routing procedure, and in Section 5 an example of a 16-node network is described. Section 6 discusses performance issues, and conclusions are given in Section 7.

2 Addressing method

In this method, network nodes are allocated addresses depending on their current physical connectivity and address availability. In general, each node is assigned a single address. However, in some conditions, a node can have more than one address as described later.

Assume that the network is segmented into p physical subnets (the distinction between physical and virtual will be clarified shortly; for the moment assume that physical subnets cover a local area) each containing up to q mobile nodes. We describe the pool of addresses over an alphabet of size $m = \max(p, q)$ containing the numbers $0, 1, 2, \cdots m-1$. Each node in the network is given a word (address) of length two, where the list significant digit (LSD) is a digit in base-q and the most significant digit (MSD) is a digit in base-p. Therefore, the total number of words (and nodes) possible is $N = pq$. Each node in this topology is affiliated with nodes whose address differs only in one digit; that is, node $x_1.x_0$ is affiliated with nodes $x_1.x'_0$, $0 \le x'_0 \le q-1, x'_0 \ne x_0$, and with nodes $x'_1.x_0$, $0 \le x'_1 \le p-1, x'_1 \ne x_1$. Thus, every node is affiliated with $p+q-2$ other nodes; we say that each node has $p+q-2$ *logical neighbors*. Next we group every q nodes that differ only in their LSD into an MSD group, and every p nodes that differ only in their MSD into an LSD group. Note that there are altogether $p+q$ groups, and each node is a member of one LSD group and one MSD group. These groups are the basic building blocks of the network as described in the next section.

3 Logical topology

Each node in the network is affiliated with a physical subnet (MSD group) and a virtual subnet (LSD group). Nodes which are members of a physical subnet are within close proximity in a local geographic area. Nodes which are members of a virtual subnet form a regional network (i.e., beyond a local area). Figure 1 depicts a mobile radio network with physical subnets (in shaded areas) and virtual subnets (e.g., in solid and dashed lines). Note that all nodes within a physical subnet have the same MSD while all nodes within a virtual subnet have the same LSD. It is assumed for the moment that nodes of a given physical subnet can reach (e.g., by adjusting their transmission power or by using a directional antenna) nodes of neighboring physical subnets. Later we deal with the case when this assumption does not hold (i.e., when a node is disconnected from its virtual subnet).

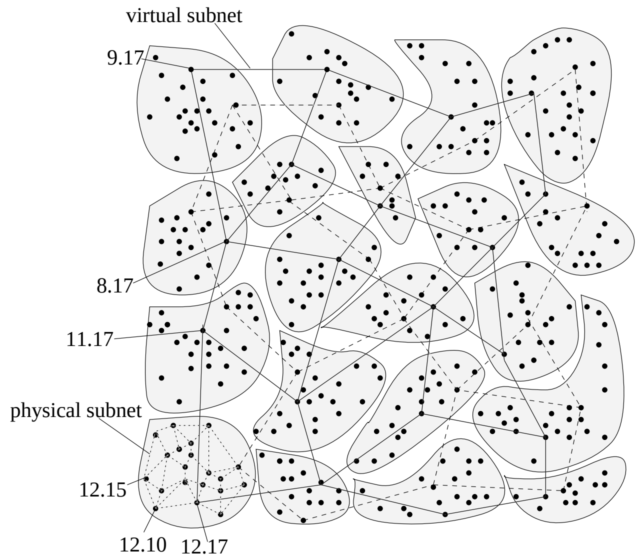

Figure 1: Physical and virtual subnets in a mobile radio network.

A node becomes a member of a physical subnet by acquiring the first available address (with the lowest LSD) in that subnet, e.g., if a node joins physical subnet 12 and there are already 10 members in this subnet it will use the address 12.10 since LSDs 0-9 are occupied already. Once a node becomes affiliated with a specific physical subnet, automatically it becomes a member of a virtual subnet defined by the LSD in its address; referring to the above example, the node will be a member in virtual subnet 10. As long as the node remains in the vicinity of physical subnet 12 (i.e., within "hearing" distance from its members) it will keep its current address. Any node in the network is updated with the current addresses used in its physical and virtual subnets (by its logical neighbors). This can be accomplished, e.g., by an advertising process where each node notifies its logical neighbors of its current address using a dedicated management channel. Therefore, a node which desires to join a specific physical subnet would contact a member(s) of this physical subnet to find out which address it can acquire, and then would advertise its newly acquired address to all of its logical neighbors. Note that if a node cannot reach any of its logical neighbors in its virtual subnet, it will use another virtual subnet via one of its logical neighbors in its physical subnet. This case will be further discussed in Section 4 dealing with routing.

When a node moves to a new location where it cannot establish a connection with its previous physical subnet's members, it will drop its previous address and join a new physical subnet whose members it can communicate with, assuming that there are available addresses in this subnet. If there are no available addresses in the new subnet, the node would seek another subnet (within a "hearing" distance) which has available addresses. If there are no available addresses in nearby physical subnets (e.g., there are already q nodes in this subnet each using a single address, or all the addresses are used by less than q nodes with multiple addresses per node and these nodes do not release an address for the mobile node), the mobile node would get a "guest" address $x.y$ from a nearby physical subnet where x is the physical subnet identification number and $y \geq q$. As soon as an address becomes available in physical subnet x, the "guest" node will acquire the released address and drop its previous "guest" address becoming a member of physical subnet x. The way a guest node communicates with other nodes is described in the next section.

Note that there is no logical connection between the virtual subnets, however, since they are "overlayed" on the same region they might interfere with each other. This interference is eliminated by the channel access scheme in use. One way, for example, is to operate each virtual subnet on a different frequency band[1]; when the number of frequency bands is less than the number of virtual subnets some form of time sharing can be used. Note that a less acute problem exists between neighboring physical subnets which have a limited degree of overlapping since they are defined over a local region. Thus, one can take advantage of spatial reuse, where only neighboring (overlapping) subnets use different frequency bands.

3.1 *Mobility management*

A mobile node which changed its subnet affiliation will notify all the nodes in its new physical and virtual subnets (i.e., its current logical neighbors) of its newly acquired address. This notification process can take place, e.g., during the establishment of links with its logical neighbors or by broadcasting in its physical and virtual subnets. In general, a source node does not know the current address of a desired destination node. The source node can determine this address by inquiring in its physical (virtual) subnet, since one of the nodes there is affiliated with the destination node virtual (physical) subnet. To clarify, let source node S and destination node D addresses be $S_1.S_0$ and $D_1.D_0$, respectively, and denote by $|S_1|$ the cardinality (number of members) of physical subnet S_1. Consider two different cases for finding node D's address; first, if $|S_1| \geq D_0$, node S would inquire in its physical subnet S_1 about node D and receives node D's address from node $S_1.D_0$ (which was notified earlier by node D, via virtual subnet D_0, regarding its current address). Second, if $|S_1| < D_0$, node S would inquire in virtual subnet S_0 about node D and receives node D's address from node $D_1.S_0$ which is affiliated with node D physical subnet. Note that node S does not know a-priori which of the above cases is valid, nevertheless, it inquires about node D first in its physical subnet and if it does not get a response it inquires in its virtual subnet (at

[1] A frequency band contains at least one frequency channel. When the band contains a single channel the subnet is operated in a pure TDMA mode, otherwise a TDMA/FDMA scheme can be used to increase throughput.

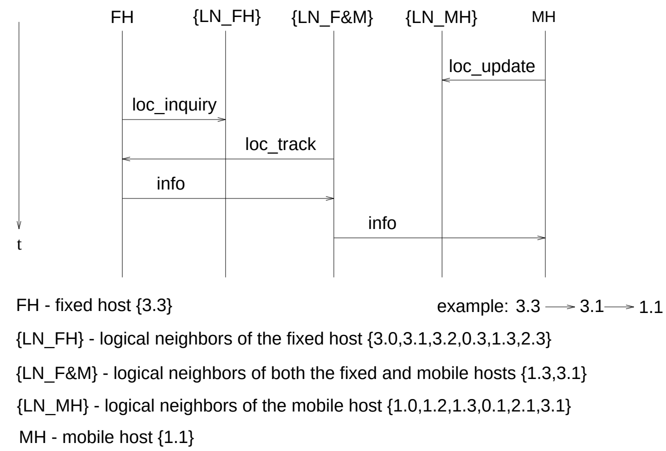

Figure 2: Location updating and tracking scheme.

least one of node S logical neighbors knows node D's address). Alternatively (instead of inquiring about node D's address), node S can broadcast its packets for node D in its physical and virtual subnets, at least one node (which is a logical neighbor of node D) will be able to forward the packets to their destination.

Figure 2 describes a simplified location updating and tracking scheme. After moving to a new location a mobile host (MH) notifies its logical neighbors by sending a location update (loc_update). A fixed host (FH) which desires to communicate with MH will inquire at its logical neighbors about MH by sending a location inquiry (loc_inquire). At least one of FH's logical neighbors is also MH's logical neighbor, thus, one of their mutual logical neighbors will provide FH with MH's address by sending loc_track. After tracking down MH's address, FH sends its data to MH via one of their mutual logical neighbors. Figure 2 shows also through a specific example in a 16-node network ($p = q = 4$) the nodes involved in the above process. Note that this location updating and tracking scheme involves only $p+q-2$ nodes which is much less than N; this results in reduced signaling/control overhead. For example, in a 400-node network with $p = 16$, $q = 25$, the location updating/tracking process involves less than 10% of the nodes in the network. Therefore, in the case of multi-hop subnets a flooding scheme can be used (assuming a contentionless access scheme, e.g., TDMA/FDMA) in the corresponding physical and virtual subnets, to broadcast the loc_update and loc_inquiry messages without overloading the whole network.

4 Routing

Several routing schemes are possible; we describe a shortest path routing procedure which is self-routing. We assume first the case of one-hop physical/virtual subnets (when using an airborne relay, satellite). In this procedure routes traverse one digit at a time in fixed order, e.g., from the LSD to the MSD. For example, when the source-node ad-

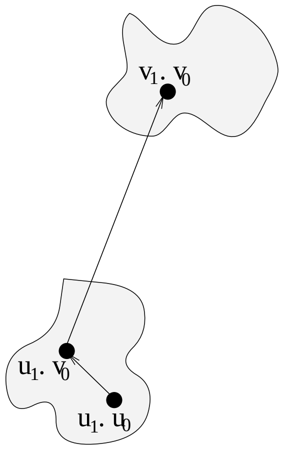

Figure 3: Shortest path routing for the case of one-hop physical/virtual subnets; physical – virtual.

dress is 12.15 and the destination-node address is 9.17, this procedure would use the path 12.15 $\longrightarrow$ 12.17 $\longrightarrow$ 9.17. Generally, the route from source-node address $u_1.u_0$ to destination-node address $v_1.v_0$ would traverse the path $u_1.u_0 \longrightarrow u_1.v_0 \longrightarrow v_1.v_0$ (see Figure 3). The length of the path equals the number of different digits in the addresses of the source and destination nodes, i.e., at most two hops. In the above procedure there is a unique path between any two nodes.

In general, the network is composed of multi-hop subnets which means that more than two hops are necessary from source to destination. In this case the routing is performed in two phases. In the first phase (Phase I) routing is performed only in the physical subnets (exchanging local traffic). Here packets are routed (e.g., using shortest path) within the physical subnet of the source node from the source node via intermediate nodes to a node having the same LSD as the destination node. In the second phase (Phase II) packets are routed in the virtual subnet, where packets reached to the last node in the first phase are routed from this node to the destination node via intermediate nodes within the virtual subnet defined by the LSD of the destination node. Preferably, during Phase I transmission power is limited to cover only the local area of the corresponding physical subnet; this would allow frequency reuse due to spatial separation. In Phase II (when virtual subnets are formed) transmission coverage is adjusted (e.g., by using a directional antenna) to reach remote physical subnets. Referring to Figure 1, the route from source node 12.15 to destination node 9.17 would traverse the path 12.15 $\longrightarrow$ 12.10 $\longrightarrow$ 12.17 $\longrightarrow$ 11.17 $\longrightarrow$ 8.17 $\longrightarrow$ 9.17, with two hops in physical subnet 12 and three hops in virtual subnet 17. Note though, that in case a node cannot reach any of its logical neighbors in its virtual subnet (i.e., when the virtual subnet is not connected) it will have to use a different virtual subnet via one of its logical neighbors in its physical subnet. For example, say that node $A.B$ would like to communicate with node $C.D$; using shortest path routing the path would usually traverse via physical subnet A to node $A.D$ and then via virtual subnet D to node $C.D$. However, if node $A.D$ is not connected to virtual subnet D it will connect to another node (say $A.E$) via physical subnet A and then via virtual subnet E to node $C.E$ and finally to the destination node $C.D$ via physical subnet C as indicated in the following,

$$A.B \longrightarrow \cdots \longrightarrow A.D \longrightarrow \cdots \longrightarrow A.E \longrightarrow$$

$$\cdots \longrightarrow C.E \longrightarrow \cdots \longrightarrow C.D$$

alternatively, the fault-tolerant routing scheme described at the end of this section can be used to overcome situations of disconnected subnets.

Usually a node will have a single address, however, it is possible under certain circumstances to have more than one address. For example, suppose that source node $S_1.S_0$ from a lightly populated physical subnet desires to communicate with destination node $D_1.D_0$ from a highly populated physical subnet and $|S_1| < D_0$; then source node $S_1.S_0$ will get another address – $S_1.D_0$ and will participate in virtual subnet D_0 in order to communicate with the above destination node. In this case the source node is affiliated with physical subnet S_1 and virtual subnets S_0 and D_0.

A guest node (with LSD greater than q) is not affiliated with any virtual subnet (there are only q such subnets); therefore, during Phase I it communicates locally exchanging intra-subnet traffic like the other nodes in the subnet, however, during Phase II it is idle. When a source node wants to send a packet to a guest node outside its physical subnet it will forward its packet via its virtual subnet to the corresponding node in the guest node's physical subnet and that node will forward the packet to the guest node during the next phase (Phase I), i.e., via intra-subnet traffic.

Finally, we mention here another self-routing scheme (called Long-path routing) which results in high fault tolerance (see Section 6.1.3). In this procedure the longest path has three hops (assuming one-hop physical/virtual subnets). Referring to Figure 4, the route from $u_1.u_0$ to $v_1.v_0$ would traverse the path $u_1.u_0 \longrightarrow u_1.u_0' \longrightarrow v_1.u_0' \longrightarrow v_1.v_0$ (Figure 4(a)), or $u_1.u_0 \longrightarrow u_1'.u_0 \longrightarrow u_1'.v_0 \longrightarrow v_1.v_0$ (Figure 4(b)), where $0 \leq u_0' \leq q-1, u_0' \neq u_0$ and $0 \leq u_1' \leq p-1, u_1' \neq u_1$. Note that each route traverses alternately physical and virtual subnets. It can be shown that between each source-destination pair there are $p+q-2$ disjoint paths, i.e., paths that do not share links or nodes (e.g., for $p = q = \sqrt{N}$ the number of disjoint paths is $O(\sqrt{N})$). Each of these paths corresponds to one of the $p+q-2$ logical neighbors of the source node. Note that a path is uniquely specified once a logical neighbor was selected by the source node. To route a packet from a source node to a destination node, the source node selects (say at random) one of the $p+q-2$ disjoint paths. In case of a path failure, the source node can select (say randomly) one of the remaining disjoint paths.

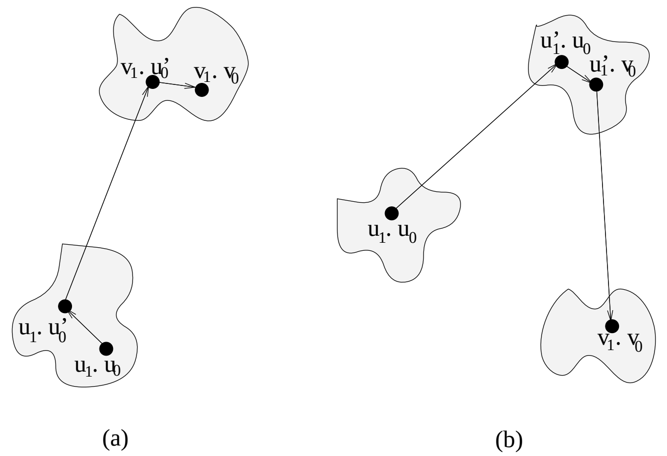

Figure 4: Long-path routing for the case of one-hop physical/virtual subnets; (a) physical – virtual – physical, (b) virtual – physical – virtual.

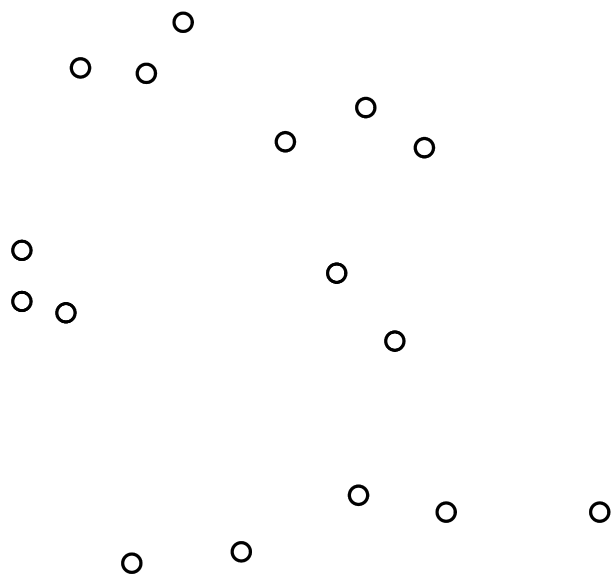

Figure 5: Example: 16-node mobile radio network.

5 Example

Consider the 16-node mobile radio network depicted in Figure 5. Assume that $p = q = 4$, i.e., there are four physical subnets and four virtual subnets of four nodes each, where each node is a member of one physical subnet and one virtual subnet. Figure 6 depicts the above 16-node network showing the links activated during Phase I forming four physical subnets of four members each with their addresses. Figure 7 depicts the same 16-node network showing the activated links during Phase II forming four virtual subnets (each virtual subnet has the same shape nodes). Each virtual subnet contains four nodes, one from each of the four physical subnets above. For clarification consider the following cases (assuming shortest path routing):

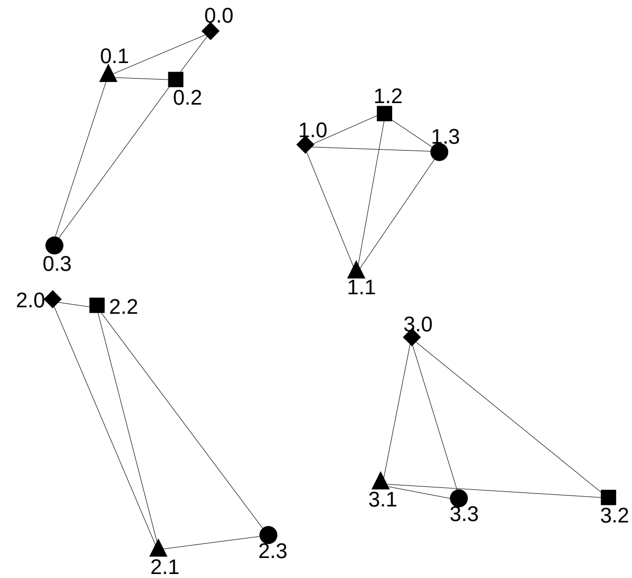

Figure 6: Example: physical subnets in a 16-node mobile radio network.

1. source-node 0.1, destination-node 3.0; the routing path is $0.1 \longrightarrow 0.0 \longrightarrow 1.0 \longrightarrow 3.0$, traversing one hop (in physical subnet 0) during Phase I and two hops (in virtual subnet 0) during Phase II, altogether three hops.

2. source-node 1.1, destination-node 2.3 and suppose that node 1.3 left physical subnet 1; since address 1.3 was dropped source-node 1.1 will acquire an additional address - 1.3 and the traversed path is of one hope (in virtual subnet 3) during Phase II $1.1, 1.3 \longrightarrow 2.3$.

3. suppose that node A of address 0.0 moved to the vicinity of physical subnet 3 (i.e., it has a physical connectivity with nodes that belong to physical subnet 3). Since there are no more available addresses in this subnet, node A will get a "guest" address 3.4. If the source-node is 2.0 and the destination-node is node A (3.4), the traversed path will be $2.0 \longrightarrow 3.0 \longrightarrow 3.4$, with one hop in virtual subnet 0 (during Phase II) and then another hop in physical subnet 3 (during Phase I).

Finally, using Long-path routing in the above 16-node network results in six disjoint paths (i.e., paths that do not share links or nodes) between any source-destination pair (see Figure 8).

6 Performance

In subsection 6.1 network performance in terms of average number of hops, delay, throughput and fault-tolerance is evaluated for two-hop networks (with one-hop physical/virtual subnets). For this case we assume that every subnet is operated in a pure TDMA mode with different

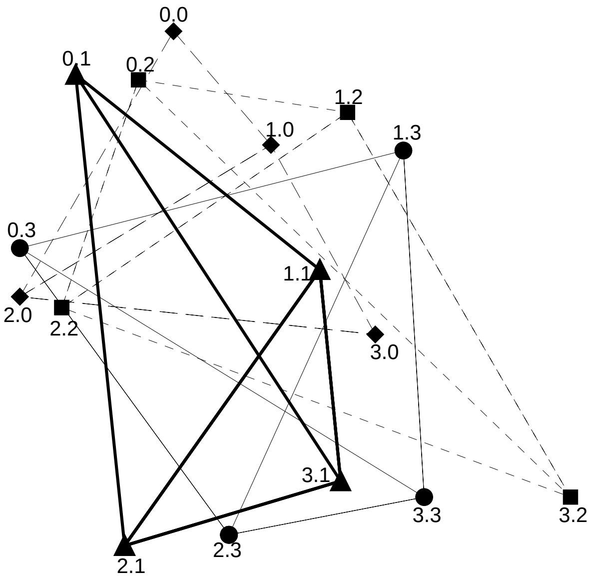

Figure 7: Example: virtual subnets in a 16-node mobile radio network.

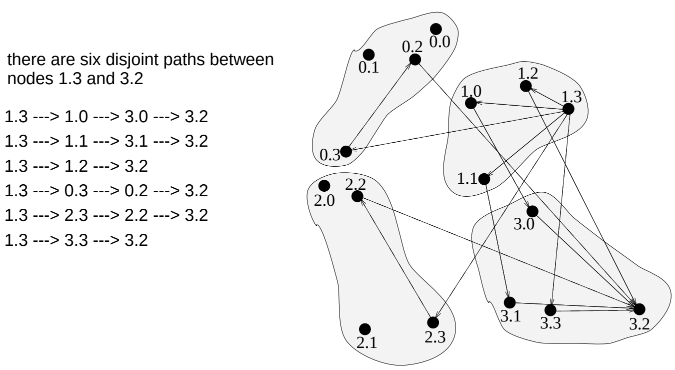

Figure 8: Example: disjoint paths in a 16-node mobile radio network using Long-path routing.

frequency bands assigned to potentially interfering subnets. Subsection 6.2 considers the general case of multi-hop networks (with multiple hops in each subnet) where subnets might operate in a TDMA/FDMA mode using multiple frequency channels in each subnet. In this subsection we evaluate the throughput of the network and compare it to that of a one large multi-hop network. We assume throughout Section 6 that the above shortest-path routing is used (i.e., any path traverses at most two subnets, one physical and one virtual).

6.1 *Two-hop networks*

6.1.1 *Average number of hops*

Each node in the network has $p+q-2$ destinations one hop away ($q-1$ and $p-1$ within its physical and virtual subnet, respectively), and $(p-1)(q-1)$ destinations two hops away (to $p-1$ nodes affiliated with each of $q-1$ virtual subnets), thus, the average number of hops in the network is given by

$$\bar{h} = \frac{2pq-(p+q)}{pq-1} \tag{1}$$

which is roughly equal to 2 for large networks.

6.1.2 *Delay and throughput*

To find the delay and throughput of the network we first derive expressions for subnet load. We define *subnet load* as the number of times a subnet is traversed by all possible $N(N-1)$ paths between source-destination pairs. It is assumed that traffic is homogeneous, where each node in the network sends λ packets/sec to any of the other $N-1$ nodes. The total traffic passing through each physical subnet is composed from $p\lambda$ packets/sec sent by each of the q nodes in the subnet to any of its $q-1$ peers in the subnet; therefore, the physical subnet load is

$$\eta_{ph} = N(q-1) \tag{2}$$

Similarly, the total traffic passing each virtual subnet is composed from $q\lambda$ packets/sec sent by each of p nodes to any of its $p-1$ peers in the virtual subnet, therefore, the virtual subnet load is

$$\eta_v = N(p-1) \tag{3}$$

thus, the maximum subnet load is

$$\eta_{max} = N(\max(p,q)-1) \tag{4}$$

The average subnet load is given by

$$\eta_{av} = \frac{p\eta_{ph}+q\eta_v}{p+q} = \frac{N(2N-(p+q))}{p+q} \tag{5}$$

In the symmetric case, where all the subnets have the same size $p=q=\sqrt{N}$,

$$\eta_{max} = \eta_{av} = N(\sqrt{N}-1) \tag{6}$$

which results in a balanced load, i.e., all subnets have the same load.

To simplify the presentation, we use an M/M/1 queueing model to describe a single subnet behavior; therefore, the average delay of a packet traversing subnet k is given by $\delta_k = \frac{1}{\mu C_k - \eta_k \lambda}$, where $1/\mu$ is the average packet length in bits, C_k is the subnet capacity in bits/sec and η_k is the subnet load. Note that if a more accurate model for the subnet behavior is used it will only result in a different expression for δ_k. Using Little's formula [14], the average queueing delay through the network is given by

$$\delta_{av} = \frac{1}{\Lambda}\sum_{k=1}^{S}\frac{\eta_k\lambda}{\mu C_k - \eta_k\lambda} \tag{7}$$

where S and Λ are the total number of subnets and offered traffic in the network, respectively; note that $S=p+q$

and $\Lambda = N(N-1)\lambda$. Since we assume a single tranceiver per node, let α be the portion of the time-frame this tranceiver is dedicated to its physical subnet during Phase I. Assume also that there are M frequency bands, and denote by R the transmitter burst-rate in bits/sec. The capacity of physical subnet k is given by $C_k^{ph} = \frac{\alpha R}{\tau_r}$, where $\tau_r = \max(r/M, 1)$ and r is the number of frequency bands required to operate all the physical subnets simultaneously (utilizing spatial reuse of frequencies at distant subnets); note that $r \leq p$ and depends on the physical layout of the subnets. Similarly, the capacity of virtual subnet k is given by $C_k^v = \frac{(1-\alpha)R}{\tau_q}$, where $\tau_q = \max(q/M, 1)$; note that q frequency bands are required to operate simultaneously the virtual subnets since they overlap. Using (7) one finds that the average queueing delay across the network is given by

$$\delta_{av} = \frac{1}{N(N-1)}\left(\sum_{k=1}^{p} \frac{N(q-1)}{\frac{\alpha\mu R}{\tau_r} - N(q-1)\lambda} + \sum_{l=1}^{q} \frac{N(p-1)}{\frac{(1-\alpha)\mu R}{\tau_q} - N(p-1)\lambda}\right) \quad (8)$$

or

$$\delta_{av} = \frac{1}{N-1}\left(\frac{p(q-1)}{\frac{\alpha\mu R}{\tau_r} - N(q-1)\lambda} + \frac{q(p-1)}{\frac{(1-\alpha)\mu R}{\tau_q} - N(p-1)\lambda}\right) \quad (9)$$

where

$$0 \leq \lambda \leq \min\left(\frac{\alpha\mu R}{N(q-1)\tau_r}, \frac{(1-\alpha)\mu R}{N(p-1)\tau_q}\right) \quad (10)$$

i.e., the load cannot exceed the subnet capacity. The maximum value of λ is achieved for α_{opt} satisfying

$$\frac{\alpha\mu R}{N(q-1)\tau_r} = \frac{(1-\alpha)\mu R}{N(p-1)\tau_q}) \quad (11)$$

solving (11) for α yields

$$\alpha_{opt} = \frac{(q-1)\tau_r}{(q-1)\tau_r + (p-1)\tau_q} \quad (12)$$

Substituting the value of α_{opt} in one of the sides of (11) one finds that

$$\lambda_{max} = \frac{\mu R}{N(q-1)\tau_r + N(p-1)\tau_q} \quad (13)$$

Using (12) and (13) in (9), and after some arrangements one gets

$$\mu R\delta_{av} = \frac{\mu R(p+q)}{N(N-1)\lambda_{max}} \frac{1}{1 - \frac{\lambda}{\lambda_{max}}} \quad (14)$$

The normalized (to burst-rate) user throughput is given by

$$\gamma = (N-1)\frac{\lambda_{max}}{\mu R} \quad (15)$$

and the normalized network throughput is given by

$$\Gamma = N\gamma \quad (16)$$

substituting (13) and (15) into (16),

$$\Gamma = \frac{N-1}{\tau_q(p-1) + \tau_r(q-1)} \quad (17)$$

Using (15) and (16), (14) becomes

$$\mu R\delta_{av} = \frac{p+q}{\Gamma}\frac{1}{1-\rho} \quad (18)$$

where $\rho \equiv \frac{\lambda}{\lambda_{max}}$ is the normalized (user/network) offered load. Observe that the left side of (14) is the average packet queuing delay across the network in units of packet transmission time.

Note that when $1 \leq M \leq \min(r, q)$ the throughput *is* linear with the number of frequency bands M, i.e., adding more frequency bands will increase linearly the throughput. Increasing M beyond $\min(r, q)$ will still increase the throughput in a sub-linear relation until a maximum is reached for $M = \max(r, q)$.

$$\Gamma_{max} = \frac{N-1}{p+q-2} \quad (19)$$

with a global maximum of $\Gamma_{MAX} = (\sqrt{N}+1)/2$ for $p = q = \sqrt{N}$. This means that for a network of size N and subnets of size $\sqrt{N}$, $\sqrt{N}$ frequencies are required to achieve maximum throughput. Figure 9 depicts the normalized throughput of different size networks for different numbers of available frequencies assuming that $p = q = \sqrt{N}$ and $r = 8$, i.e., eight frequencies are necessary and sufficient for simultaneous operation of all the physical subnets.

Figure 10 depicts the effect of subnet size on network throughput. In accordance with (19), higher throughput is expected when $p+q$ decreases till its minimum value of $2\sqrt{N}$ for $p = q = \sqrt{N}$. Thus, better throughput performance is achieved when the physical subnet size is closer to the virtual subnet size with the best performance for identical size. This is because for $p = q = \sqrt{N}$ the average subnet load is equal to the maximum subnet load (see (6)). This means that for *any* other routing scheme $\eta_{max} \geq N(\sqrt{N}-1)$; therefore, the current routing procedure achieves maximum throughput and minimum delay for all loads from zero till the maximum throughput is achieved.

Figure 11 depicts the average end-to-end packet queuing delay versus user/network offered load for a 1024-node network with different size subnets; it was assumed that $M = r = 8$. In Figure 12 the dependency of the average

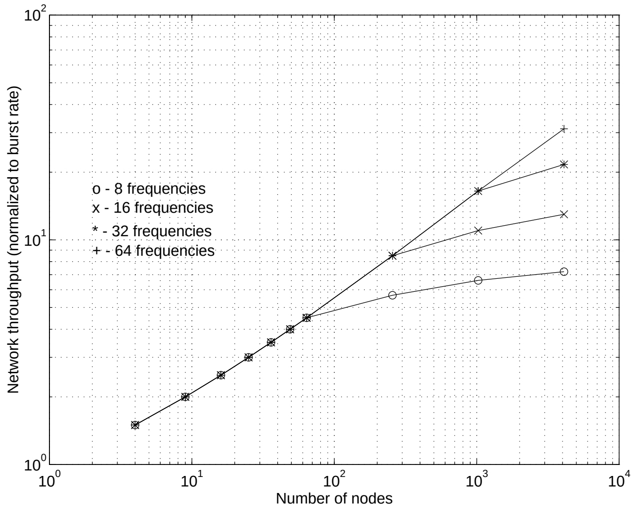

Figure 9: Effect of the number of frequencies on the throughput in two-hop networks.

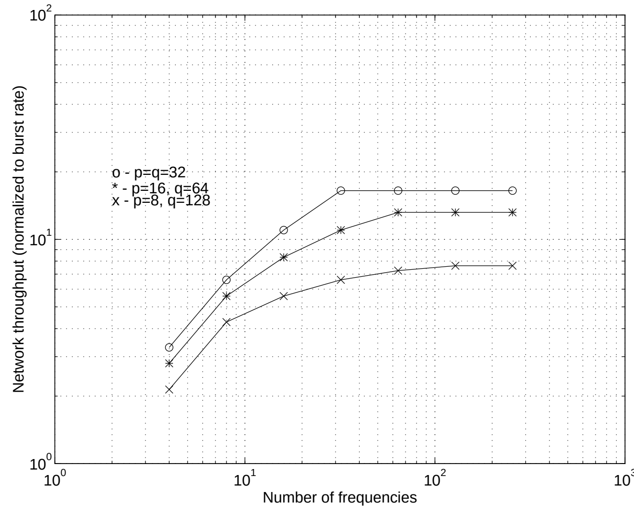

Figure 10: Effect of subnet size on the throughput in a two-hop 1024-node network.

delay on the offered load for a 1024-node network is depicted for different values of available frequencies; it was assumed that all subnets have the same size ($p = q = 32$) and $r = 8$. Figure 13 depicts the average delay vs. network size (number of nodes) for different values of offered load; it was assumed that ($p = q = \sqrt{N}, M = 8, r = 8$).

6.1.3 *Fault-tolerance*

To evaluate the fault-tolerance of the architecture we use two metrics, *node-connectivity* and *link-connectivity*. We define *node-connectivity* – κ as the minimum number of faulty nodes that creates a disconnected network. Since each node in the logical topology is connected directly to $p + q - 2$ other nodes, the node-connectivity of the network is $\kappa = p + q - 2$. For example, consider a 1024-node network composed of 32 subnets of 32 nodes each, then any 61 nodes can be faulty before the network becomes disconnected. Define the network *link-connectivity* – σ as the minimum number of node-disjoint paths between any source-destination pair (i.e., paths that do not share links or pass through the same node). The link-connectivity of the topology is $\sigma = 2$ for path-lengths of not more than two hops. If we allow path-lengths to be up to three hops (using Long-path routing, see Figure 4), it can be shown that the link-connectivity reaches its maximum value of $\sigma = p + q - 2$. Referring to the above example, there are at least 62 node-disjoint paths, i.e., alternative paths between any source-destination pair with path-lengths of at most three hops. Since the network has high node-connectivity and high link-connectivity it is therefore very reliable. A possible technology that can benefit from this is wireless ATM, where a virtual path established between a given source-destination pair has many alternative disjoint routes which can be used in case of node or link failures due to mobility, interference etc.

6.2 *Multi-hop networks*

In general, the subnets are not fully connected and multiple hops are necessary in each physical or virtual subnet. We assume that in each subnet a link-activation TDMA/ FDMA (multiple frequency channels per subnet are possible) scheme is used where each link is activated at least once during each time frame. The analysis in this subsection is true for any routing procedure *within* the multi-hop subnets.

6.2.1 *Throughput*

In the following analysis we use these additional notations:

T number of time-slots used in one large multi-hop network.

T_1 number of time-slots used in Phase I.

T_2 number of time-slots used in Phase II.

L total number of links activated in one large multi-hop network.

L_1 total number of links activated in the physical subnets during Phase I.

L_2 total number of links activated in the virtual subnets during Phase II.

η^i load of link i in one large multi-hop network, i.e., the number of times link i is traversed by all possible $N(N-1)$ paths in the network.

η_1^j load of link j in its physical subnet, i.e., the number of times link j is traversed by all possible $q(q-1)$ paths in its physical subnet.

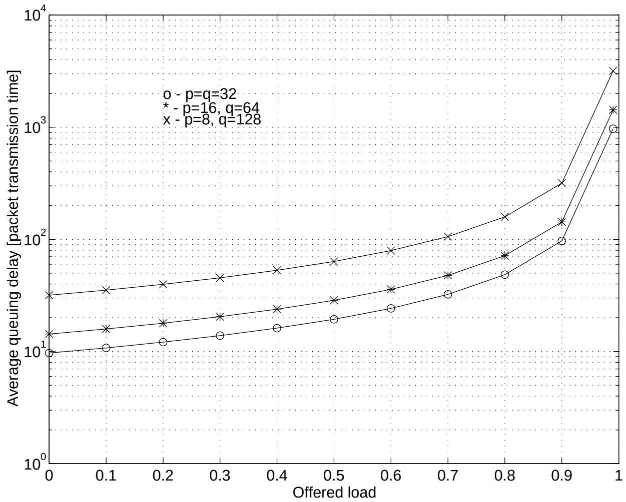

Figure 11: Average queuing delay vs. offered load for a 1024-node network with different size subnets ($M = 8, r = 8$).

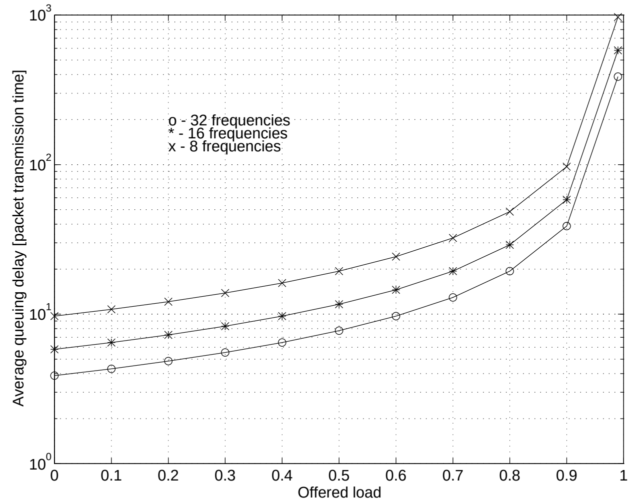

Figure 12: Average queuing delay vs. offered load for a 1024-node network for different values of available frequencies ($p = q = 32, r = 8$).

η_2^k load of link k in its virtual subnet, i.e., the number of times link k is traversed by all possible $p(p-1)$ paths in its virtual subnet.

η maximum link-load in one large multi-hop network (i.e., $\max_i(\eta^i)$).

η_1 maximum link-load during Phase I (i.e., $\max_j(\eta_1^j)$).

η_2 maximum link-load during Phase II (i.e., $\max_k(\eta_2^k)$).

We assume again homogeneous traffic and an M/M/1 queueing model to describe the behavior of each activated link. Using Little's formula and summing over all the activated links in the network, the average queueing delay across the network is given by

$$\delta_{av} = \frac{1}{N(N-1)} \left(\sum_{j=1}^{L_1} \frac{p\eta_1^j}{\frac{\alpha\mu R}{T_1} - p\eta_1^j \lambda} + \sum_{k=1}^{L_2} \frac{q\eta_2^k}{\frac{(1-\alpha)\mu R}{T_2} - q\eta_2^k \lambda} \right) \tag{20}$$

Note that the capacity of the links is inversely proportional to the number of time slots (in the corresponding phase) which depends on the number of frequency channels used. The maximum traffic between any two nodes during phase I is $\lambda_1 = \frac{\alpha\mu R}{T_1\eta_1 p}$. Similarly, the maximum traffic between any two nodes during phase II is $\lambda_2 = \frac{(1-\alpha)\mu R}{T_2\eta_2 q}$. To find the optimal value of α we solve $\lambda_1 = \lambda_2$ and find that

$$\alpha_{opt} = \frac{T_1\eta_1 p}{T_1\eta_1 p + T_2\eta_2 q} \tag{21}$$

therefore, the maximum traffic between any two nodes in the network is

$$\lambda_{max} = \frac{\mu R}{T_1\eta_1 p + T_2\eta_2 q} \tag{22}$$

The normalized (to burst rate) network throughput is given by

$$\Gamma = N(N-1)\frac{\lambda_{max}}{\mu R} = \frac{N(N-1)}{T_1\eta_1 p + T_2\eta_2 q} \tag{23}$$

Note that η_1 and η_2 depend on the logical topology and the routing procedure used within the subnets.

Observe that (17) can be derived as a special case (a two-hop network) from (23) by substituting $T_1 = q(q-1)\tau_r$, $T_2 = p(p-1)\tau_q$ and $\eta_1 = \eta_2 = 1$ (note that $q(q-1)$ and $p(p-1)$ are the number of links in one-hop fully connected physical and virtual subnet operated in a pure TDMA mode, respectively).

We are interested in comparing the performance (e.g., throughput) of the proposed architecture and that of a one large multi-hop radio network. First, we find the throughput of the one large network. In a similar way to the above, the queueing delay across one large multi-hop network is given by

$$\delta'_{av} = \frac{1}{N(N-1)} \sum_{i=1}^{L} \frac{\eta^i}{\frac{\mu R}{T} - \eta^i \lambda} \tag{24}$$

thus, the throughput of one large multi-hop network is given by

$$\Gamma'_{max} = \frac{N(N-1)}{T\eta} \tag{25}$$

The throughput values for the proposed architecture (23) and for one large multi-hop network (25) depend strongly

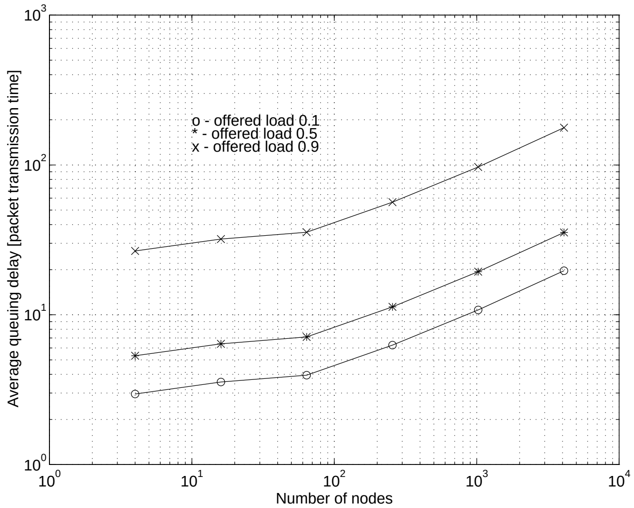

Figure 13: Average queuing delay vs. number of nodes for different values of offered load ($p = q = N^{1/2}, M = 8, r = 8$).

on the physical and logical topologies, routing procedure and the number of frequencies. We observed – especially in large networks with random topology (a characteristic of ad-hoc sporadic networks) – that the maximum link traffic-load in one large network is significantly higher than the maximum link traffic-load in the subnets, i.e., $\eta \gg p\eta_1,\ q\eta_2$. Therefore, partitioning of the network may reduce congestion in the network which in effect can improve the performance, e.g., results in higher throughput, lower delay. Note that there is a penalty associated with large transmission radii resulting in reduced link capacities (because more time-slots are needed). However, since the link loading may be significantly reduced, the total effect may result in increased throughput. Also, while one large multi-hop network cannot take advantage of many frequencies (if available) because of spatial reuse, the current architecture can (to separate the overlayed virtual subnets), which also results in increased throughput.

The following example of a 16-node network illustrates the strength of the proposed architecture. Figure 14 depicts a network with a well controlled topology of maximum degree six composed of equilateral triangles. According to the proposed architecture the network may be partitioned into four physical and four virtual subnets ($p = q = 4$). Figure 15 and Figure 16 show the links activated in the physical and virtual subnets, respectively.

Using TDMA/FDMA link-activation assignment (see Appendix A) and shortest-path routing, we find the values needed to calculate throughput performance depending on the number of available frequencies (see Table 1).

The throughput performance for one large multi-hop network and for a network using four virtual subnets is depicted in Figure 17. For this particular example, one can see that for a given number of frequencies the proposed

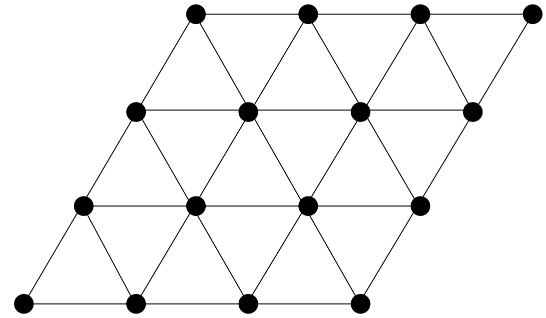

Figure 14: 16-node network with a well controlled topology.

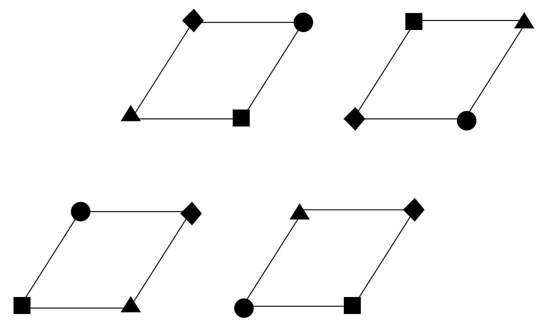

Figure 15: Links activated in the physical subnets.

architecture always has a better throughput performance than the one large multi-hop network. Note than not more than three frequencies are required to achieve the maximum throughput of the one large network, i.e., adding more frequencies will not increase the throughput. This is because of the limited transmission range of the nodes, taking advantage of spatial reuse. However, in the proposed network it is possible to further increase the throughput by adding more frequencies (up to eight frequencies). Observe that the average and maximum link loading in the one large multi-hop network is 8.33 and 16, respectively, while for the network using four virtual subnets the corresponding values are identical – 8.0, i.e., the network has a balanced load. The average and maximum number of hops in the one large network is 2.29 and 6.0, respectively, while for the network using four virtual subnets the corresponding values are 2.13 and 4.0.

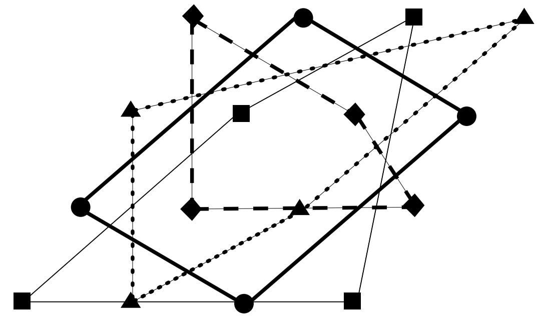

Figure 16: Links activated in the virtual subnets.

Number of frequencies	T	η	T_1	T_2	η_1	η_2
1	22	16	8	32	2	2
2	14	16	4	16	2	2
3	12	16	4	11	2	2
4	12	16	4	8	2	2
8	12	16	4	4	2	2

Table 1: Values for calculating throughput performance depending on the number of available frequencies.

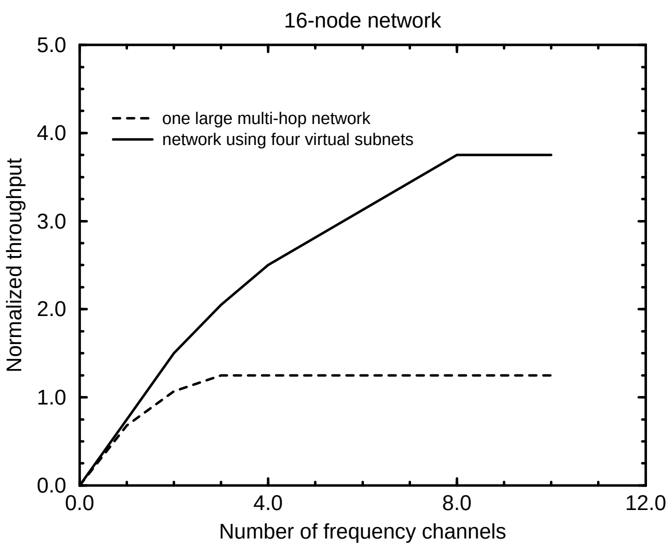

Figure 17: Throughput performance for one large multi-hop network and for a network using four virtual subnets.

7 Conclusions

An architecture comprising a logical topology of physical and virtual subnets, and corresponding addressing, mobility management and routing schemes were described. This architecture is applicable to mobile radio networks and accommodates dynamic topology changes due to relative movement of network nodes. The architecture partitions a mobile network into logically independent subnetworks. Network nodes are members of physical and virtual subnets and may change their affiliation with these subnets due to their mobility. Each node is allocated an address based on its current subnet affiliation. We observed – especially in large networks with random topology – that partitioning of the network may result in significantly more balanced load than in one large multi-hop network, an attribute that can significantly improve the network's performance. The architecture is highly fault-tolerant, has a relatively simple location updating and tracking scheme, and by virtue of its load balancing feature, typically achieves a network with relatively high throughput and low delay. The addressing method, logical topology, mobility management and routing procedure were described, and network performance for one-hop and multi-hop physical/virtual subnets was evaluated.

Acknowledgments

The author would like to thank A. C. Sevdinoglou and F. M. Torre for critically reading the manuscript and their comments.

Appendix A

In the following we describe a distributed on-line TDMA/FDMA link-activation algorithm for assigning a directed link between two nodes (say i and j). It is assumed that nodes i and j have knowledge of existing assignments at their neighboring nodes (i.e., nodes within "hearing" distance). The output of the algorithm is the assignment of link $i \longrightarrow j$, i.e., an ordered pair of time-slot and frequency – (t, f) or a time-slot – t when only one frequency is available (pure TDMA). We use the following notations:

T set of available time-slots.

F set of available frequencies.

T_r set of time-slots in use by node r.

$F_{r,h}^{in}$ set of frequencies in use by links incoming to nodes neighboring node r at time-slot t_h.

$F_{r,h}^{out}$ set of frequencies in use by links outgoing from nodes neighboring node r at time-slot t_h.

The sets T and F are ordered arbitrarily at the outset.

Algorithm

1. choose $t_k \in T - (T_i \cup T_j)$, where k is the least order number; if no time-slot is available go to step 4,
2. choose $f_l \in F - (F_{i,k}^{in} \cup F_{j,k}^{out})$, where l is the least order number; if no frequency is available, $T \leftarrow T - \{t_k\}$ and go to step 1,
3. stop, the pair (t_k, f_l) is the $i \longrightarrow j$ link assignment.
4. stop, link $i \longrightarrow j$ cannot be assigned.

Note that time-slot assignment involves only logical links, while frequency assignment considers both logical and physical links.

References

[1] I. Chlamtac and A. Lerner. A link allocation protocol for mobile multihop networks. In *Proc. IEEE Globecom '85*, December 1985.

[2] I. Cidon and M. Sidi. Distributed assignment algorithms for multihop packet radio networks. *IEEE Trans. Comp.*, C-38(10):1353–1361, 1989.

[3] M. S. Corson and A. Ephremides. A distributed routing algorithm for mobile wireless networks. *Wireless Networks*, 1(1), January 1995.

[4] O. S. de Souza, P. Sen, and R. R. Boorstyn. Congestion based routing in packet radio networks. In *Proc. IEEE ICC '89*, volume 3, pages 51.3.1–51.3.5, Boston, June 1989.

[5] E. M. Gafni and D. P. Bertsekas. Distributed algorithms for generating loop-free routes in networks with frequently changing topology. *IEEE Trans. Commun.*, COM-29:11–18, 1981.

[6] M. Gerla and J. T. Tsai. Multicluster, mobile, multimedia radio network. *Wireless Networks*, 1(3), October 1995.

[7] R. L. Hamilton, Jr. and H. C. Yu. Optimal routing in multihop packet radio networks. In *Proc. IEEE Infocom '90*, June 1990.

[8] T. Hou and V. O.K. Li. Transmission range control in multihop packet radio networks. *IEEE Trans. Commun.*, COM-34(1), January 1986.

[9] L. Hu. Distributed code assignments for CDMA packet radio networks. *IEEE Trans. Network.*, 1(6), December 1993.

[10] L. Hu. Topology control for multihop packet radio networks. *IEEE Trans. Commun.*, 41(10), October 1993.

[11] C. Jones and T. Anwar. Two routing strategies for land-based mobile packet radio networks. In *Proc. 3rd International Conference Information Network and Data Communication; INDC-90*, Lillehammer, Norway, March 1990.

[12] L. Kleinrock and Silvester J. Optimum transmission radii for packet radio networks or why six is a magic number. *Nat. Telecom. Conf.*, December 1978.

[13] L. Ruston and Sen P. A rule-based system for packet radio network design. In *Proc. IEEE Infocom '89*, pages 512–519, 1989.

[14] M. Schwartz. *Telecommunications Networks: Protocols, Modeling and Analysis.* Addison-Wesley, 1987.

[15] J. Sharony and A. C. Sevdinoglou. Distributed TDMA–FDMA–CDMA link assignment in mobile radio networks with/without flexible directivity. To be submitted for publication, 1996.

[16] H. Takagi and L. Kleinrock. Optimal transmission ranges for randomly distributed packet radio terminals. *IEEE Trans. Commun.*, COM-32, March 1984.