

# Topological Design of 3G Wireless Backhaul Networks for Service Assurance\*

Chalermpol Charnsripinyo<sup>1,2</sup> and David Tipper<sup>1</sup>, *Senior Member, IEEE*

<sup>1</sup>Department of Information Science and Telecommunications, University of Pittsburgh, Pittsburgh, PA, USA, 15260

<sup>2</sup>National Electronics and Computer Technology Center, Klong Luang, Pratumthani 12120, Thailand

E-mail: [chalerm.pol.charnsripinyo@nectec.or.th](mailto:chalerm.pol.charnsripinyo@nectec.or.th), [tipper@tele.pitt.edu](mailto:tipper@tele.pitt.edu)

**Abstract** – In this paper, we propose an optimization based model for the design of survivable third generation wireless access backhaul networks using a mesh topology. The network design model seeks to minimize the cost of the backhaul network while meeting quality of service and survivability requirements. The design model includes the packet based nature of network traffic and incorporates the effects of user mobility after a failure. We adopt a two-phase design methodology. The first phase provides a minimum-cost, initial network design meeting QoS requirements. The second phase augments the network topology from phase one in order to satisfy survivability requirements. In order to scale the design with network size a computationally efficient heuristic based on iterative minimum cost routing is proposed. Numerical results are given illustrating the network design approach and the quality of the heuristic solution method.

*Index Terms* – network design; survivability; wireless networks

## I. INTRODUCTION

Third-generation (3G) wireless networks are expected to support voice and data services for a range of applications requiring different aspects of quality of service (QoS). Providing QoS in 3G wireless mobile networks has been widely addressed in the literature [1,2]. QoS is generally referred to as a set of service requirements to be met by the network while transporting a traffic stream from source to destination. It may be specified in terms of delay, bit error rate, packet loss rate, guaranteed bit rate, and so on. As user dependence on mobile services increases, a network failure that inhibits communication or results in loss of critical data will be problematic. However, typical cellular networks have a tree-like (root-branch-leaf) topology, which is vulnerable to failures. Examples of failure scenarios in cellular networks include failure of a base station and loss of the link between a base station and mobile switching center. The effects of failure in cellular networks will vary and depend on the specific failure scenario, which can be determined by the network component that fails and its location. Recent survivability analysis studies [3-5], show that a relatively small failure in a cellular network (e.g., loss of 4 base stations out of a group of 100 attached to the same MSC) can have a large impact on network performance (e.g., call blocking, location update delay) throughout the entire service area. This is because, unlike wired networks disconnected users in wireless networks are mobile and a portion of disconnected users will simply move and attempt to reconnect, resulting in an increase in signaling traffic and call attempts. In order to protect network services against failures, network survivability strategies (i.e., redundant network components, fault-tolerant topology with spare capacity, traffic management, and restoration mechanism, etc.) must be incorporated into the network infrastructure during the design process.

In comparison to the literature on wired network design relatively little work exists on the design of cellular networks. Dutta and Kubat [6] considered the design of partially survivable backhaul networks for cellular systems. Self-healing rings were used for the backbone network. A diversity requirement, which forces a cell site to interconnect with more than one hub on the ring, was specified to ensure survivability. The design problem was formulated as an integer programming model and a heuristic based on Lagrangian relaxation was applied to solve the problem. Cox and Sanchez [7] considered the least-cost backhaul network design while meeting survivability and capacity constraints. The problem studied was to locate hubs and interconnect nodes in the backhaul network such that the total cost is minimized. A survivability constraint was specified as a routing-diversity which forces a cell site that has more than one link to connect with different hubs. The problem was formulated as a large integer programming problem and a Tabu Search heuristic was applied to solve the optimization problem. Kubat, Smith and Yum [8] consider the problem of cellular network design with diversity and capacity constraints. The objective was to find a cell-to-hub interconnection and backbone link capacity so that the diversity requirement is met and the cost of interconnection is minimized. In the design, the topology of the backbone is given as a tree topology (or a forest) with the MSC at the root. The diversity requirement is used to ensure that a cell site is connected to more than one hub node of different subtrees for partial traffic protection if one of the connections fails. The authors present mathematical models for three different variations on the cellular network design problem. Heuristic algorithms based on LP relaxations were presented to solve the problems. In [9], we proposed the use of a survivable mesh type of backhaul topology which is any single link fault tolerant and formulated a integer programming optimization model to determine the network design. A routing based heuristic was proposed to solve the optimization model.

Availability based approaches to network design were studied in [10,11]. Ma, et. al., [11] presented techniques for the computation of the availability and steady state performance of 2G cellular network architecture. Malloy, et. al, in [11] develop a framework for evaluating the availability of cellular networks. The framework consists of decomposing the cellular network into interconnected modules based on the facility hierarchy of a 2G cellular network. Given the reliability of network components, availability calculations are conducted to approximately determine the network availability and the impact of various failures. A network design procedure is proposed whereby redundant modules and fault tolerant interconnection of the modules are added to the network (by a network designer) and the improvement in availability determined. In [12], the authors considered reliable 3G UMTS based backhaul network design. The problem was to find a minimum cost sub-network between the radio network controller (RNC) and its assigned base stations (BS) while meeting availability requirements. The requirements were based on the availability of the working path and the traffic loss parameter of the base stations. A two-phase heuristic method was presented to solve the reliable

---

\*This work was supported in part by NSF grant ANIR 9980516 and a NIST Critical Infrastructure Protection grant

network planning problem. In the first phase, a tree-like topology between RNC and BS is designed while taking reliability into account. In the second phase, heuristics are proposed to add new links for protecting the most failure-sensitive parts of the topology. The traffic in the network was modeled as an aggregate fixed demand flow.

While the current literature provides insight into the design of survivable cellular backhaul networks, the approaches and assumptions used were largely based on existing voice based circuit-switched networks which limit their application to emerging 3G networks. Specifically, the current literature does not consider packet based traffic in the backhaul network and user mobility effects after a failure. Here we propose a novel design approach that incorporates these issues into the design of 3G backhaul networks. We adopt a two-phase design methodology to solve the mesh-based survivable 3G backhaul network design problem. The first phase provides a minimum cost network design meeting QoS requirements. The second phase uses an incremental approach to augment the network topology from phase one in order to satisfy survivability requirements. The design problem is formulated as a series of two mixed integer-programming models. Note that, optimization techniques for solving mixed integer programming problems do not scale for large network sizes. Hence, we develop a heuristic method based on iterative minimum cost routing to solve the network design model within a reasonable computational time while obtaining a near optimal solution. The remainder of this paper is organized as follows. In section 2, we discuss the critical issues in 3G backhaul network design including strategies for service assurance. In section 3, we present the mathematical formulation of optimization problems for the topological design of survivable 3G backhaul networks. In section 4, we present an iterative routing based heuristic algorithm to solve the design model. In section 5, we present numerical results for sample networks comparing solutions obtained by solving the optimization problem using standard techniques with our heuristic method. Also, results are given illustrating the effects of user mobility patterns after a base station failure on the network design cost. Finally, section 6 presents our conclusions.

## II. 3G BACKHAUL NETWORK DESIGN ISSUES

A typical 3G wireless network architecture based on the 3G partnership project (3GPP) for the evolution of UMTS to an all packet switched network [1,2] is illustrated in Figure 1. As shown in Figure 1 the radio access network (RAN) is a set of radio network subsystems (RNS) which connect to the core network. Each RNS consists of one radio network controller (RNC) which controls the radio resources and services in its domain, and one or more access point entities called Node B which are connected to the RNC. A Node B corresponds to a base station in the 2G GSM network, serving as a fixed access point for all mobile terminals (MT) within the cell. Similarly the RNC corresponds to a base station controller in GSM. The wireless link between a Node B and mobile is packet based using wideband CDMA (WCDMA) as the air interface standard. Serving GPRS support nodes (SGSNs) and a gateway GPRS support node (GGSN), which are enhanced versions of those in the General Packet Radio Service (GPRS) networks, are included in 3GPP networks. The primary functions of the SGSN are to detect and register new mobile terminals in its serving area, send/receive data packets to/from the mobiles, and to track the location of the mobiles within its service area. The GGSN is the interface between the 3G network and external networks. The functions of the GGSN include translation of data formats, protocol conversions, and address translation for routing of incoming and outgoing packets to the Internet and other data networks. The Call State Control Function (CSCF) performs call control, service switching, address translation and coding type negotiation functions. The Home Subscriber Server (HSS) is a superset of

GSM home location register (HLR) functionality which includes responsibility for maintaining user subscription and location information. Network gateways are used to connect with public switched telephone networks (PSTN) and legacy networks. Additional network database elements, such as, an Equipment Identity Register (EIR) (not shown in figure) may be included in the network architecture. Unlike current cellular networks, the architecture is based on packet switching and the signaling initiation protocol (SIP) is expected to be used for exchanging signaling messages between network entities.

The network architecture in Figure 1 is organized into a facility hierarchy. Each Node B transmits traffic that represents demands from all mobiles within the cell. Each Node B is connected to an RNC and the RNC is then connected to the SGSN and so forth. The communications links between the Node B, RNC, and SGSN are typically wireline or microwave links. This hierarchical structure suggests that traffic in the wireless backhaul network should be concentrated into high capacity links to the central locations. Hence, a tree-like network topology has been traditionally used. The tree topology is typically the least-cost network design due to the minimum connectivity and the economy of scale on high capacity links. However, the tree topology is vulnerable to failures. For the network to be survivable, alternate routes with sufficient spare capacity and appropriate traffic restoration techniques must be incorporated into the network infrastructure.

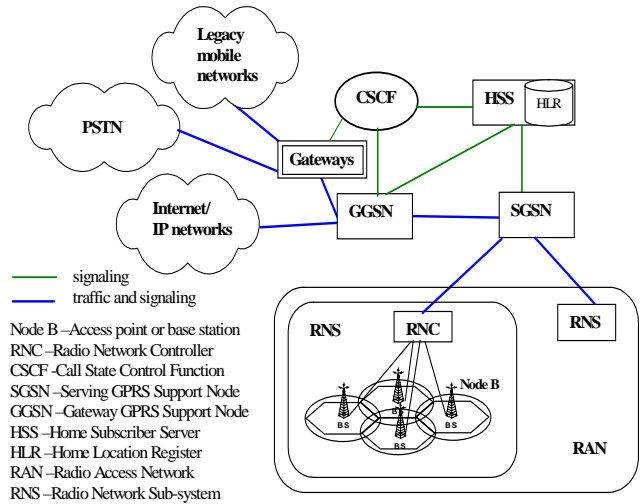


Figure 1. 3GPP wireless network architecture.

In survivable network design, several strategies can be used to protect network services. Following the wireless network survivability framework proposed in [4], survivability techniques can be deployed at different network layers. At the radio layer, the primary failure to be guarded against is failure of the wireless link to the mobile user. Due to the constraint of limited frequency spectrum, allocation of spare radio channels for use in case of the failure is not economically feasible. As shown in Figure 2, the effect of a BS (Node B) failure is to disconnect mobiles in the geographic area serviced by the BS. Techniques proposed in the literature to provide radio coverage in the disconnected area include designing the network with an overlapping or partially overlapping cell site architecture along with a dynamic channel allocation algorithm and adaptive power control to provide dual-homing at the wireless link level [4,13]. Figure 2 shows an architecture with partial overlapping coverage areas for radio level survivability. After the failure of Node B-2, the neighbor Node Bs increase their power level and expand their cell size, thereby providing a radio link for some mobiles in the failed cell. For mobiles unable to attach due to poor radio coverage, we propose that

they reconfigure into a ad-hoc network mode with mobiles that are in radio range of a Node B acting as relays to mobiles that are out of range as shown in Figure 2. However, this may overload the traffic in the cell sector, requiring forced handoffs of some users in the cell sector to neighbor cells (if possible) or forcing users in the overloaded sector to downgrade their data rate or to connect to another sector of the cell via an ad-hoc relay. Network protocols and signaling to achieve such radio link survivability is on-going research.

In the backhaul network, the primary survivability concern is link failure. Hence, we propose to adapt a mesh-based restorable network topology with link (span) restoration. Any link that is vulnerable to failure is augmented with a backup path with sufficient spare capacity between its two end-nodes. If a link failure occurs, the end nodes of the failed link will automatically restore network services by re-routing traffic through the backup path within a short period of time. Note that network protection for link failures can generally be classified into two schemes, namely link restoration and path restoration. For link restoration, the interrupted traffic is rerouted locally around the failed link whereas for path restoration the interrupted traffic can be rerouted on a new route from the source to destination nodes. Although using path restoration is known to be more cost-efficient, it is difficult to use in wireless networks due to the facility hierarchy.

Unlike wired networks where some portion of traffic is lost after network node failures, a portion of affected traffic in cellular networks will simply move to neighbor cells after a BS/Node B failure. Hence, it is essential to take into account such failure scenarios and their effects when designing the backhaul network. In the event of a BS failure, communication in the area served by the failed BS will be terminated. Mobile users in the area covered by the failed BS can not access the network unless a radio link survivability technique is employed. Alternatively, mobile users may move into neighbor cells and receive the radio signal from the neighbor BS. In any event, neighbor BS will have to serve increasing traffic demands from affected mobile users. Mobile users whose calls were prematurely terminated may attempt to reconnect to the network in a near simultaneous fashion. Simulation results of sample cellular networks show that BS failures will increase the traffic demand, signalling and connection blocking probability in cells adjacent to the failed area [4,5]. Note that, measurements from a cellular network operator [21], indicate that BS failures are the most common type of failure.

To mitigate the effect of single BS failures, extra bandwidth must be allocated at each BS to absorb increased traffic due to user movements from a failed neighbor cell. Each BS node may have different extra bandwidth requirements depending upon the geographical location of the BS in the service area. We estimate the amount of extra bandwidth for mitigating user mobility effects at each BS node as follows. Let  $\alpha_{ij}$  denote the mobility factor indicating the proportion of traffic demands from a neighbor BS node  $i$  served at BS node  $j$  after BS node  $i$  fails (note that  $0 \leq \alpha_{ij} \leq 1$  and  $\sum_j \alpha_{ij} \leq 1$ ). Let  $a^i$  be the total traffic demands served at BS node  $i$  in a normal operational state. Then, the amount of traffic at BS  $j$  in the event of BS  $i$  failing is  $\alpha_{ij} \cdot a^i$ . Let  $\beta_j$  denote the amount of spare bandwidth required at BS node  $j$  for any single BS  $i$  failure scenario. Then,  $\beta_j = \text{MAX}_i (\alpha_{ij} \cdot a^i)$ , that is,  $\beta_j$  is the maximum bandwidth required to serve traffic from mobile users at BS  $j$  due to any single neighbor BS node  $i$  failure.

### III. NETWORK DESIGN PROBLEM FORMULATION

Survivable network design involves three main tasks which are determining a fault-tolerant network topology, dimensioning links between network nodes, and routing traffic demands subject to QoS and survivability requirements. Here, we propose a two-phase network design approach similar to recent work on survivable wired backbone network design [14,17]. In the first phase, we formulate a

minimum-cost network design problem which is solved to yield an initial network. The problem is to determine the interconnection links, and their capacity to satisfy traffic demand with minimum cost. Note that, the traffic here is packet based and the capacity allocation is based on an aggregate equivalent bandwidth calculation.

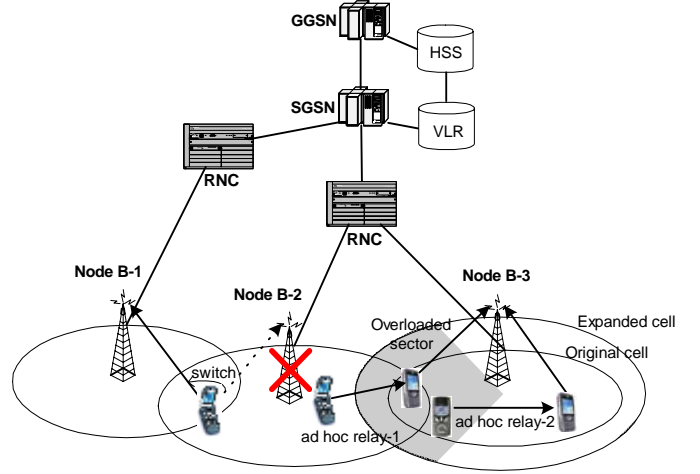


Figure 2. Radio link survivability techniques

In the second phase, we consider the survivable incremental design problem where the minimum-cost network (from phase one) is modified to satisfy survivability requirements. Specifically, links in the phase one network design will be protected using link restoration to guard against any single link failure. The new edges required in the link restoration design, together with existing edges in the minimum cost network provide a mesh restorable network topology. Also, to mitigate the effects of user mobility in the event of BS failures, additional spare capacity is allocated in the backhaul network. In the following, we present mathematical formulations for the two-phase network design procedure. We first present variable definitions and the equivalent bandwidth calculation.

$B$	Set of Node B's in the network service area
$H$	Set of potential hub nodes
$C$	Set of radio network controller (RNC) nodes, $C \subseteq H$
$M$	Set of 3G switching centers (i.e., 3G SGSNs)
$N$	Set of all nodes, $B \cup H \cup M$
$E$	Set of potential edges (links) between nodes in $N$
$N(i)$	Set of neighbor nodes $j \in N$ of node $i$ such that edge $(i, j)$ belongs to the edge set $E$
$D$	Set of all traffic demands between nodes indexed by $k$ . Each demand node-pair in $D$ is associated with an origin node and a target node
$O[k]$	The origin node of the $k^{\text{th}}$ demand node-pair in $D$
$T[k]$	The target node of the $k^{\text{th}}$ demand node-pair in $D$
$a^k$	The traffic demand associated with the $k^{\text{th}}$ demand pair in $D$ (i.e. $a^k$ is the data rate or bandwidth required)
$w_{ij}$	Minimum required capacity of working link between node $i$ and $j$ , $w_{ij} \in R^+$

$e_{ij}^{k,l}$	Routing binary decision variable that equals 1 if edge $(i, j) \in E$ is traversed by the $k^{\text{th}}$ demand pair in $D$ ; 0 otherwise. The index $l \in \{1,2\}$ represents the levels in the facility hierarchy, level 1 is from the origin node $O[k]$ to the RNC and level 2 is from the RNC to the corresponding target SGSN node $T[k]$
$\varphi_{ij}$	Edge binary decision variable ( $\varphi_{ij} = \varphi_{ji}$ ) that equals 1 if potential edge $(i, j) \in E$ between node $i$ and node $j$ is selected for connection; 0 otherwise
$VC_{ij}$	Variable cost of adding one unit of capacity to edge $(i, j) \in E$ between node $i$ and $j$
$FC_{ij}$	Fixed cost for adding an edge $(i, j) \in E$ between node $i$ and $j$
$y_{kc}$	Binary parameter that equals 1 if the $k^{\text{th}}$ traffic demand in $D$ must be processed at RNC node $c \in C$ ; 0 otherwise
$EC(bw)$	Equivalent capacity of aggregate traffic with bandwidth $bw$
$HP$	Maximum number of hops for each traffic route
$\Psi$	Positive constant that is much larger than any link capacity in the network

Note that, in 3G networks, the traffic between Node Bs and the switching center (SGSN) is not necessarily symmetric. Thus the traffic in each direction of a node-pair may have different bandwidth requirements. The set of traffic demands  $D$  is defined for traffic between all base stations and the SGSN. In order to allow asymmetric traffic demand between node-pairs, traffic demand in each direction of a node-pair can be separately defined. For example, the  $k^{\text{th}}$  demand in  $D$  is associated with an origin BS node  $O[k]$  and a target SGSN node  $T[k]$ . Similarly, the  $(k+1)^{\text{th}}$  demand in  $D$  could be associated with an origin SGSN node  $O[k+1]$  and a target BS node  $T[k+1]$ . The amount of traffic associated with the  $k^{\text{th}}$  demand node-pair is aggregate traffic  $(a^k)$  which estimates the data rate needed. The  $k^{\text{th}}$  traffic demand node-pair is also associated with a RNC node and the traffic must be processed at the assigned RNC node  $c \in C$  if  $y_{kc}$  equals 1.

Since the traffic demand of each node-pair in  $D$  is an aggregate traffic, it is required that the amount of capacity given to an aggregate traffic routed through a link in the backhaul network must satisfy the demand and QoS requirements. A simple approach to accommodate those requirements is to set the link capacity equal to the total peak rate. However, the cost of over dimensioning the network can be expensive and we adopt a more efficient approach based on an equivalent bandwidth guarantee to satisfy QoS requirements [15]. A typical type of equivalent bandwidth calculation characterizes source traffic by its peak rate ( $R$ ), utilization factor ( $\rho$ ), and mean burst period ( $b$ ). The equivalent capacity estimation ( $\hat{c}_i$ ) for each traffic source is then estimated by

$$\hat{c}_i = \frac{\omega b(1-\rho)R - q + \sqrt{(\omega b(1-\rho)R - q)^2 + 4q\omega b\rho(1-\rho)R}}{2\omega b(1-\rho)}$$

Where  $\omega = \ln(1/\varepsilon)$ ,  $q$  is the buffer size, and  $\varepsilon$  is the packet loss ratio due to buffer overflow. These parameters are specified as part of the QoS requirements. The equivalent capacity (EC) for aggregate traffic is then equal to

$$EC = \min\{m + \theta \cdot \sigma, \sum_{i=1}^{\eta} \hat{c}_i\} \quad \text{with} \\ \theta = \sqrt{-2\ln(\varepsilon) - \ln(2\pi)}, \quad m = \sum_{i=1}^{\eta} m_i, \quad \sigma^2 = \sum_{i=1}^{\eta} \sigma_i^2$$

where  $\eta$  is the number of multiplexed connections,  $m$  is the mean aggregate bit rate and  $\sigma$  is the standard deviation of aggregate bit rate corresponding to all the connections routed on the same link. Note that other formulations for the equivalent capacity calculation can also be used in the network design model. The equivalent capacity of aggregate traffic at a Node B can be estimated by using the service area's user profiles.

### A. 3G Minimum-cost Network Design Problem

The inputs to the network design model include a set of nodes ( $N$ ), a set of potential edges or links ( $E$ ), a set of traffic demands ( $D$ ) for each node-pair with aggregate traffic ( $a^k$ ), a set of traffic flow to RNC assignment parameters ( $y_{kc}$ ), a traffic route hop count limit ( $HP$ ), the cost for establishing each potential link ( $FC_{ij}$ ), and the cost of adding units of capacity to a link ( $VC_{ij}$ ). In addition, the input parameters for the equivalent bandwidth calculation are specified as part of the QoS requirements. The output from the design model is a minimum-cost network topology which includes a set of selected links ( $\varphi_{ij}$ ), capacity of each link ( $w_{ij}$ ), and working paths of traffic demands ( $e_{ij}^{k,l}$ ). Given the notation above, the network design problem can be formulated as.

$$M1: \quad \text{Minimize} \quad \sum_{ij \in E} VC_{ij} \cdot w_{ij} + \sum_{ij \in E} FC_{ij} \cdot \varphi_{ij} \\ e_{ij}^{k,l}, \varphi_{ij}, w_{ij}$$

Subject to:

$$\sum_{j \in N(i)} e_{ij}^{k,l} - \sum_{j \in N(i)} e_{ji}^{k,l} = 1 \quad ; \forall k \in D, i = O[k], l = 1 \quad (1.1)$$

$$\sum_{j \in N(c)} e_{cj}^{k,l} - \sum_{j \in N(c)} e_{jc}^{k,l} = -y_{kc} \quad ; \forall k \in D, \forall c \in C, l = 1 \quad (1.2)$$

$$\sum_{j \in N(i)} e_{ij}^{k,l} - \sum_{j \in N(i)} e_{ji}^{k,l} = 0 \quad ; \forall k \in D, \forall l \in \{1,2\} \\ i \in N - (\{O[k], T[k]\} \cup C) \quad (1.3)$$

$$\sum_{j \in N(c)} e_{cj}^{k,l} - \sum_{j \in N(c)} e_{jc}^{k,l} = y_{kc} \quad ; \forall k \in D, \forall c \in C, l = 2 \quad (1.4)$$

$$\sum_{j \in N(i)} e_{ij}^{k,l} - \sum_{j \in N(i)} e_{ji}^{k,l} = -1 \quad ; \forall k \in D, i = T[k], l = 2 \quad (1.5)$$

$$\sum_{l \in \{1,2\}} \sum_{ij \in E} e_{ij}^{k,l} \leq HP \quad ; \forall k \in D \quad (1.6)$$

$$w_{ij} - EC \left( \sum_{k \in D, l \in \{1,2\}} e_{ij}^{k,l} \cdot a^k \right) \geq 0 \quad ; \forall ij \in E \quad (1.7)$$

$$w_{ij} - \Psi \cdot \varphi_{ij} \leq 0 \quad ; \forall ij \in E \quad (1.8)$$

$$\varphi_{ij} - \varphi_{ji} = 0 \quad ; \forall ij \in E \quad (1.9)$$

$$\varphi_{ij} \in \{0,1\} \quad ; \forall ij \in E \quad (1.10)$$

$$e_{ij}^{k,l} \in \{0,1\} \quad ; \forall k \in D, l \in \{1,2\}, \forall ij \in E \quad (1.11)$$

$$w_{ij} \in R^+ \quad ; \forall ij \in E \quad (1.12)$$

In model M1, the objective function is to minimize the total cost which includes the cost of establishing new edges and working capacity on each edge in the topology. Constraint sets (1.1)-(1.5) are traffic flow constraints between Node B and the switching center. Each  $k$ th traffic demand pair (O[k], T[k]) in set D will be routed from the source node O[k] to the corresponding target node T[k] to determine the working path. Note that the traffic flow between the uplink and downlink in 3G networks is not necessarily symmetric and the working path in one direction (e.g., Node B to SGSN) may be different from the other direction (e.g., SGSN to Node B). Nevertheless, the traffic flow must be routed through its assigned RNC. This traffic routing is modeled with a two-level approach. In the first level ( $l=1$ ), constraint sets (1.1) and (1.2) ensure that the  $k$ th traffic flow is routed from the source node O[k] to its assigned RNC  $c \in C$ , if the given parameter  $y_{kc}$  is set to 1. In the second level ( $l=2$ ), constraint sets (1.4) and (1.5) ensure that the  $k$ th traffic flow continues from its assigned RNC to the corresponding target 3G GPRS support node T[k]. Constraint set (1.6) ensures that the route of each traffic flow can not exceed a maximum number of hops given by HP. Constraint set (1.7) defines the required edge (link) capacity for all traffic flows over the edge, ensuring the amount of capacity given to an aggregate traffic satisfies traffic demands and their QoS requirements. Constraint sets (1.8)-(1.9) ensure that the required capacity is assigned on an edge that exists in the network. Constraint sets (1.10)-(1.12) express the binary requirements and non-negativity of variables. The solution of problem formulation M1 will give the minimum-cost network design and is used as the input to the second phase where the mesh restorable network design problem is solved.

### B. Survivable Network Design Problem

In the survivable network design problem, we modify the minimum-cost network to protect against any single link failure. An alternate route for traffic with sufficient spare capacity is determined between two end-nodes of each protected link (edge) to provide a backup path in the event of a link failure. We define the following additional variables for the problem formulation.

$W$  Set of working edges to be protected between nodes in  $N$ ,  
 $W \subseteq E$

$b_{ij}^{st}$  Backup path binary decision variable that equals 1 if edge  $(i, j) \in E$  from node  $i$  to  $j$  is used for restoration due to the failure of edge  $(s, t) \in W$ ; 0 otherwise

$s_{ij}$  Spare capacity of edge from node  $i$  to  $j$ ,  $s_{ij} \in R^+$

$\alpha_{ij}$  Mobility factor indicating the estimated proportion of traffic demands served at BS node  $j$  due to user mobility after its neighbor BS node  $i$  failed,  $0 \leq \alpha_{ij} \leq 1$  and  $\sum_j \alpha_{ij} \leq 1$

$\beta^k$  Amount of extra bandwidth assigned to a BS node associated with the  $k^{\text{th}}$  demand pair in  $D$  to absorb increased traffic demands due to user mobility after a neighbor BS failure,

$$\beta^k = \text{MAX}_{k' \in D: i=O[k'] \in B} (\alpha_{ij} a^{k'}) \quad ; \forall k \in D, j = O[k] \in B$$

$HB$  Hop count limit for backup routes

The inputs to the survivable network design model include a set of working links (edges) to be protected, a set of working paths for all traffic demand node-pairs in the existing network, the estimated mobility factors, and the backup route hop count limit. Given the notation defined previously, the span-restoration problem can be formulated as follows.

$$M2: \text{Minimize}_{b_{ij}^{st}, s_{ij}, \varphi_{ij}} \sum_{ij \in E} VC_{ij} \cdot s_{ij} + \sum_{ij \in \{E-W\}} FC_{ij} \cdot \varphi_{ij}$$

Subject to:

$$\sum_{j \in N(s)-\{t\}} b_{sj}^{st} - \sum_{j \in N(s)-\{t\}} b_{js}^{st} = 1 \quad ; \forall st \in W \quad (2.1)$$

$$\sum_{j \in N(i)-\{s\}} b_{ij}^{st} - \sum_{j \in N(i)-\{t\}} b_{ji}^{st} = 0 \quad ; \forall st \in W, \forall i \notin \{s, t\} \quad (2.2)$$

$$\sum_{j \in N(t)-\{s\}} b_{ij}^{st} - \sum_{j \in N(t)-\{s\}} b_{jt}^{st} = -1 \quad ; \forall st \in W \quad (2.3)$$

$$\sum_{ij \in E} b_{ij}^{st} \leq HB \quad ; \forall st \in W \quad (2.4)$$

$$s_{ij} - (b_{ij}^{st} \cdot w_{st} + b_{ij}^{ts} \cdot w_{ts} + \sum_{k \in D \setminus \{1,2\}} \beta^k \cdot e_{ij}^{kl}) \geq 0 \quad ; \forall st \in W, \forall ij \in E - \{st, ts\} \quad (2.5)$$

$$s_{ij} - \Psi \cdot \varphi_{ij} \leq 0 \quad ; \forall ij \in E - W \quad (2.6)$$

$$\varphi_{ij} - \varphi_{ji} = 0 \quad ; \forall ij \in E - W \quad (2.7)$$

$$\varphi_{ij} = 1 \quad ; \forall ij \in W \quad (2.8)$$

$$\varphi_{ij} \in \{0, 1\} \quad ; \forall ij \in E \quad (2.9)$$

$$b_{ij}^{st} \in \{0, 1\} \quad ; \forall st \in W, \forall ij \in E - \{st\} \quad (2.10)$$

$$s_{ij} \in R^+ \quad ; \forall ij \in E \quad (2.11)$$

The objective of M2 is to minimize the total cost of spare resources needed to satisfy survivability requirements. Constraint sets (2.1)-(2.3) are flow-balance constraints for each backup route of edges in the original topology. Constraint set (2.4) limits the number of hops for each backup route. This is to limit the path delay and reduce the computational complexity. Constraint set (2.5) defines the required spare capacity to satisfy all simultaneous flows over the edge  $(i, j) \in E$  in the direction from node  $i$  to  $j$  due to a failure of any link  $(s, t) \in W$ . The spare capacity on the edge includes extra bandwidth to absorb additional traffic due to user mobility from adjacent BS failures or backhaul link failures that may cause additional traffic on the edge. Constraint sets (2.6)-(2.7) ensure that the spare capacity is assigned on an edge in the new topology. Constraint set (2.8) guarantees that links in the phase one design are included in the mesh restorable topology. Constraint sets (2.9)-(2.11) express the binary requirements and non-negativity of decision variables. The solution of problem M2 yields a mesh network topology, with a backup route for each protected link having sufficient spare capacity.

Mixed integer programming problems M1 and M2 can be solved using a standard branch and bound technique. While the optimization models can be solved for small network sizes providing a useful benchmark solution, in general the problems M1 and M2 are NP-hard and the branch and bound technique will not scale.

## IV. HEURISTIC ALGORITHM

This section presents a heuristic algorithm to solve the design models M1 and M2 presented above. We propose a heuristic based on a randomized iterative minimum-cost routing approach. The basic idea of the heuristic method is to sequentially route each traffic de-

mand using a shortest-path routing algorithm with link-cost metric calculated from the routed traffic demand for all potential links. The process of sequentially routing each traffic demand is randomly repeated until there is no change in the network topology or the maximum iterations is reached. A similar routing based heuristic approach has been efficiently applied to determine spare capacity requirements in survivable network design problems in [19,20].

The proposed heuristic is applied to both design phases, with some modifications as detailed in [18]. In the phase 1 minimum-cost design heuristic algorithm, input data consist of a set of all traffic demands ( $D$ ), required equivalent bandwidth of each traffic demand ( $TD[1..|D|]$ ), a set of all potential links ( $E$ ), fixed cost ( $FC[1..|E|]$ ) and variable cost ( $VC[1..|E|]$ ) of potential links. Output from the algorithm includes the working path of each traffic demand ( $WPath[1..|D|]$ ) and the bandwidth on each link ( $WLinkBW[1..|E|]$ ). The phase one heuristic algorithm is shown in Figure 3 and begins with randomly generating an ordering of traffic demands. From the generated order, each demand is sequentially routed with a minimum-cost routing algorithm based on the link-cost metric for each potential link. The link cost is calculated from the fixed charge and the traffic based variable charge. For each traffic demand, it is first checked whether a working path has been previously assigned to the traffic demand. If a working path exists for the traffic demand, its required bandwidth will be removed from working links before a new path is searched for the demand. Next, the link cost metric for each potential link is calculated from the fixed cost and variable cost of using the link by the routed traffic demand. Then, a path will be determined using a shortest-path routing algorithm based on the calculated link-cost metric. If the computed path is a new minimum-cost path, it is recorded as the working path for the traffic demand. Finally, the required bandwidth of the traffic demand is allocated on every link along the working path. The link-cost metric is recalculated for each flow and the network topology is updated once a flow is routed using a new path. After all traffic demands are routed, the process is repeated with a new randomly generated order of traffic flows until there is no update in the network topology or the maximum number iterations is reached. At the end of the algorithm, the network topology will be determined from the potential links that have non zero traffic demands, along with the capacity of each link. The solution found in the phase 1 design heuristic is used as the input for the survivable network design heuristic in the second phase.

The phase 2 heuristic is similar to the minimum-cost design heuristic algorithm. However, each working link from the minimum-cost network topology will be considered as a failure scenario and traffic demands passing through the link will be protected by a backup path. Also, spare capacity allocated to each link in the link restoration design can be shared over different disjoint failure scenarios. A spare provision matrix [18-20] in the algorithm maintains information about spare capacity requirements on each potential link due to any working link failure scenario. The spare capacity required on each potential link is the maximum bandwidth among all failure scenarios that use the link in the backup path. Each backup path is computed using a shortest path routing algorithm with calculated link-cost metric for all potential links. In the link-cost metric calculation, the amount of bandwidth required on each potential link is equal to the capacity of working link to be protected minus the spare capacity already reserved on that potential link. If the reserved spare capacity on a potential link is greater than the capacity of protected link, there is no additional cost of using the potential link in the backup path. Note that the working (protected link) and backup path must be disjoint so that traffic can be rerouted to the backup path in the event of working link failure. A hop count limit is included in the routing algorithm to avoid choosing backup paths with large number of hops which may introduce an undesirable delay. Lastly, extra bandwidth

## Variable Definition

$D$  :set of traffic demands or equivalent bandwidth flows;  
 $E$  :set of all potential links;  
 $TD[1..|D|]$  :required bandwidth of each traffic demand;  
 $WLinkBW[1..|E|]$  : bandwidth on each link;  
 $LinkCost[1..|E|]$  :cost of each potential link;  
 $FC[1..|E|]$  :fixed cost of each potential link;  
 $VC[1..|E|]$  :variable cost of each potential link;  
 $WPath[1..|D|]$  :working path of each traffic demand;  
 $Seq[1..|D|]$  :sequence of each traffic demand;

## Initialization

```
iteration ← 0;
for each traffic demand  $k$  in  $D$ 
     $WPath[k] \leftarrow \text{NULL}$ ;
for each link  $e$  in  $E$ 
     $WLinkBW[e] \leftarrow 0$ ;
```

## Begin

```
repeat {
    no_update ← TRUE;
    randomly generate a sequence  $Seq[1..|D|]$  of
    traffic demand node-pairs in  $D$  to be routed;
    //sequentially route traffic demands in  $D$ 
    for each traffic demand  $k$  in  $Seq[1..|D|]$ 
    {
        if ( $WPath[k] \neq \text{NULL}$ ) {
            //remove  $TD[k]$  on links in  $WPath[k]$ 
            for each link  $e$  in  $WPath[k]$ 
                 $WLinkBW[e] \leftarrow WLinkBW[e] - TD[k]$ ;
        }
        //calculate new link cost metric  $LinkCost[1..|E|]$ 
        for each link  $e$  in  $E$  {
            if ( $WLinkBW[e] > 0$ )
                 $LinkCost[e] \leftarrow VC[e]*TD[k]$ ;
            else
                 $LinkCost[e] \leftarrow FC[e] + VC[e]*TD[k]$ ;
        }
        //find the minimum-cost path for demand  $k$ 
         $NewPath[k] \leftarrow \text{ShortestPath}(k, LinkCost[1..|E|])$ ;
        if ( $NewPath[k] \neq WPath[k]$ ) {
             $WPath[k] \leftarrow NewPath[k]$ ;
            no_update ← FALSE;
        }
        //add  $TD[k]$  to links in  $WPath[k]$ 
        for each link  $e$  in  $WPath[k]$ 
             $WLinkBW[e] \leftarrow WLinkBW[e] + TD[k]$ ;
    }
    iteration ← iteration + 1;
} until (no_update = TRUE or iteration > max_iteration);
```

## End

Figure 3. Minimum-cost design heuristic algorithm.

for mitigating the mobility effects due to BS failures is added to links along the working path of each traffic demand. At the end of the algorithm, the solution obtained from the phase 2 heuristic includes backup paths for protected links, new established links and spare capacity in each link to recover from a network failure. Due to space limitations the details of the phase 2 algorithm are given in [18]. Also the computational complexity of the heuristic algorithms described above is shown to be polynomial in nature in [18].

## VI. NUMERICAL RESULTS

We present numerical results of solving the network design model using our two phase optimization model formulation ( $M1$  and  $M2$ ) and the corresponding two phase heuristic for different sample networks. Here we highlight some of the extensive numerical results from [18] for the sample networks with parameter values shown in Table 1. In the sample networks, BS are placed in the network service area using a distance based method so that each BS covers a small region and has a minimum normalized distance of 3 units to its neighbor BS. Note that the distance unit is normalized and can be represent any appropriate unit such as km. We assume that the fixed cost ( $FC_{ij}$ ) for adding a new edge is a linear function  $FC_{ij} = c_1 + c_2 u_{ij}$ , where  $c_1$  is a fixed termination charge ( $c_1$  is set to 15),  $c_2$  is the charge per distance unit ( $c_2$  is also set to 15), and  $u_{ij}$  is the Euclidean-distance between two end-nodes ( $i$  and  $j$ ) of the edge. The variable cost ( $VC_{ij}$ ) of adding a bandwidth unit (e.g., 64 kbps) to an edge is also a linear function  $VC_{ij} = c_3 u_{ij}$ , where  $c_3$  is set to 1. **Table II** lists the characteristics of traffic sources used for randomly generating the amount of bandwidth required at each BS in the experiments.

In our experiments, we first compare the results obtained from solving the optimization models  $M1$  and  $M2$  using the CPLEX 7.1 optimization solver implementing a branch and bound solution technique [16] with results from the heuristic algorithm. The CPLEX 7.1 solver was run on a Sun Blade1000 workstation with 750 MHz processor and 2 GBs of memory and the heuristic algorithm implemented in C++ was run on a PC with Intel Pentium III 750 MHz processor and 128 MBs of memory. Note that due to the random ordering of flows in the heuristic, the algorithm can give different results depending on the initial starting point. Hence the results reported here are the best solution found from 64 different runs using different random number seeds to generate the ordering of the traffic demands for routing.

Following the two phase design approach one first solves the minimum-cost network design problem. Table III shows results of solving the phase 1 model  $M1$  using the CPLEX solver and the heuristic algorithm for two sample networks with different traffic demands given as total peak bandwidth at BS nodes. Traffic demands are divided into three groups: 2-6 Mbps, 12-18 Mbps and 30-36 Mbps. These total peak bandwidths are estimated from the maximum radio bandwidths available for UMTS 3G systems. All traffic demands are assumed to be equal in both directions of demand node-pairs. The numerical results in the table show that the proposed heuristic can find near-optimal solutions to the  $M1$  model. It is worth noting that the computational time to solve  $M1$  for larger networks was prohibitive.

The solution obtained from the phase 1 design is used as the input for the second phase survivable network design problem  $M2$ . In order to fairly compare the results obtained from the heuristic algorithm versus those of the CPLEX solver for the survivable network design problem, we use the optimal solution obtained from the CPLEX solver for the minimum-cost design  $M1$  as the input to both. **Table IV** shows the results of solving the phase 2 design problem with different number of protected links, where links are randomly selected to be protected, for the N30 network. Since the  $M2$  optimization problem is NP-hard and can not be solved for large problem size,

a LP relaxation technique is used to give a lower bound infeasible solution for comparison. From the results shown, the heuristic method can give good solutions. Note that the computational time of the heuristic algorithm in **Table IV** is the total time to find the best solution from all 64 runs using different random seed numbers to generate the sequence of routing traffic demands.

To examine the computational time of the heuristic method for larger problem sizes, we tested the algorithm with our three largest sample networks. **Table V** shows the average computational time of one run. The time is based on the average of 64 runs with different random seed numbers over three different traffic loads. One can see that the heuristic can scale to solve practical size network problems

In the results discussed above, we have not considered the cost of extra bandwidth to absorb the traffic due to user mobility (i.e.,  $\alpha_{ij} = 0$  for all BS) after a BS failure. Table VI shows the total network design cost of solving the two-phase design models using the heuristic method for different mobility factor parameter settings in the N50 and N100 networks. The random movement column in Table VI refer to cases where mobile users can move in any direction after a failure and the deterministic movement column refers to cases where a large number of mobile users move in the same direction or follow the same path. The deterministic movement is comparable to a highway travel in the real environment. To represent the deterministic movement environment, a road map was specifically defined in the network service area as illustrated in Figure 4. In this paper, we set the mobility factor ( $\alpha_{ij}$ ) associated with each BS to a value within a given range by inspecting the location of each BS in the network map. If a BS is located close to its neighbor BS, a higher mobility factor value is assigned. For the case of deterministic movement, BSs serving mobile users along the road are be assigned a high mobility factor value. From the numerical results shown, the cost of the network design incorporating extra bandwidth to mitigate the effect of user mobility in the event of BS failures increases. The results also show that the deterministic movement environment needs special attention as the percentage of cost increase is about twice that of the random movement case.

## VII. CONCLUSIONS

In this paper, we have considered the problem of designing survivable 3G wireless packet based backhaul networks. A network design model was proposed that aims to find a wireless backhaul network topology that can provide acceptable quality of service and survivability of services to mobile users, while minimizing the network design cost. The network design strategy is based on mesh-based restorable network design. Our approach to the problem is a two-phase design. In the first phase, the minimum-cost network design to satisfy traffic demands and QoS requirements is determined. In the second phase, an incremental network design problem is considered for the minimum-cost network where the topology is augmented with new links for backup routes and spare capacity is allocated to satisfy survivability requirements. The spare capacity includes the bandwidth needed to recover from any single link failure as well as the spare bandwidth to mitigate the effects of user mobility after base station failures. In order to scale the design model with network size a computationally efficient heuristic method based on a iterative randomized minimum cost routing method was proposed. Numerical results showed the computationally efficiency and accuracy of the heuristic.

### VIII. REFERENCES

- [1] 3GPP, Universal Mobile Telecommunication System (UMTS): QoS Concept and Architecture, TS 23.107 v.5.5.0, <http://www.3gpp.org>, 2002.
- [2] B. Walke, P. Seidenberg, M. Althoff, UMTS The Fundamentals. John Wiley & Sons, Ltd, West Sussex, England, 2003
- [3] A. Snow, U. Varshney, and A. Malloy, "Reliability and survivability of wireless and mobile networks," *IEEE computer*, vol. 33, pp. 49-55, July 2000.
- [4] D. Tipper, T. Dahlberg, H. Shin, and C. Charnsripinyo, "Providing fault tolerance in wireless access networks," *IEEE Communication Magazine*, vol. 40, no. 1, pp. 58-64, January 2002.
- [5] D. Tipper, C. Charnsripinyo, H. Shin, and T. Dahlberg, "Survivability analysis for mobile cellular networks," *Proceedings Communication Networks and Distributed Systems Modeling and Simulation Conference 2002*, San Antonio, Texas, Jan. 27-31, 2002.
- [6] Amitava Dutta, and Peter Kubat, "Design of partially survivable networks for cellular telecommunication systems," *European Journal of Operational Research*, vol. 118, pp. 52-64, 1999.
- [7] Louis A. Cox, Jr., and Jennifer R. Sanchez, "Designing least-cost survivable wireless backhaul networks," *Journal of Heuristics*, vol. 6, pp. 525-540, 2000.
- [8] P. Kubat, J. M.G. Smith, and C. Yum, "Design of Cellular Networks with Diversity and Capacity Constraints," *IEEE Trans. On Reliability*, Vol. 49., No. 2., June 2000
- [9] C. Charnsripinyo and D. Tipper, "Topological Design of Survivable Wireless Access Networks," *Proceedings Fourth International Workshop on the Design of Reliable Communication Networks, (DRCN 2003)* Oct. 20-22, 2003, Banff, Canada
- [10] Y. Ma, J. Han, K. Trivedi, "Composite Performance and Availability Analysis of Wireless Communications," *IEEE Trans. on Vehicular Technology*, Vol. 50, No. 5, pp. 1216 – 1223, Sept., 2001
- [11] A. Malloy, U. Varshney, and A. Snow, "Supporting Mobile Commerce Applications Using Dependable Wireless Networks," *Mobile Networks and Applications*, Vol. 7, pp. 225-234, June, 2002
- [12] A. Szlovenscak, I. Godor, and J. Harmatos, "Planning reliable UMTS terrestrial access networks," *IEEE Communication Magazine*, vol. 40, no. 1, pp. 66-72, January 2002.
- [13] T. Dahlberg, D. Tipper, B. Cao, and C. Charnsripinyo, "Survivability in wireless mobile networks," in *Reliability, Survivability and Quality of Large Scale Telecommunications Systems*, P. Stavroulakis, Ed. John Wiley & Sons, Ltd., West Sussex, England, pp. 81-118, 2003.
- [14] Wayne D. Grover, and John Doucette, "Topological design of survivable mesh-based transport networks," *Annals of Operations Research*, vol. 106, pp. 79-125, 2001.
- [15] R. Guérin, H. Ahmadi, and M. Naghshineh, "Equivalent capacity and its application to bandwidth allocation in high-speed networks," *IEEE Journal on Selected Areas in Communications*, vol. 9, no. 7, pp.968-981, September 1991.
- [16] ILOG, "ILOG CPLEX 7.1", <http://www.ilog.com/>.
- [17] W. D. Grover, *Mesh-based Survivable Networks: Options and Strategies for Optical, MPLS, and ATM Networking*, Prentice Hall PTR, 2003
- [18] C. Charnsripinyo, "Design of Survivable Wireless Access Networks," *Ph.D. Telecommunications Dissertation*, University of Pittsburgh, Dec., 2003
- [19] Y. Liu, D. Tipper, and P. Siripongwutikorn, "Approximating Optimal Spare Capacity Allocation by Successive Survivable Routing," *ACM/IEEE Transactions on Networking*, Vol. 13., No. 1, pp. 198-211, Feb., 2005
- [20] Y. Liu and D. Tipper, "Apparatus and Method for Spare Capacity Allocation", *U.S. Patent 6,744,727 B2*, June 1, 2004
- [21] S. Matz, L. Volta and M. Malkawi, "Analysis of failure and recovery rates in a wireless telecommunications system," *Proceedings of IEEE DSN*, pp. 687-693, 2000

Table I. Characteristics Of Sample Networks

Network	Number of BS/Node B	Number of RNC	Number of potential links	Service area (unit <sup>2</sup> )
N30	30	3	202	400
N50	50	5	548	900
N100	100	10	979	2500
N200	200	15	1268	4900

Table II. Characteristics Of Traffic Sources

Peak rate (kbps)	Utilization	Mean Burst Time	Buffer (kilobits)	Effective Bandwidth (kbps)
15	0.8	0.5s	32	12.954
30	0.8	0.5s	64	25.907
60	0.6	1s	128	51.598
120	0.6	1s	256	103.196
240	0.5	1.5s	512	213.960
480	0.5	1.5s	1024	427.921
960	0.4	2s	2048	876.691
1920	0.4	2s	4096	1753.382

Table III. Phase 1 Design Cost

Network	Peak Traffic Rate (Mbps)	Minimum Cost Design $M_I$		
		CPLEX	Heuristic	Gap (%)
N30	2-6	52690	53421	1.387
	12-18	176397	177548	0.652
	30-36	358904	361718	0.784
N50	12-18	269713	271461	0.648
	30-36	552286	554116	0.331

Table V. Average Computational Time

Network	Phase 1 Design	Phase 2 Design
N50	0.4s	11.3s
N100	2.1s	90.9s
N200	10.1s	539.6s



Table IV Results of Survivable Network Design Phase for N30 Network with Peak Bandwidth 2-6 Mbps

Network	Number of Protected Edges	CPLEX LP Relaxation		CPLEX Branch & Bound		Heuristic	
		Cost	Time (hh:mm:ss)	Cost	Time (hh:mm:ss)	Cost	Time (hh:mm:ss)
N30	5	7169	7s	7591	56s	7819	22s
	10	24502	13s	27242	1h:59m:10s	28022	28s
	15	28786	44s	N/A	72h:00m:00s	35491	36s
	20	38766	1m:25s	N/A	72h:00m:00s	54291	50s
	25	41462	2m:46s	N/A	72h:00m:00s	58165	1m:5s
	33(all edges)	49932	11m:23s	N/A	72h:00m:00s	75463	1m:38s

Table VI. Total Cost of Wireless Backhaul Network Design

Network	Traffic Demand (Mbps)	Total Network Design Cost						
		No Extra Bandwidth ( $\alpha_{ij} = 0$ )	Random Movement				Deterministic Movement	
			$\alpha_{ij} = 0.02-0.1$	$\alpha_{ij} = 0.1$	$\alpha_{ij} = .01-0.15$	% Increase	$\alpha_{ij} = 0.01-0.3$	% Increase
N50	2-6	157559	167674	170171	171409	8.79	179324	13.81
	12-18	521720	557440	564077	571391	9.52	602582	15.50
	30-36	1045987	1123371	1133295	1154865	10.41	1220938	16.73
N100	2-6	464601	499118	505760	509036	9.56	527506	13.54
	12-18	1483756	1595630	1612712	1634223	10.14	1699954	14.57
	30-36	2982122	3224469	3247966	3314447	11.14	3462832	16.12

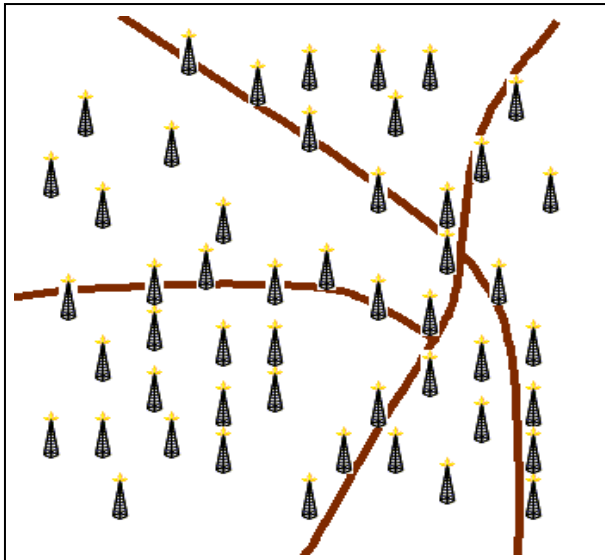


Figure 4(a) NodeB locations and road map in N50 network.

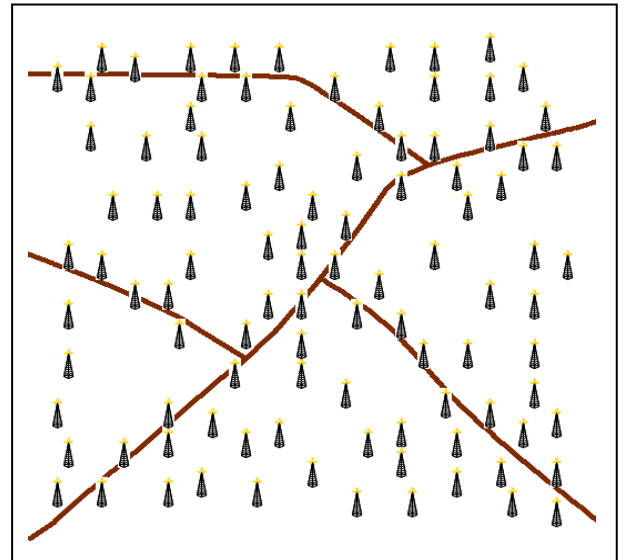


Figure 4(b) NodeB locations and roads in N100 network.