

An Architecture for IP-based Next Generation Radio Access Network

Ram Dantu, and Parthasarathy Gudur,

*Department Of Computer Science & Engineering and Department of Electrical Engineering
University of North Texas, Denton, Texas, USA*

ABSTRACT

High call volumes due to novel mobile data applications necessitate development of next generation wireless networks centered on high performing and highly available radio access networks (RANs). In this paper, we present an innovative IP-based wireless routing architecture (for a RAN) with mechanisms for seamless handoff operations and high Quality of Service (QOS). Algorithms for dynamic configuration of the RAN, and efficacious network bandwidth management through traffic control are also presented. We establish the superiority of our system with real-life data indicating significant cost and availability improvements with our system over the traditional networks

KEYWORDS

IP-based RAN, QOS, bandwidth management, traffic management, next generation networks.

INTRODUCTION

Explosive growth in mobile data applications in recent times has motivated the mobile operators to explore efficacious and cost effective solutions for handling the backhaul traffic (from base-stations to their core networks). These solutions for the B3G (Beyond 3rd Generation) wireless networks include bandwidth optimizations in the current radio access networks (RANs) as well as IP (Internet Protocol)-based RANs over the T1/E1 and Ethernet backhaul for 1X-EVDO CDMA (1 time radio transmission technology Code Division Multiple Access) (TIA/EIA/IS-2000.1.A -2, 2002) and HSPA W-CDMA (High Speed Packet Access Wideband CDMA) networks (3GPP2 C-S0033 Rev 0 Ver 2.0, 2003). We propose here IP-based RANs which replace legacy mobile wireless systems with innovative wireless routers (WRs) equipped with traffic management and self-configuration capabilities. These attributes together with high system availability and high QOS makes our IP RAN an ideal choice for next generation mobile wireless networks.

The proposed IP-based RAN architecture can be adapted to new long term evolution (LTE) standard for mobile communications (Ekström *et al.*, 2006, Ergen, 2009, Fazel & Kaiser, 2009) based on orthogonal frequency division multiple access (OFDMA) technology developed by 3rd generation partnership project (3GPP). The evolved packet system (EPS) for the evolved universal mobile telephone system (UMTS) terrestrial radio network (E-UTRAN), which forms the access side of the LTE architecture, will not impact the IP-architecture proposed here. However, our proposal could be compatible with the IP-based flat core network architecture of LTE, variously known as extended packet core (EPC) and System Architecture Evolution (SAE). As a matter of fact, since LTE release 8's air interface is intended for use over any IP-network, compatibility of our IP architecture is no issue at all. Now that LTE has been ratified in March 2009, adaptation of the current work to LTE can be carried out.

Worldwide Interoperability for Microwave Access (WiMAX) is another competitive next generation technology for mobile communications (Ergen, 2009, Fazel & Kaiser, 2009). It is based on the IEEE 802.16 standard for broadband wireless access. The WiMAX forum proposed an architecture for integration of a WiMAX network with an IP core network. This makes it possible to adapt our IP architecture to WiMAX networks also.

BACKGROUND

Since IP-based RANs is an emerging technology, the literature on flexible and reconfigurable next generation RAN architectures based on versatile wireless routers is rather sparse. An interesting paper in this direction, Ghosh, Basu & Das (2005) provides an insight into a flexible, reliable, and cost effective IP-based RAN architecture with BSs connected to form a mesh. Bu, Chan & Ramjee (2006) discuss connectivity, performance and resiliency of IP-based CDMA RANs focusing on a star topology between the BSs and RNC (Radio Network Controller). Vassiliou *et al* (2002) propose a RAN for next generation wireless networks based on multi-protocol label switching (MPLS) and hierarchical mobile IP. Kempf & Yegani (2002) propose a new architecture called open-RAN for mobile wireless IP RANs. Yasukawa, Nishikido, & Hisashi (2001) propose an IPV6-based wireless routing architecture with scalable mobility and quality of service (QOS) support. Chen & Hamalainen (2003) present a method for handovers in IP RANs.

The proposed wireless router (WR) architecture, its mechanisms for efficacious traffic and bandwidth management, and dynamic self-configuration are based on our earlier patented research work (Dantu, 2005; Dantu et al., 2006; Patel et al., 2006).

WIRELESS ROUTER ARCHITECTURE

1. Issues, Controversies, Problems

The legacy mobile wireless architecture shown on the right half of Fig. 1 with a hierarchy of BS (Base Station)s, BSC (Base Station Controller)s, and MSC (Mobile Switching Center)s has a number of shortcomings with respect to handling of large volumes of data in B3G networks: i) wireless frame selection for handoff management is done at BSCs resulting in the duplicate traffic flow on the backhaul, ii) even during the ideal periods of a call, transmission resources are reserved resulting resources wastage in contrast to the IP-networks equipped with the statistical multiplexing scheme, iii) only 15% of the BS-BSC traffic is payload, and the rest is overhead, iv) BSs forward erroneous frames also BSC and this results in dead payload on the backhaul, v) uneven utilization of links makes the system inefficient, cost-ineffective, and unsuitable for deployment of new data

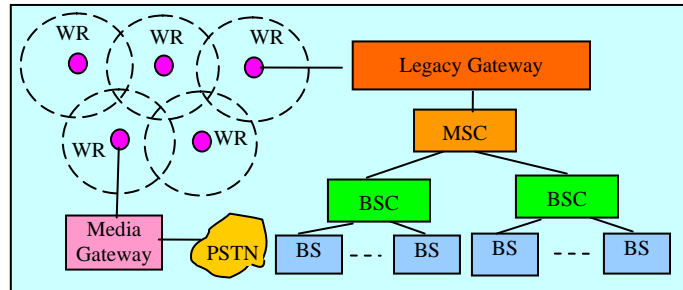


Fig.1. A novel distributed wireless router based RAN communicating with a legacy network (each pink dot represents a wireless router).

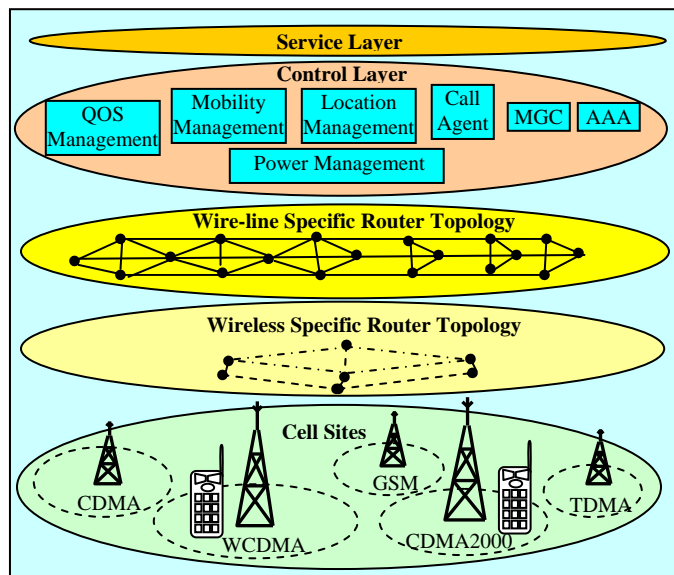


Fig.2. Several overlaying features that need to be integrated into a WR

intensive services, vi) transmission delays in long BS-BSC links could cause soft-handoff failures and call drops and thereby contribute to performance degradation, and vii) single-point failures of the legacy system entities or the links between them results in low system availability.

The demand for high quality services despite large volumes of call traffic necessitates a drastic reduction in the long backhaul (BS-MSC) control and data paths used in the legacy systems. An elegant approach to address the problem is to build the next generation RANs as distributed configurations of simple but functionally comprehensive wireless router (WR)s that could replace and at the same time inter-work, as shown in Fig. 1, with the hierarchy of legacy wireless system entities *i.e.* the BSs, BSCs, and MSCs. Thus these new versatile WRs, need to have integrated into them several overlaying features of an all-IP 3G wireless network depicted in Fig. 2. Even though the functional modules of the service layer are not detailed out in the figure in view of the ever growing number of wireless/wire-line services, typical entities that provide network services are call servers, bandwidth brokers, SLA (Service Level Agreement) managers, billing servers, HLR (Home Location Register)s, HSS (Home Subscriber Server)s, MGW (media Gateway)s, SGW (Signaling Gateway)s, legacy servers, DNS (Domain Name Servers)s, and so on. The control layer supports the services with entities such as QOS (Quality of Service)/Mobility/Location/Power managers, Call Agents, and AAA (Authentication, Authorization, and Accounting) managers. In the legacy systems, these two functional layers roughly correspond to the MSC and BSC functions, respectively. For execution of these functions, the legacy system entities (BSs, BSCs, and MSCs) need to communicate with one another. This is supported by backhaul IP networks with wire-line topologies. A part of the inter-BS communication is supported by wireless routing. Finally, the communication between the BSs at different cell sites and the mobile device constitutes the physical layer functionality of the wireless system.

Since the WR replaces the mobile network entities in the proposed RAN architecture, it needs to incorporate the control and routing functionalities of the legacy systems. In particular, it should have the following features: i) it should support all the data and signaling protocols for inter-router communication as well as communication with various service and control entities, ii) it should facilitate effective hand-off management to achieve nearly zero call drop rate, iii) it should provide high QOS by effective bandwidth management through traffic shaping, and iv) it should be capable of dynamically configuring its operational parameters in collaboration with its neighbors, and adapt itself to RF topology and other changes. Overall, it should render the RAN both high performing and highly available at a low cost.

2. Solutions and Recommendations

In the proposed architecture, the wireless routers (WRs) situated in different cells perform message routing in addition to the traditional Base Station (BS) functions. In fact, the traditional functions of the base stations, BSC (Base Station Controller), and MSC (Mobile Switching Center) in the legacy systems are lumped together and distributed across various WRs in the new wireless architecture. Fig. 3 is a very high level depiction of various inter-router links as well as the control links between the proposed WRs, and the traffic and control interfaces. The inter-router links include: i) wireless specific virtual tunnels *i.e.* Multi Protocol Label Switch (MPLS) paths based on IP packets, ii) RSVP (Resources Reservation

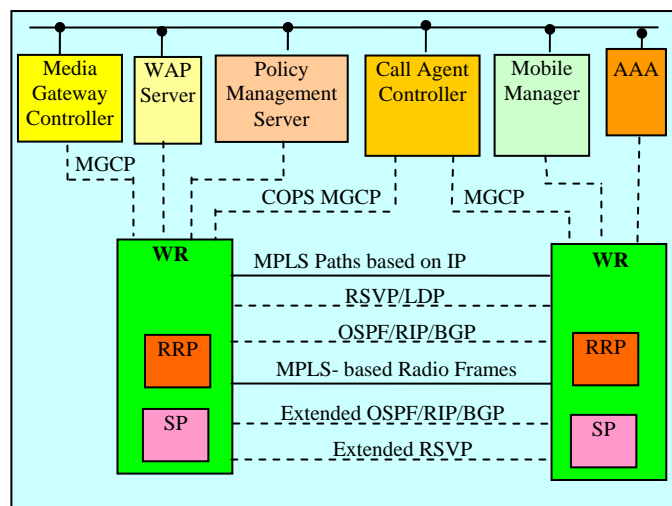


Fig.3. Proposed wireless router links to peer routers and various control interfaces.

Protocol)/LDP (label Distribution Protocol) signaling channel, iii) routing message channel, iv) wireless specific virtual channel to carry MPLS based radio frames, and v) one or more wireless-specific control channels for call setup and maintenance including a signaling channel for usage by any signaling protocol (SP) such as extended RSVP and a routing channel for usage by any radio routing protocol (RRP) such as extended OSPF (Open Shortest Path First), RIP (Routing Information Protocol), or BGP (Boarder Gateway Protocol). The bold and dashed lines in the figure indicate the data and control channels, respectively.

For supporting wireless traffic services, the WRs access traffic and control interfaces that include media gateway controllers, WAP (Wireless Application Protocol) servers, policy management servers, call agent controllers, mobility managers, and AAA (Authentication, Authorization, and Accounting) servers. WRs communicate with these interfaces through MGCP (Media Gateway Controller Protocol), COPS (Common Open Policy Service), and other suitable protocols.

The WR architecture proposed herein facilitates wireless-access technology (*e.g.* CDMA or TDMA) independent network routing with the help of wireless interfaces to disparate wireless peripherals as shown in Fig. 4, and hence is pivotal to an all-IP (Internet Protocol) radio access network (RAN) that

seamlessly inter-works with the backhaul IP network with interfaces to various network peripherals. For communication with the backhaul networks at the other end, the WR includes wire-line interfaces to various network peripherals as well. At the heart of the WR lies the traffic control with various modules as follows: i) Quality of Service (QOS) engine for traffic conditioning and effective management of transmission resources, ii) Selection and Distribution Unit (SDU), iii) Central Processing Unit (CPU), iv) Call Processor, v) Timing Unit for synchronization purpose, vi) Communication Module with various traffic-controller

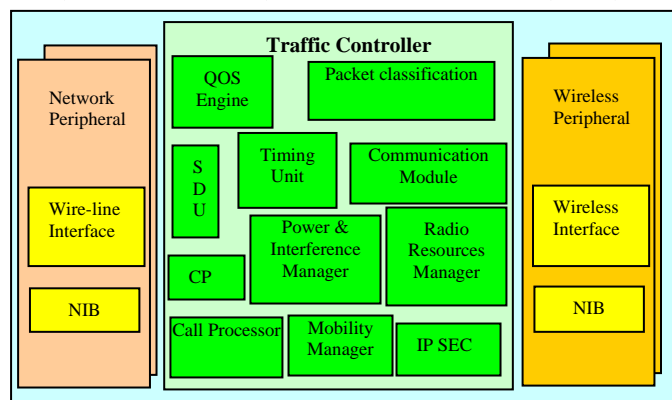


Fig. 4. Internal structure of the proposed wireless router

interfaces to gateways, services, policy managers, IP routers, base-stations, call agents, and other remote nodes and resources, vii) Power and Interference Manager, viii) Radio resources Manager, ix) Mobility Manager, X) Packet Classification Unit, and xi) Security (IP SEC) Module.

In the traditional mobile networks, an SDU is placed centrally at the Base Station Controller (BSC) to manage the call processing at a number of base stations. It selects the best frame from a number of incoming radio frame instances from the mobile via different base-stations (BSs) for onward transmission to the intended destination through backhaul network, and distributes similarly the messages received from the backhaul to the target mobiles. Based on the quality of the frames received from different base-stations, an SDU also manages soft-handovers (that is, call redirections from one BS to the other). In our architecture, BS is replaced by the more versatile WR module. Additionally, it incorporates an innovative concept of a distributed SDU with every WR housing an SDU. These distributed SDUs in the present architecture facilitate distribution of intelligence, and switching and control functionality to individual cell sites.

The main advantage here is that the radio frames need not be transmitted to the BSC over the backhaul links. A very efficacious use of the backhaul network bandwidth this way, in turn, results tremendous cost savings for the customer. Additionally, this approach helps in averting traffic congestion. Common switching points leading to delayed/dropped traffic are reduced if not altogether eliminated. However, these advantages could be realized only by effective inter-router communication methods for mobility management by call redirection (soft handoff) and MPLS path reconfiguration as the mobile device transitions between cells. The proposed innovation should also be implemented without compromising

the quality of service (QOS). In the following subsections, we describe the mechanisms for soft handoff and QOS in the proposed architecture.

A. Call Processing and Soft handoff Mechanism

Figure 5 depicts interactions among various network elements for ingress data (*i.e.* data flowing from the mobile device to the edge router). Figure 6 depicts the interactions for the egress data flowing in the reverse direction. These figures are OO (object-oriented) style sequence diagrams for call flow in an UML (Universal Modeling Language) like notation.

In both the scenarios described here, the first step is designation of one of the routers accessible to the mobile device as the primary router. In case of mobile originated calls, the mobile informs the WR from whom the strongest signal is received to take the responsibility as prime WR. On the other hand, in case of mobile terminated calls, the edge router determines the prime router based on the mobile location after locating and successfully paging the mobile. Once the primary router is determined, it initiates the process for setting MPLS tunnels between itself and the edge router as well as the secondary routers. The key innovation here is

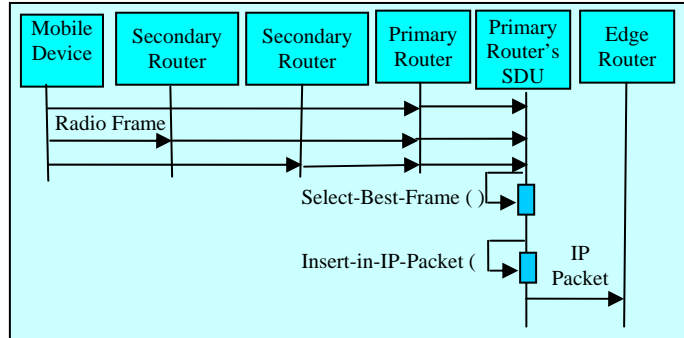


Fig. 5. Data Transfer from the mobile device to an edge router

to employ these MPLS tunnels to emulate the BS-BSC and BSC-MSC (mobile switching center) A3/A7 interfaces in the legacy wireless systems. Distribution of call flow control this way among various routers this way results in enormous cost savings for the customers due to effective utilization network links by reduction of the call control and data paths. Figures 5 and 6 depict the call flows after creation of MPLS tunnels.

In case of ingress data, the radio frames transmitted by the device are received by all active routers including a primary router and a number of secondary routers as shown in Fig. 5. The secondary routers simply forward the radio frames to the primary router. The SDU of the primary WR selects the best one among all such frames including the one directly received, inserts that into an IP packet, and transmits it to a back-haul network via an edge router for onward transmission to the other party.

The flow in case of egress data is naturally in the opposite direction as shown in Fig. 6. The core network hands over the IP packets to an edge router for onward transmission to the primary router through a pre-established or dynamic MPLS path. The primary WR segments the packets into radio frames and multicasts them to all the secondary WRs in the active set via dynamically configured Label Switch Paths (LSPs). The primary and secondary WRs then transmit each one of these received radio frames after different amounts of delay offset to the mobile device so that the

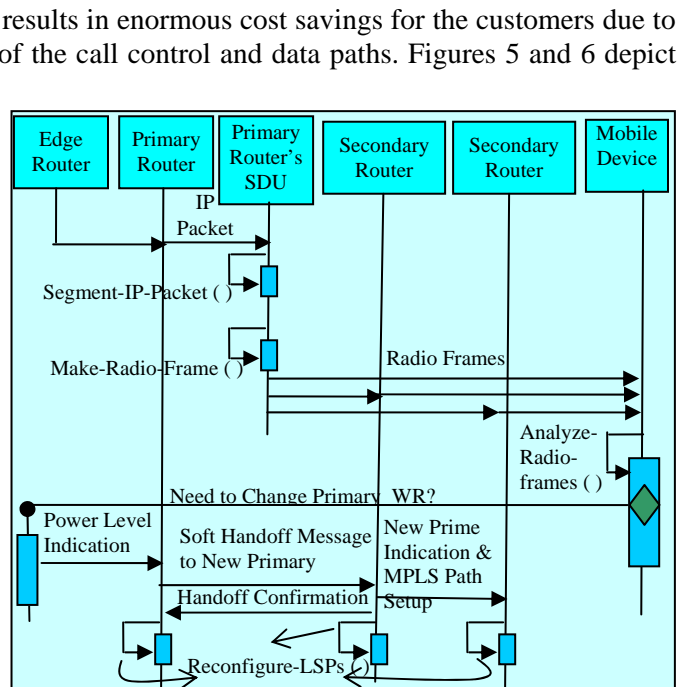


Fig. 6. Soft handoff and data transfer from an edge router to the mobile

primary WR segments the packets into radio frames and multicasts them to all the secondary WRs in the active set via dynamically configured Label Switch Paths (LSPs). The primary and secondary WRs then transmit each one of these received radio frames after different amounts of delay offset to the mobile device so that the

replicas of individual frames from different WRs arrive simultaneously at the destination. These synchronous radio frames received from different WRs are analyzed by the mobile device to not only obtain the best (correct) radio frame, but also assess the power levels of the frames. As the mobile device moves away from the primary WR, the power level of the radio frames from the mobile at the primary as well as that of the frames from the primary at the mobile drop. When the power level drops below a pre-configured threshold, the mobile device sends a control message to the primary WR indicating the power level of the prospective primary. The new primary WR could be one of the previous secondary wireless routers in an active set for the call. To achieve micro mobility, the current primary WR would signal the new primary an indication of the handover of its responsibility as primary WR. It would also supply to the latter the list of active WRs in the same control message. After receiving this message, the new primary WR receiving the strongest signals first confirms to the current primary that it is ready to take control of the traffic distribution, and then establishes multicast MPLS paths (LSPs) for the secondary routers in the active list. The LSPs provide synchronized framing for distribution and selection between neighbors of wireless traffic and fast rerouting for soft handoff using RSVP.

B. High QOS with Traffic Control for Effective Bandwidth Management

Overloading of a wireless network could occur because of the heavy data traffic. Traffic flow control for effective management of transmission resources, particularly the bandwidth, facilitates high quality of service (QOS). For example, by shaping (that is, spacing) the data traffic which usually comes in bursts unlike the voice traffic, the network bandwidth could be utilized more effectively. On the other hand, overloading of the wireless network even for short durations of time could result in degradation of QOS due to increased bit error rates.

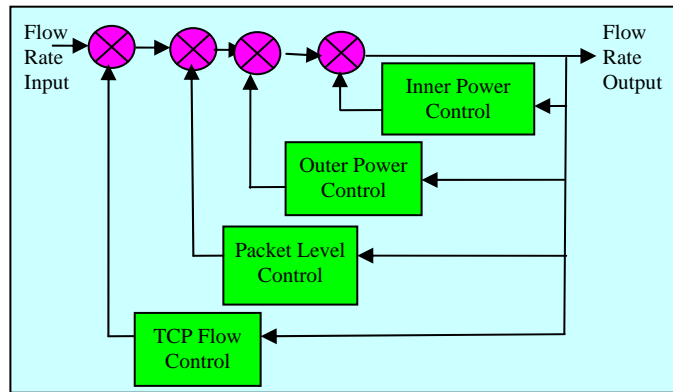


Fig. 7. Feedback mechanism for traffic flow control

Fig. 7 depicts the different feedback controls used in our system for traffic flow adjustment for effective data transmission. Signal power of the mobile device or any other direct indicator of power for the RF link constitutes the inner power control just as in the traditional wireless systems. This feedback is provided every 1-2 milliseconds. Similarly, outer power comprises a link error rate and/or interference indicator for the wireless link and may account for soft handoff power. This feedback may be provided every 50-100 milliseconds. The inner and outer power control loops conjointly provide feedback based on the signal strength of the RF link. The packet level control is provided every several hundred milliseconds

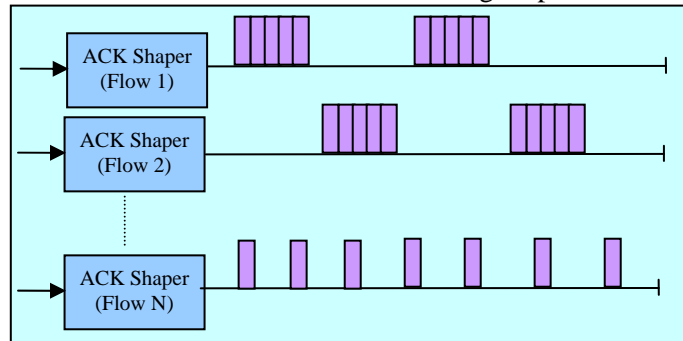


Fig. 8: Gang Flow Control for Sectors of a wireless network

by the queuing system in the WR based on the congestion status of the queues. Finally, our repertoire of traffic control mechanisms included the well known and well studied TCP flow control, which is provided through the acknowledgment messages between the two end points of the TCP flow. Based on the acknowledgment messages received, the source (WR) adjusts its transmission rate. Thus, with this mechanism, traffic flow is controlled by the congestion and/or interference state of the wireless links.

In our WR, there is also a provision for the traditional end-to-end rate control with a queue mechanism to shape up bursty traffic from a source into a smooth traffic flow of radio frames into the sink (mobile). The ACKs from the sink are also similarly queued up, and used as feedback for the source so that it can control its egress traffic flow.

An innovative flow control mechanism in the present work is Gang (or Group) flow control which seeks to shape the TCP flows from various sectors of the wireless network simultaneously, shown as Fig. 8 with N acknowledge shapers corresponding to N sectors of the WR. Each shaper accepts the packets from wireless network and stores the packets in the acknowledge queues inside the WR. The acknowledge shaper can transmit acknowledge message over time to change the traffic flow for the sector based on the flow's power indicator of the RF link from wireless network.

Fig. 8 illustrates two kinds of shaped acknowledge messages. In the first type shown for flows 1 and 2, the acknowledge messages are arranged in small groups and the groups are dispatched periodically. In the second type depicted for Flow N , on the other hand, the ACKS are evenly distributed over time and transmitted. The difference of these two arrangements is that they have different transmit time. The first type of ACK shaping will affect the offsetting bursts for the traffic flow whereas the second type will affect the steady flow rate for the traffic flow.

The flows with unused bandwidth or lack of bandwidth will be identified for each TCP group for the interval related to retransmission time out for the TCP flow. The fair share of each TCP flow within the group will be calculated and used in adjusting the speed of acknowledge messages for the flows inside the gang. The fair share of each TCP flow may be used together with the RTT and arrival time for each traffic flow.

C. Mechanisms to Dynamically Configure the Router

The proposed WR has been designed for dynamic configuration of its operational parameters. Configuration parameters of a WR are typically related to the site and technology used. Site parameters may be classified as geo-location, network operation, service configuration, and antenna parameters. The technology specific parameters depend upon whether the CDMA or GSM is supported and include technology-specific site parameters, but may be broadly classified into coverage, spectrum, channel, interference, control, and threshold parameters. At a finer level of detail, site Id, number of sectors/beams, sector/beam ID, latitude and longitude, sector/beam location, maximum radius of influence are typical geo parameters. Similarly, network configuration parameters include network interfaces (*e.g.* T1, SONET, T3, *etc.*), site capacity, and network capacity. In the service configuration, we may have the list of various services supported, and the related directory agent (DA) addresses. The antenna parameters listed on a per sector/beam basis include the antenna type, digitized pattern, horizontal/vertical beam widths, max gain, and mechanical and electrical down tilts.

In the technology parameters class, maximum RTD (round trip delay), PER (packet error rate), FER (frame error rate), and percentages of blocked calls, access failures, dropped calls constitute the threshold parameters subclass. The coverage parameters subclass includes environment (*e.g.* rural or urban), path loss margin, technology specific hardware losses and gains, RF coverage prediction models, and traffic distribution maps. The spectrum parameters subclass consists of channel bandwidth, channel mask, channel number range, and maximum transmit power per channel. The channel parameters include the number of channels in the range, air capacity/bandwidth, minimum channel spacing, frequency use, frequency grouping, and hopping sequences. The interference parameters include interference thresholds, power control thresholds, channelization and sequencing, channel scheduling algorithms, RF interference prediction models, traffic distribution maps, and adjacent channel interference threshold. The control parameters subclass includes access parameters, intra-technology and inter-technology handoff parameters, and timing parameters.

In the following subsection, we present a very high level procedure by which a WR learns its parameters and configures itself in collaboration with its neighbors. The control logic and the operations

performed in each high level state could be quite complex with several states for error paths and exceptions. For example, application of RF or IP discovery protocols in the startup state involves considerable information exchange between a WR and its neighbors. Hence, we present only an overview of our automatic WR configuration state machine here below, and refer to our patent [10] for finer details.

D. Automatic Router Configuration

Automatic configuration of the WR is performed using a 4-state machine including a start up state, a learning state, an operational state and a site down state as depicted in Fig. 9.

The WR configures the RF/IP topology in the startup state, and refines the topology in the learning state. In the operational state, the wireless router handles a full traffic load, and continues to check if it meets the operational thresholds. Scheduled or unscheduled maintenance leads the WR into site-down state.

1) *Start up state*: In the starting state, connectivity is first established between the WR and the wire-line routers in the network. The wire-line connectivity is then used to establish connectivity between WR and its WR neighbors. Subsequently, MPLS paths or other suitable virtual circuits or IP tunnels are established among the WR neighbors to facilitate inter-router communication. The WR uses wire-line connectivity to learn from its neighbors the RF topology in the neighborhood, and establish wireless-specific connectivity with its neighbors.

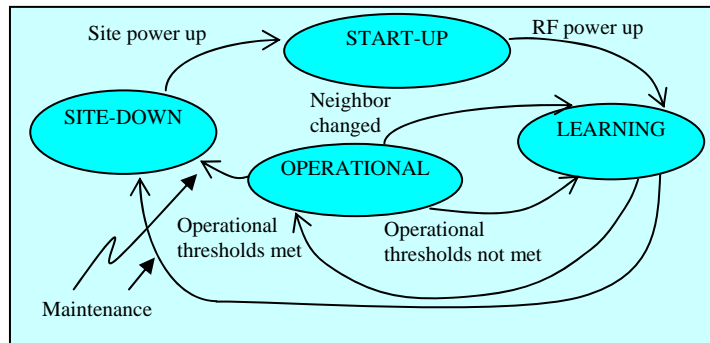


Fig. 9: A state machine for dynamic configuration of a wireless router.

After establishing the wireless connectivity with its WR neighbors, a WR exchanges RF impact information, including some or all of the parameters mentioned earlier. By exchanging this information, and negotiating the various operation parameters, the WR is able to determine or estimate a set of operating parameters that will help maximize radio coverage, minimize interference, and aid in providing a seamless coverage from cell to cell with smooth handovers. If no coordination could be achieved between the wireless routers in a neighborhood, Operation, Administration and Maintenance (OAM) server is contacted for resolving the differences. The OAM server then performs the RF impact analysis and responds with the operational parameters for the new site and the neighboring sites. The OAM server may also re-identify neighbor sites after parameters are agreed to, and store them in the configuration and parameter tables in the routers. The routers transition to the Learning state as shown the figure. Now, a RF system and network has been established by activating the wireless routers.

In the start up state, chores of the WR are: i) identification of the neighbors of the WR and preparation neighbor list, ii) interference impact, coverage, and other parameter analysis, iii) configuration of LSPs with neighbors, and iv) exchange and negotiation of power and handoff parameters with neighbors. In this way, the WR automatically configures itself for operation in the wireless network. Once these operations are completed, it transitions to the learning state.

2) *Learning State*: In the learning state, the WR continues to analyze, exchange and negotiate parameters, in order to minimize interference in the wireless network and to ensure that all the operational thresholds are met. The WR transitions into this state from the start-up state when RF power is up, and from operation state when either operational parameters change, or the WR neighbors change, or the operational thresholds are not met. In the learning state, parameters are re-negotiated and re-estimated based on the information given during transition from the operational state. Once operational thresholds are met for a specific period of time, the WR transitions to the operational state.

3) *Operational State:* In the operational state, the WR continues to monitor its operational thresholds periodically or otherwise exchanges information with its WR neighbor to ensure maximum efficiency and minimum RF interference within the wireless network. If the operational thresholds are not met, the wireless router transitions from the operational state back to the learning state for detailed analysis and evaluation of the configuration parameters and reconfiguration, as required, so that the operational thresholds can be met.

Also, if any of the neighboring routers change, affecting the topology of the network, such as a neighboring router failure, or a new router is added to the wireless topology, the router transitions from the operational state to the learning state, to reconfigure itself to suit to the new topology. In addition, if any parameters are changed due to any requests from its neighbors, the WR transitions to the learning state, for analysis and evaluation of operation using the new parameters.

4) *Site Down state:* The wireless router may enter the site down state from the learning state or the operational state if it requires either scheduled or unscheduled maintenance. Upon power up, the router will again transition back to the start up state for reloading and reconfiguration of the operational parameters. In this way, the wireless routers automatically adjust and account for changing conditions in the network to optimize operation of the network.

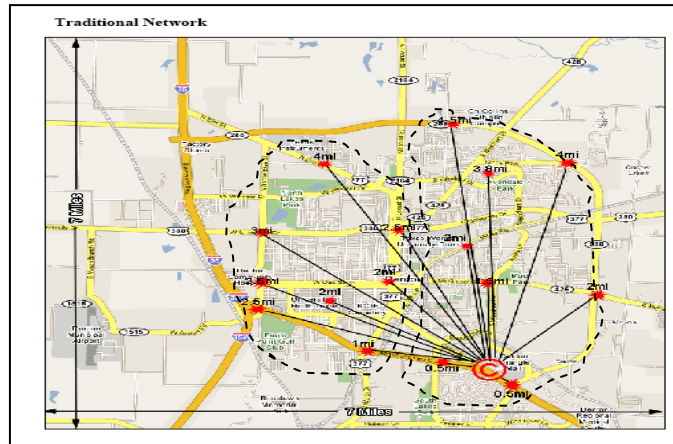


Fig. 10. Typical deployment of base-stations in the Denton area

E. RESULTS

Based on our analysis on a possible field deployment, we present in the following sections the results on cost and system availability improvements.

i) Improvement on Annual Costs

For this analysis, we use a typical configuration of traditional wireless network in the Denton area of Dallas metropolis, Texas, USA, with 16 base stations deployed at the “*” marks as shown in Fig. 10. The big encircled “C” in the picture represents a central office (CO). We consider three wireless network architectures as shown in Fig. 11 with identical deployment for our comparative analysis: i) a traditional network (Fig. 11 (a)) with the CO in Denton connected to BSC in Dallas via two more Cos; MSC is assumed to be

co-located with BSC, ii) a star network (Fig. 11 (b)) in which the base-stations are partitioned into two groups as shown hatched line contours and connected to one CO cum WR each, and iii) a

distributed WR mesh network (Fig. 11(c)) with WRs replacing base stations and the same two groups of WRs as in the star network each connected to an edge router (ER). We assumed traditional T1 and T3 leased lines for communication between the COs, CO (CO cum WR) and BSC or BS, and WR and ER. In configuration (iii), WR-WR links are assumed to be fixed cost internet *e.g.* DSL (Digital Subscriber Loop) links. **Table 1** depicts the cost structure used in our analysis. In order to make a fair comparison, we presumed that in configurations (b) and (c), there is another CO/WR (ER, in (c)) at a distance 35 miles off (closer to BSC in (a)) for communication with users at that end, even though this is

Table 1: Leased Monthly Class Structure Used In Our Analysis

Link	Channel Termination Cost	Inter-Office (Fixed Cost)	Cost/Mile
T1	150\$	50\$	15\$
T3	1500\$	500\$	150\$

not depicted in the figure in order to keep it simple and clear.

Even though the inter-office connections depicted by IOL (Inter Office Links) in 11(a) and the links depicted by double arrows in 11(b) and 11(c) seem to have an identical functionality, there is a classic difference between them. The links represented by double arrows carry IP traffic and hence they take advantage of statistical multiplexing. Assuming this multiplexing factor to be 3, these links effectively carry only one-third of the traffic ensuing from the BSs. Further, since the handoff processing is performed at the CO/WR (or WR) level thanks to the distributed SDUs therein, the redundant traffic resulting from the need for simultaneous monitoring of the same call via multiple BSs for facilitating handoffs particularly at the cell boundaries, will not be present on the double arrow links. Assuming that the handoff factor is 1.75 (*i.e.* handoff processing results in 75% increase in the BS traffic), we will get a further reduction by a factor of 1.75 for the traffic on the double arrow links with a total reduction by a factor of 5.25. This assumption is consistent with the available statistical information that only 15% of the BTS-BSC traffic is pay load. Looking from a different angle, we need lesser number of T1/T3 (double arrow) links to handle a fixed BS traffic.

In our cost analysis, we used the maximum number of simultaneous subscribers at a BS at any time (shortly denoted by N_s as an independent variable and carried out for this variable varying in multiples 100, the cost estimates based on the T3/T1 links required to bear the traffic on different segments of the three network configurations under study. We performed our analysis for both 2G traffic with exclusively voice traffic, and 3G traffic with 70% voice traffic and 30% data traffic. Since bandwidth for a voice call is 64 Kbps, and that for a data call is on the average 700 kbps, it is straightforward to verify that the bandwidth requirements in mbps for N_s subscribers of a BS for these two cases is given by the two following two formula:

$$BW_{2G} = N_s * 0.064 \quad \dots(1)$$

$$BW_{3G} = N_s * (0.7 * 0.064 + 0.3 * 0.700) \quad \dots(2)$$

We estimated the number of T3s required by considering the multiples of 45 mbps in the estimated bandwidth. We tried to cover the residual bandwidth by T1s if it is less than 10 mbps. Otherwise, we used an additional T3 because, from Table 1, it is apparent that T3 is more cost effective if the number of T1s exceeds 9. As discussed before, the bandwidth estimates above have been scaled with handoff and statistical multiplexing factors as required in the relevant segments of the network. Now, the cost C_T for the traditional network, as per the ILEC (Incumbent Local Exchange Carrier) rate basis, is given by:

$$C_T = N_{BS} \left[\sum_{Link=T1,T3} N_{Link} (2.T_{link} + F_{Link} + M_{Link} D) \right] \quad \dots(3)$$

Here N_{BS} is the number of BSs, N_{Link} is the number of links of link type “link” = T1 or T3, and T_{Link} , F_{Link} and M_{Link} , are the termination, fixed, and per mile costs, respectively, associated with the link. D is the average total distance from a BS to BSC (or farthest CO/WR or ER in the other two non-

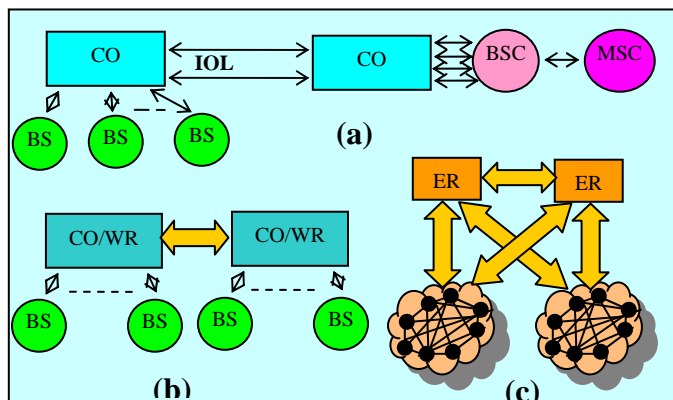


Fig.11: Three network configurations with identical BS/WR deployment used in our analysis: a) Traditional network, b) Star network, and c) Distributed WR mesh network.

traditional configurations) and is assumed to be 35 miles. For computing C_S , the cost per star network, we need to consider the fact that, out of and N_{Link} s emanating from the BSs, only N_{Link}^* s are required to carry the traffic reduced by a factor 5.25 along the double arrow route as discussed above. Hence, C_S is given by:

$$C_S = N_{BS} [\sum_{Link=T1,T3} \{ N_{Link} (T_{link} + F_{Link}) + N_{Link}^* M_{Link} D + (N_{Link} - N_{Link}^*) M_{Link} d_{avg} \}] \quad \dots(4)$$

Finally, C_M , the mesh network cost, may be estimated as the sum of the costs of the WR-WR links, WR to local ER links, WR to farther ER links (for better reliability), and ER-ER links. Upon aggregation, the total cost is given by:

$$C_M = 0.25 N_{BS} (N_{BS} - 2) C_{Fixed-Link-Cost} + N_{BS} (\sum_{Link=T1,T3} N_{Link}^* (T_{link} + F_{Link} + M_{Link} D)) \quad \dots(5)$$

In the above calculations, the N_{BS} WRs (in place of BSs) are assumed to be partitioned into two equal sized groups with total connectivity in each group. The first term on the right hand side represents the cost of the WR-WR links with a fixed cost $C_{Fixed-Link-Cost}$ (of 50\$) for each link. The other three types of links across the double arrows can be treated as continuous pipes of total length D from WR to the farthest ER in downtown Dallas. The second term in the above equation represents the cost for these $N_{BS} N_{Link}^*$ IP-traffic bearer links.

Fig. 12 (a) and (b) are the graphs showing variations of the total annual costs with the number of subscribers (N_s) in the three network configurations for the 2G and 3G scenarios, respectively. Obviously, the star and mesh configurations with the proposed router outperform the traditional network configuration. The difference is much more pronounced in the 3G scenario compared the 2G one. In both the situations, the mesh is more cost effective compared to the star obviously because of handoff management at WR/BS level and inter-router IP communication.

B. Improvement on System Reliability

It is known that the availability of an individual system entity or link is given by $MTTF/(MTTF+MTBF)$ where $MTTF$ and $MTBF$ are the mean time to failure and mean time between failures, respectively. In this analysis, we presume the individual availability values and compute the overall system availability by aggregating these availabilities. It is obvious that the traditional network topology is the least reliable of all due to its susceptibility to single point failures along the long dedicated link paths from BS to MSC. Assuming there are M legacy entities

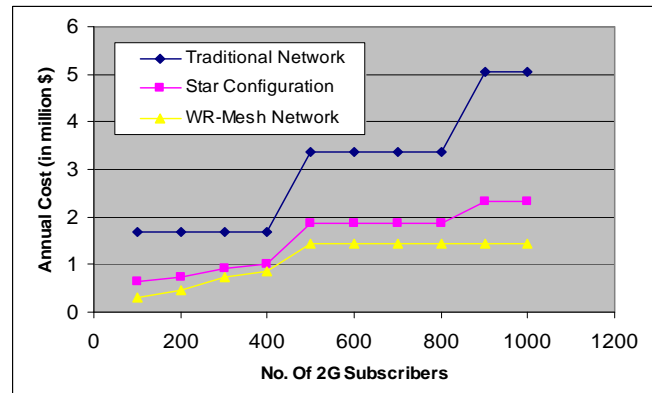


Fig. 12 (a) Annual cost variations of the three networks with the size of the 2G subscribers that could access a base station simultaneously.

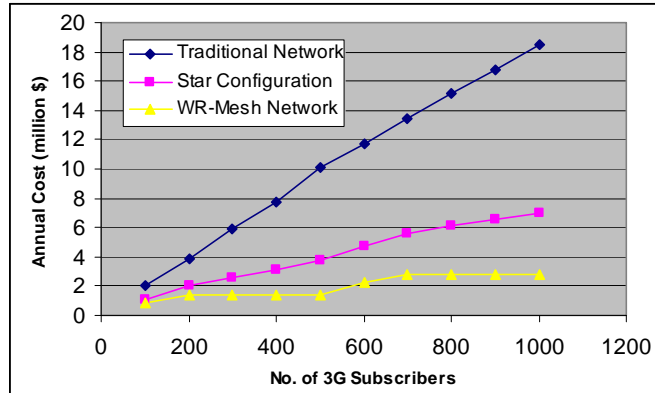


Fig. 12(b) Annual cost variations of the three networks with the size of the 3G subscribers that could access a base station simultaneously.

(e.g. BSC, CO, MSC) on this path, each of availability A_E , interconnected by links each of availability A_L , the overall availability of the system can be estimated, based on the serial connectivity, to be $(A_E A_L)^M$. In the star topology, the dedicated path terminates at CO/WR. Hence, the availability of this topology is given by $A_E A_L$. For

availability analysis of our mesh network, we need to consider three kinds of failures: (i) failures of individual WRs, and WR-WR links, (ii) WR-ER link failures, and (iii) ER failures. Multiple parallel paths in WR-mesh, and from WR-meshes to ERs make the effect failures of type (i) and (ii) on the overall availability negligible. Hence, we need to consider only the dominant ER failures. If A_{ER} is the availability of an ER, and each mesh has multiple parallel links to N ERs, the overall availability of the mesh configuration may be approximated as $(1 - (1 - A_{ER})^N)$. With $A_E = A_L = A_{ER} = 0.99$, $M = 3$, Table II gives overall availability of the topologies

Table 2: Availability of Different Wireless Network Topologies

considered above. The 4 to 6 9s reliabilities (availabilities) of the WR-mesh networks as indicated in **Table 2** satisfy the high availability requirement of next generation RANs.

Traditional	Star	Mesh (N=2)	Mesh (N=3)
0.94	0.98	0.9999	0.999999

Future Research Directions

In addition to the few network topologies explored in this section, a number of generic and application-specific topologies need to be explored in the future and evaluated based on the cost effectiveness, fault-tolerance, and performance. With the emergence of voice-over-IP (VOIP) technologies for multimedia communications, application of the IP wireless router technologies for development of next generation 911-service infrastructure becomes an important future direction for the research proposed herein. Analytical and simulation studies on quality of service and bandwidth efficiency tradeoffs with various wireless router configurations are an interesting direction for future research.

CONCLUSION

We present here a wireless IP-RAN architecture based on innovative wireless routers that provide message routing in addition to control functions of legacy system entities such as BSCs and MSCs. The tremendous savings these routers offer on backhaul network costs without sacrificing the QOS make the highly available and scalable IP-RANs based on a distributed configuration of these routers indispensable for next generation wireless networks.

REFERENCES

3GPP2 C-S0033, Rev 0, Ver 2.0 (2003) Recommended Minimum Performance Standards for cdma2000 High Rate Packet Data Access Terminal. *3rd Generation Partnership Project* (also categorized as TIA-866-1 standard).

Bu, T., Chan, M. C., & Ramjee R. (2006) Connectivity, performance, and resiliency of IP-based CDMA radio access networks. *IEEE Trans. Mobile Computing*, 5 (8), 1103-1118.

Chen, T., & Hamalainen, S. (2003) Handover in IP RAN. *Proc. IEEE International Conf. Communication Technology*, 2, 812-815.

Dantu, R., Patel, P. R., Choksi, O. T., Patel, A. R., Ali, M. R., Miernik, J., & Holur, B. S. (2006) Wireless Router and method for processing traffic in wireless communication network. *US patent # 7068624 B1*.

- Dantu, R. (2005) Method and System of Integrated Rate Flow Control for a Traffic Flow across Wireline and Wireless Networks. *US Patent # 6904286 B1*.
- Ekström, H., Furuskär, A., Karlsson, J., Meyer, M., Parkvall, S., Torsner, J., & Wahlqvist, M. (2006) Technical Solutions for the 3G Long-Term Evolution. *IEEE Commun. Mag.*, 44 (3), 38–45.
- Ergen, M. (2009) Mobile Broadband - Including WiMAX and LTE. *Springer, NY*.
- Fazel, K., & Kaiser, S. (2008) Multi-Carrier and Spread Spectrum Systems: From OFDM and MC-CDMA to LTE and WiMAX. *2nd Edition, John Wiley & Sons*.
- Ghosh, S., Basu, K., & Das, S. K. (2005) What a Mesh! An Architecture for Next Generation Radio Access Networks. *IEEE Network*, Sept./Oct. 2005, pp. 35-42.
- Kempf, J., & Yegani, P. (2002) OpenRAN: a new architecture for mobile wireless Internet radio access networks. *IEEE Communications Magazine*, 40 (5) 118-123.
- Patel, P. R., Choksi, O. T., Davidson, K. W., & Dantu, R. (2006) Method and System for Configuring Wireless Routers. *US Patent# 7031266 B1*.
- TIA/EIA/IS-2000.1.A -2, “The CDMA2000 Family of Standards for Spread Spectrum Systems,” Feb. 11, 2002.
- Vassiliou, V., Owen, H. L., Barlow, D. A., Grimminger, J., Huth, H. -P., & Sokol, J. (2002). A radio access network for next generation wireless networks based on multi-protocol label switching and hierarchical Mobile IP. *Proc. 56th IEEE Vehicular Technology Conf.*, 2, 782-786.
- Yasukawa, S., Nishikido, J., & Hisashi, K. (2001) Scalable mobility and QoS support mechanism for IPv6-based real-time wireless Internet traffic. *Proc. IEEE GLOBECOM Conf.*, 6, 3459 – 3462.