Base Station Controlled Load Balancing in Mobile WiMAX

Thomas Casey

Supervisor: Professor Riku Jäntti
Instructor: Nenad Veselinovic Dc.S.

Thesis conducted for Elektrobit Corporation in
co-operation with HUT Communications Laboratory

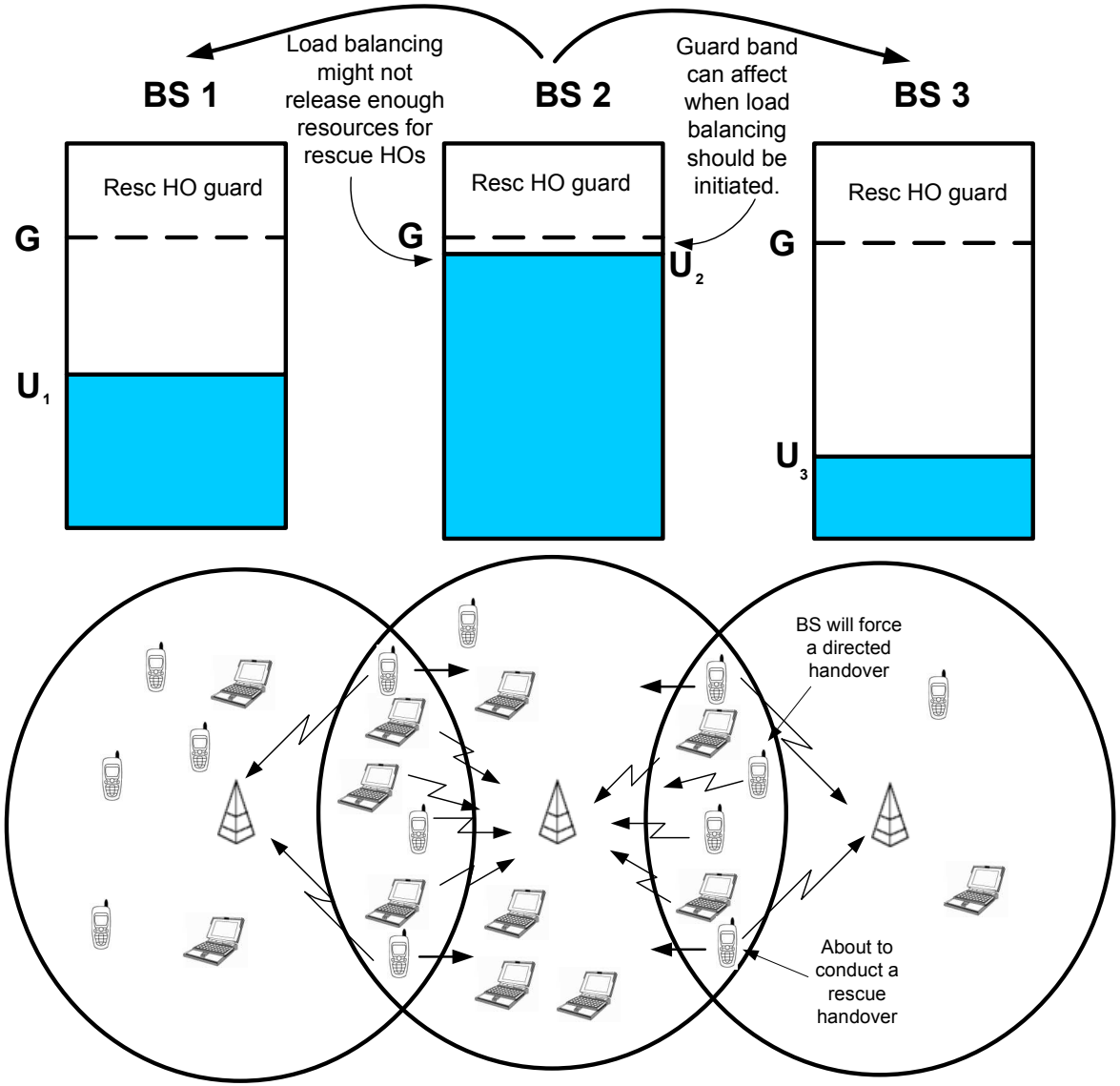
Outline

- Objectives
- IEEE 802.16e and Mobile WiMAX framework
- Load Balancing and handover prioritization
- Load Balancing scheme design and enhancements
- Evaluation
- Conclusions and Future Work

Objectives of the thesis

- To examine how Load Balancing with Base Station initiated directed handovers could be conducted in Mobile WiMAX and its potential
  - To enhance Resource Utilization
  - To improve QoS system wide in a Mobile WiMAX access network
- Goal also to conduct preliminary research on
  - How resources could be reserved for the admittance of rescue handovers (and higher priority traffic)
  - How this would affect Load Balancing
  - How these two approaches could be combined

1. Load Balancing based BS initiated **”directed handovers”**
   - Triggered when a threshold in Resource Utilization or Reservation passed
2. Signal quality based MS initiated **”rescue handovers”**
   - A guard band can be reserved in the Target BS to minimize connection drops
3. Different kinds of terminals and traffic



IEEE 802.16e radio interface

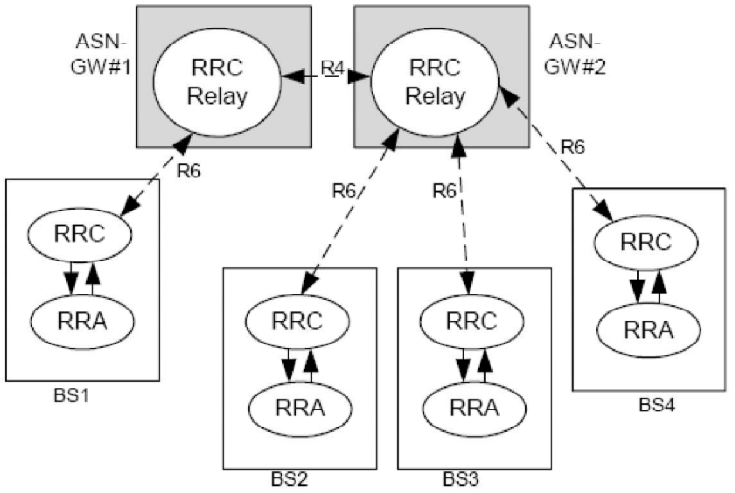
- Strong QoS framework enabled by agile MAC scheduling
  - non-BE services
    - UGS and ertPS for delay and jitter sensitive connections (e.g. VoIP)
    - rtPS for more jitter tolerant real-time connections (e.g. streaming video)
    - nrtPS for delay tolerant and elastic TCP based connections (e.g. HTTP and FTP)
  - BE service (any traffic)

- Wide support for different types of handovers (from an uncontrolled handover to a network controlled and optimized handover)
  - Simple hard handovers (also possibility for optimization)
  - Pre-association to the target BS to speed up ranging
  - FBSS (seamless handover), MDHO (soft handover)
    - Targeted especially for delay sensitive and high mobility connections

- **How could the QoS framework be taken into consideration and the variety of handovers be utilized in Load Balancing?**

WiMAX Forum Access Service Network (ASN)

- WiMAX Forum defines an access network based on the IEEE 802.16e technology
- Load Balancing logic in an ASN is distributed to the Base Stations
  - Radio Resource Agent (RRA) responsible for local Radio Resource Management and triggers Load Balancing if necessary



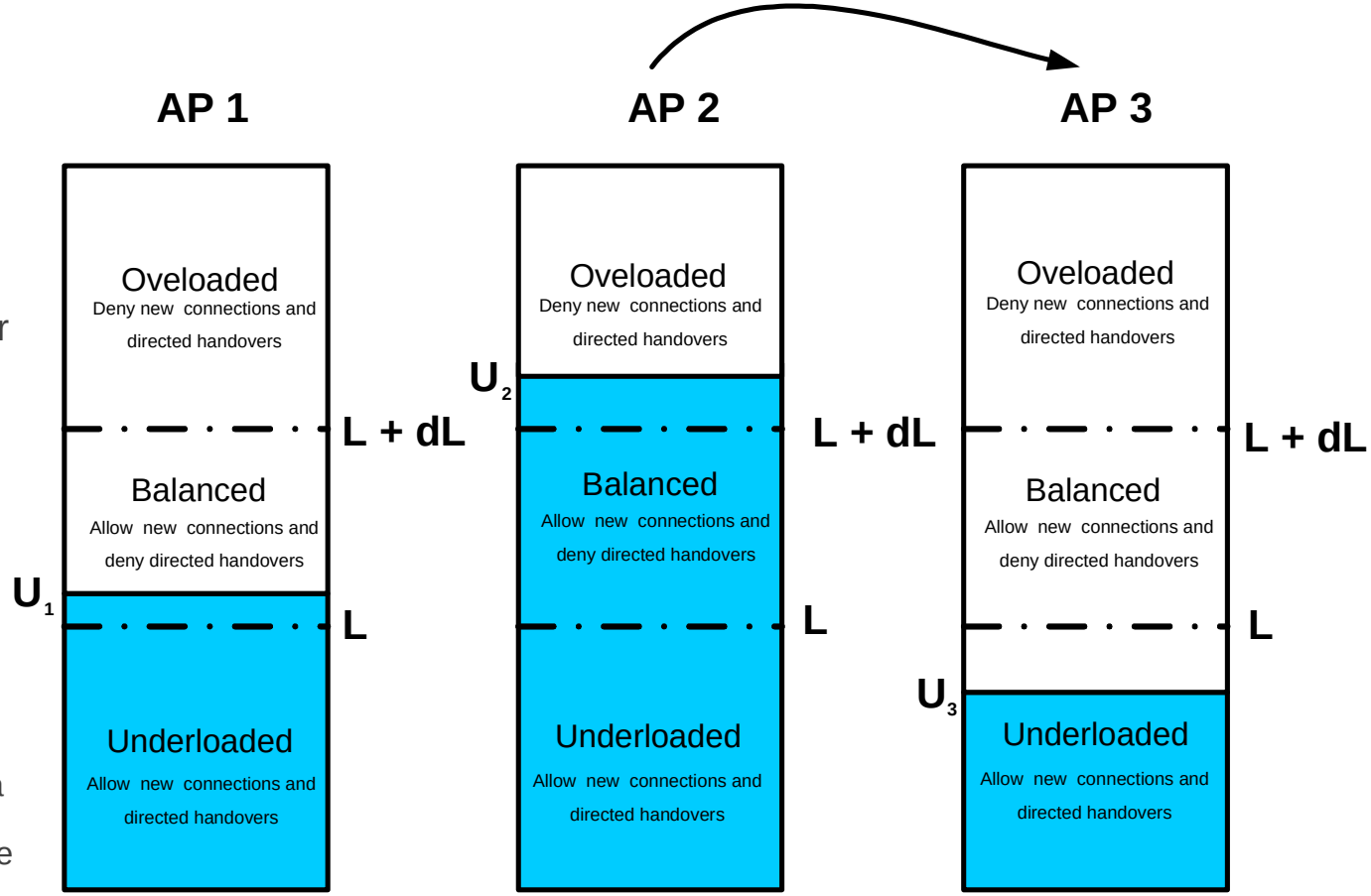
- Spare capacity report (SCR) used to communicate the resource usage status between BSs
  - The Radio Resource Controller (RRC) takes care of the communication between and across RRAs
  - Available resources communicated only for non-BE connections !
  - BE terminals -> assume MS initiated Load Balancing
- Admission control
  - Protects existing connections
  - Can be used to reserve the guard band for rescue handovers and higher priority traffic

Basis for Load Balancing

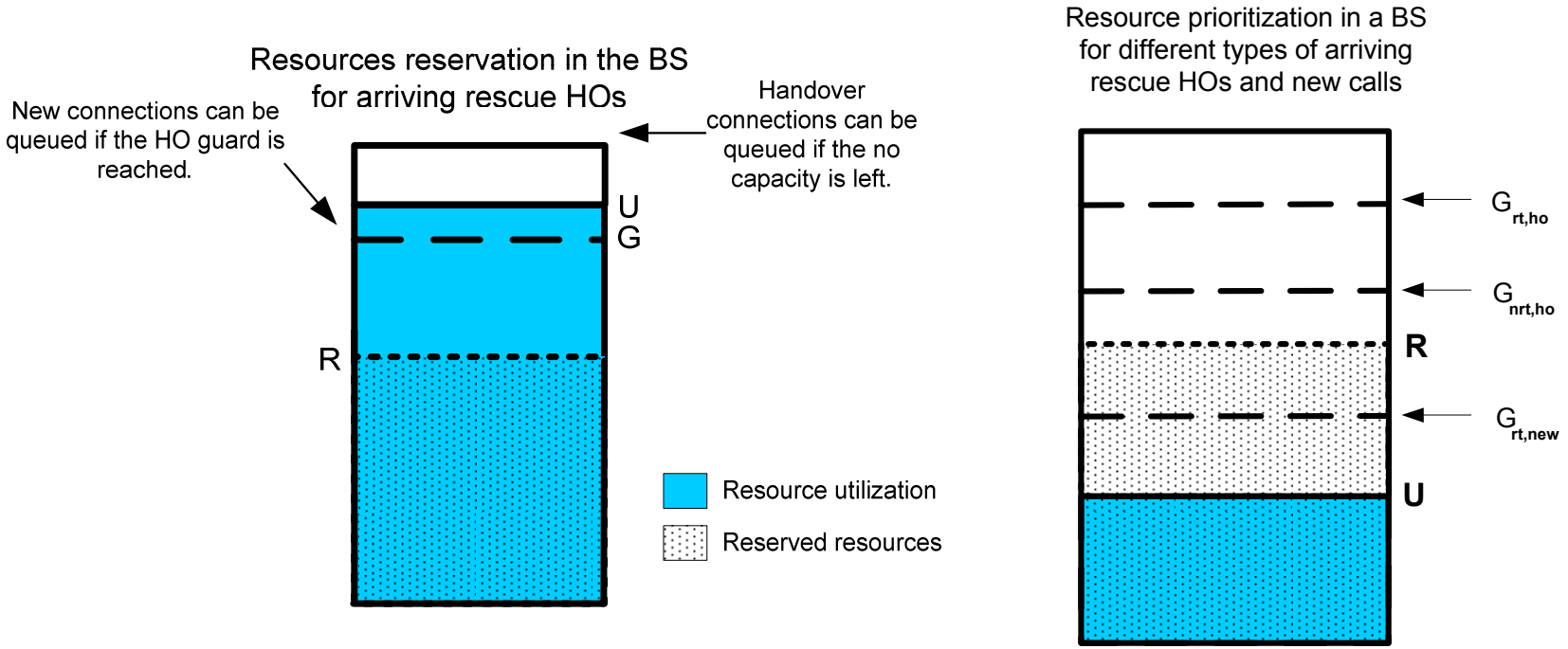
- Load Balancing methods used in previous telecommunications systems
  - Resource Reservation
    - Bring the resources (unoccupied frequencies) to where most of the users are located
    - More popular in traditional cellular networks, no framework exists in Mobile WiMAX
  - Load distribution
    - Direct traffic to where resources are
      - Directed handover
      - If connection blocked -> Network Directed retry (or roaming)
    - Used in Mobile WiMAX for Load Balancing
- Slot (base time-frequency unit) used as the main Load Balancing metric
  - Describes well both traffic load and channel capacity (i.e. used MCS)
  - Delay and packet drops can be used as supporting variables

A basis for development from previous research

- Not much research conducted on Resource Utilization based BS initiated load distribution schemes
- Most suitable scheme for Mobile WiMAX from [Vel04]
  - Simple scheme
  - LB triggered in overloaded state
  - Directed HOs conducted only to underloaded peers
  - Avoids the ping-pong effect caused by fluctuating Resource Utilization **U** by using a hystersis margin **δ** in relations to the average Resource Utilization **L**.



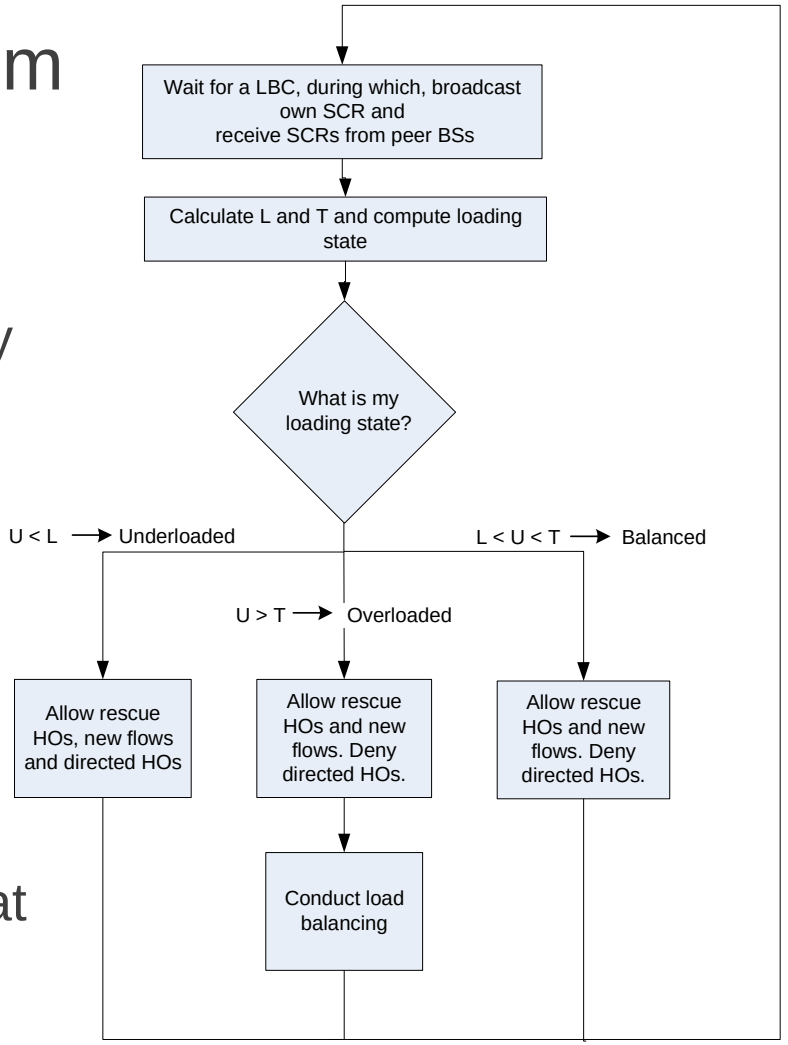
Handover and traffic prioritization



- Guard bands set to protect handovers or higher priority traffic
  - E.g. after passing G_rt,new, new non-real-time calls are blocked and new real-time connections (and all handover traffic) still admitted
  - Can be fixed or adjusted dynamically according to the loading situation
- Resource Utilization U can be different from Resource Reservation if fluctuating traffic with idle periods (e.g. HTTP)
  - There should be a possibility to trigger Load Balancing also on Resource Reservation level before the guard bands are passed.

Basic Load Balancing algorithm

- Based on [Vel04]
  - Some modifications
- Hysteresis value fixed and set manually
- Distinguishes between
  - Rescue HOs
  - Directed HOs
  - Rescue HOs always admitted
- Complemented by admission control
  - A new flow can be accepted also in an overloaded TBS
- LB conducted only for (“static”) MSs that are likely to reside in the overlapping areas for their whole session



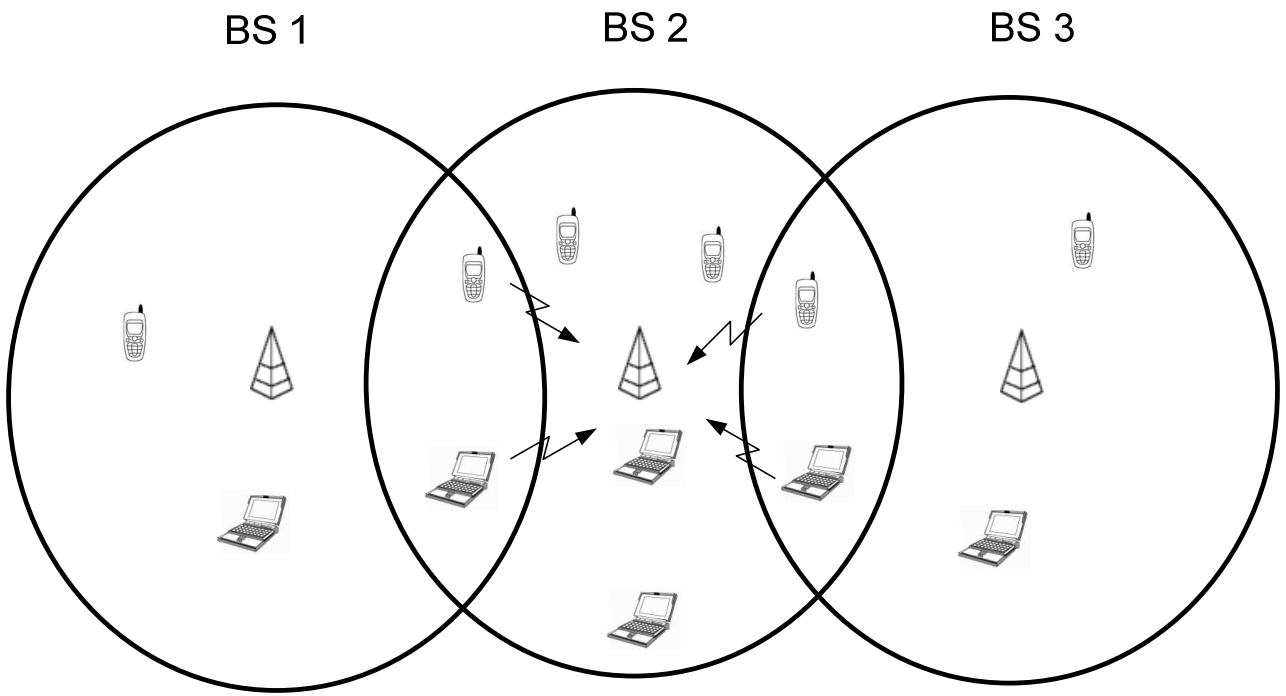
When Load Balancing triggered

- BS should decide
  - Which MSs?
    - Which MSs feasible -> in overlapping areas
    - Likely to stay in the overlapping area
  - In what order?
    - MSs with higher priority and worse channel conditions first
  - How many at a time?
    - Possible to conduct handovers in parallel to speed up Load Balancing
- Challenges
  - Hard to set one fixed margin that is good for all traffic profiles
    - A need for a scheme that dynamically adjusts the threshold
  - How to resolve static MSs in overlapping areas
    - Requires heavy scanning -> especially bad for e.g. VoIP connections

Enhancements

- Enhancement proposals
  - Automatic tuning of the triggering threshold
  - BS controlled Load Balancing for BE traffic
  - Multiple threshold triggering
    - In a fluctuating environment
  - Resource reservation triggering
    - Triggering Load balancing in relations to the guard band

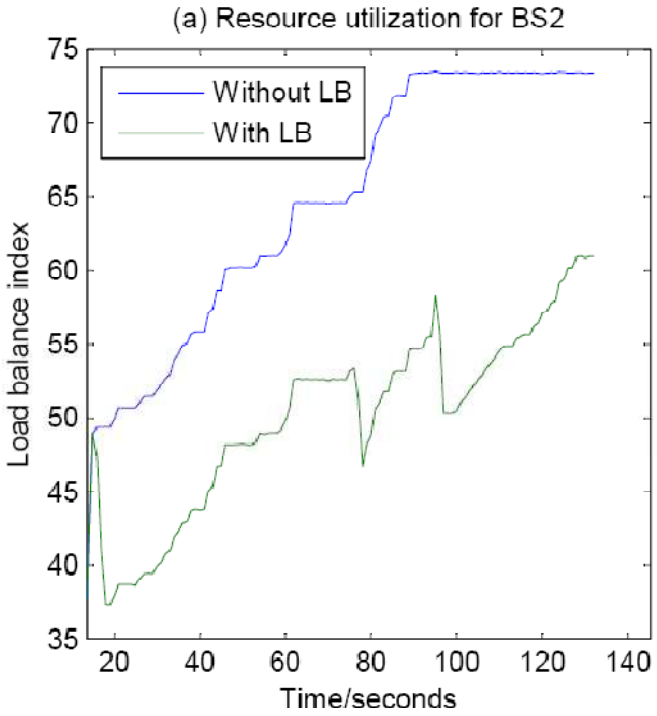
Evaluation: configuration

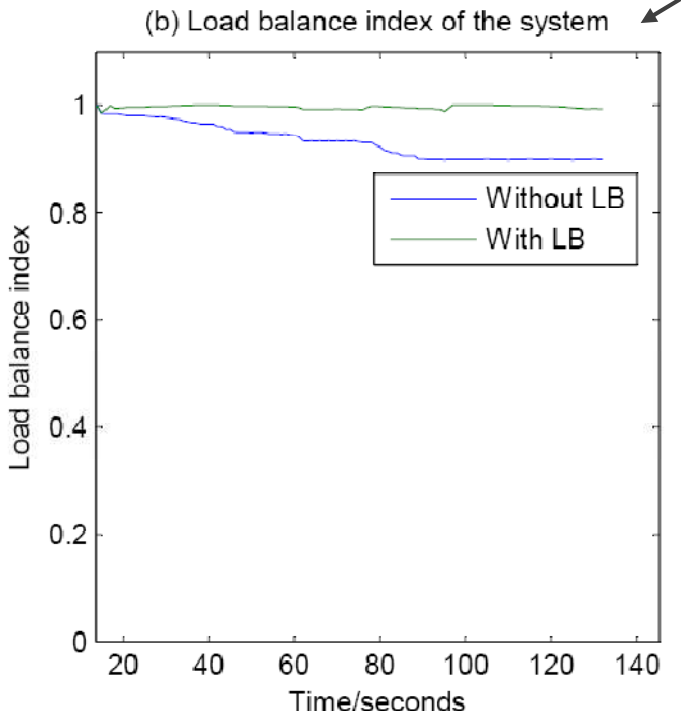


- Environment
  - Static terminals used (fixed MCSs)
  - Traffic
    - Static VoIP based non-BE traffic (UGS and ertPS service used)
    - FTP and HTTP based BE traffic
- Overload BS 2 -> use basic Load Balancing algorithm (enhancements not evaluated)
  - Case 1: without LB vs with LB
  - Case 2: Hysteresis margin optimization
  - Case 3: Load Balancing Cycle length optimization

With and without LB

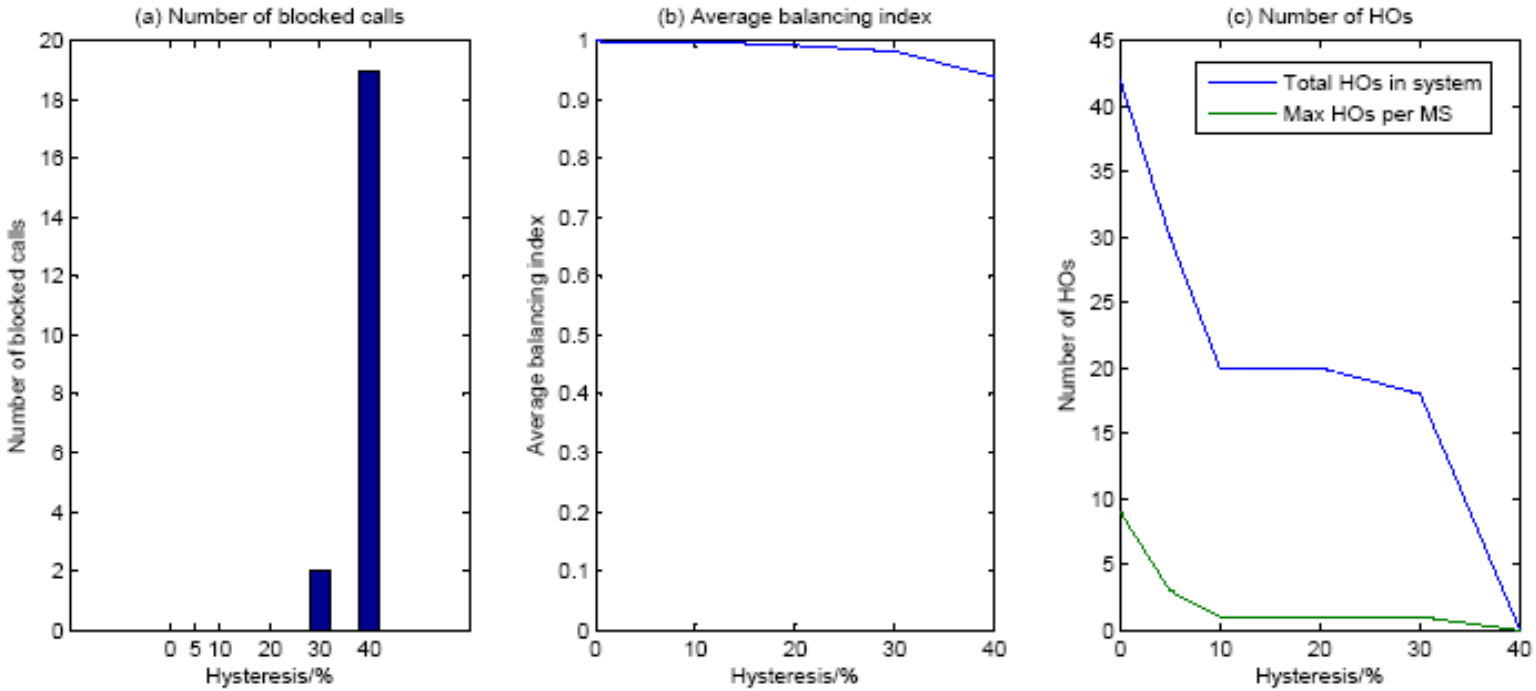
The closer to 1 the more balanced the system is!



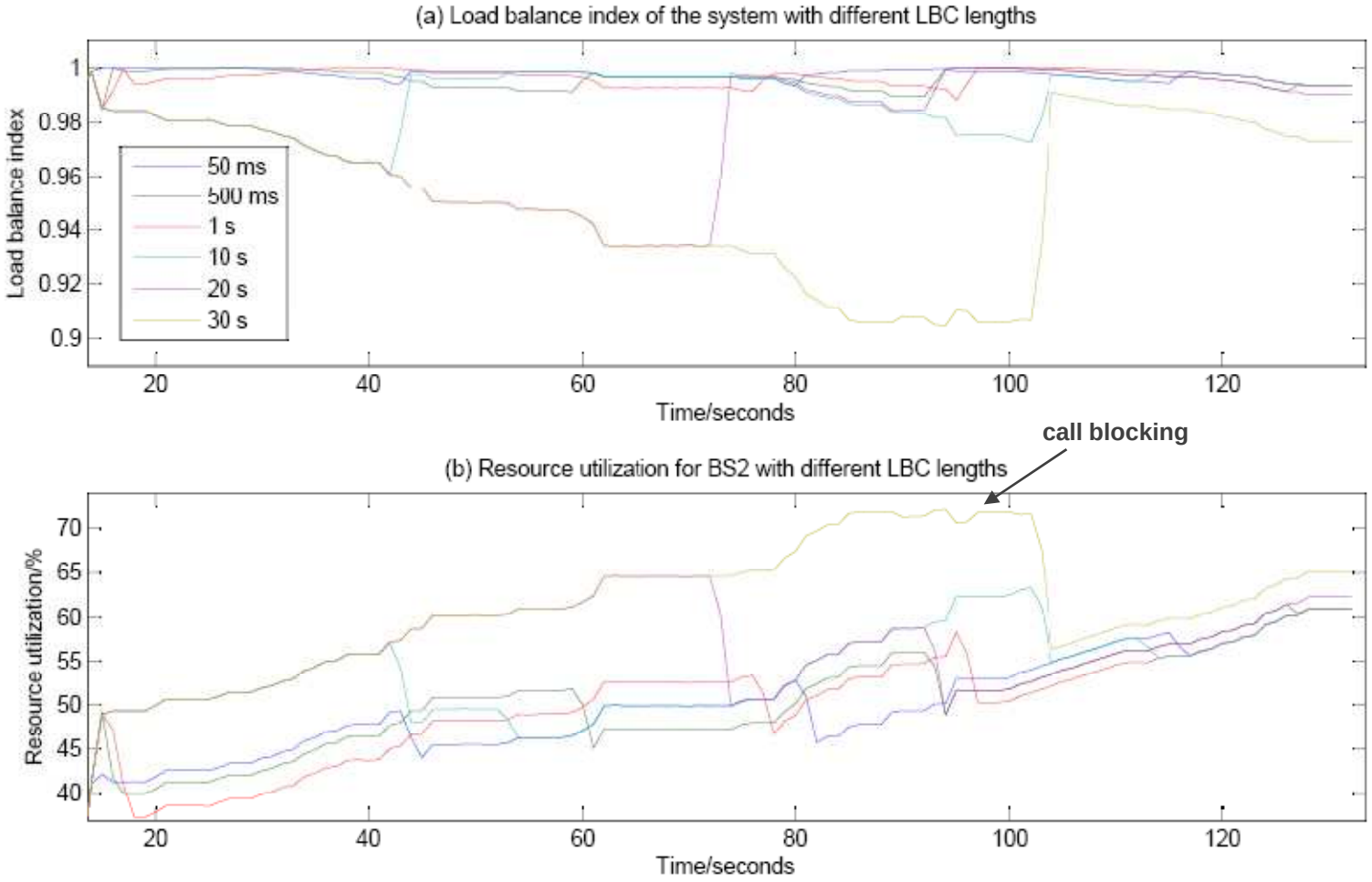


- When LB not used
  - BS 2 became overly loaded
  - 19 VoIP calls had to be blocked
  - Considerable decrease in BE throughput (UL acks did not go through)
- When LB was used
  - Load distributed across the system well
  - Call blocking in BS 2 avoided altogether
  - BE throughput improved

Hysteresis and LBC optimization



- Hysteresis value 20 % was concluded as the best value to avoid both
  - Call blocking and
  - The ping-pong effect



- An LBC value even up to 20 seconds was sufficient to avoid call blocking in the simulations
- With this static profile LBC not a very important issue -> with more dynamic profiles more challenging

Conclusions

- Load Balancing with directed handovers can
  - Be a very efficient way to enhance system wide Resource Utilization
  - Enhance the possibility to fulfill QoS guarantees in Mobile WiMAX
- However Load Balancing cannot itself ensure that enough resources released for incoming high priority traffic (e.g. handovers) -> the use of guard bands should be considered
- The basic algorithm performed well in the (rather static) simulated environment
  - Could be deployed as such
  - Still a clear need for the enhancements introduced earlier, especially when
    - Mobility increases
    - Traffic becomes more fluctuating

Possible future work

- More elaborate evaluation
  - Enhancements (auto tune, BE load, multiple threshold, guard band triggered)
  - rtPS and nrtPS scheduling services
  - Corresponding more fluctuating traffic
  - Impact of mobility
    - rescue handovers (prioritization)
    - different handover mechanisms
- Supporting fields of research
  - Location and velocity estimation (i.e. identifying static/mobile MSs)
  - Effect of transmission power and interference
  - Admission control and resource consumption estimation
- Future Load Balancing research
  - From micro to macro cells
  - To other parallel systems (e.g. UMTS)
  - Relay stations (IEEE 802.16j)

Questions?

References

- [IEEE05] Air interface for fixed broadband wireless access systems - amendment for physical and medium access control layers for combined fixed and mobile operation in licensed bands. IEEE Standard 802.16e, Dec. 2005.

- [ASN3] WiMAX Forum Network Architecture (Stage 3: Detailed Protocols and Procedures - Release 1.1.0)

- [Vel04] H. Velayos, V. Aleo, and G. Karlsson, ”Load Balancing in overlapping wireless LAN cells,” 2004 IEEE International Conference on communications, Volume 7, pp. 3833 - 3836, June 2004.