Network Selection Simulations

Future Internet Program of TIVIT,

Activity 2.4 Deliverable DA2.4.3, 1H2009

Contributors:

Pasi Seittenranta
Sami Heinonen
Timo Ketonen
Janne Tervonen

Abbreviations and Terminology

AC	Admission Control
AP	Access Point
BSS	Basic Service Set (WLAN network)
CQI	Channel Quality Information
FTP	File Transfer Protocol
HO	Handover
HSDPA	High Speed Downlink Packet Access
IKR	Institut für Kommunikationsnetze und Rechnersysteme (Stuttgart University)
MHN	Mastering Heterogeneous Networks
MIP	Mobile IP
NodeB	Base station in 3GPP network
NWDS	Network Discovery and Selection
PBRM	Policy-Based Resource Management
QoS	Quality of Service
RAT	Radio Access Technology
RNC	Radio Network Controller
TCP	Transmission Control Protocol
UMTS	Universal Mobile Telecommunications System
UE	User Equipment
WCDMA	Wideband Code Division Multiple Access
WLAN	Wireless Local Area Network

1. Introduction

This document is deliverable DA2.4.3 for activity 2.4 of Future Internet program of TIVIT. This document covers the task for verifying and simulating the effects of different kinds of network selection information provided by PBRM server.

The PBRM concept itself is defined in earlier FI deliverable [1]. As described on that document, PBRM server can provide to the UEs both network discovery and network selection information. Basically, network discovery information can assist UE's scanning procedure by defining how, when and what networks to scan. Network selection information is a tool for the operators or service providers to influence what networks UEs are using.

This document concentrates on network selection information by studying what kind of effects different network selection information may have for a single UE as well as for the whole network. This work was conducted by developing a simulator environment consisting of different radio access technologies – WLAN and 3G – together with an implementation of PBRM server. A number of network selection algorithms and scenarios were tested in the simulator environment.

The purpose of this paper is to describe the results of the simulations for different network selection algorithms in a heterogeneous network environment. All the algorithms considered are mobile based: a UE selects a network according to the policies it received from PBRM server, using also load and signal quality/strength as network selection parameters where applicable.

The content of the documentation is as follows:

Chapter 2: General background information on simulation setup. Also, the basic algorithms used in the simulations are described.

Chapter 3: Simulation results of the basic algorithms are presented (one WLAN and one 3G network).

Chapter 4: Simulation results when an additional WLAN network has been added (two different WLAN networks and one 3G network).

Chapter 5: Simulation results for the modified algorithms and a new "greedy" algorithm introduced.

Each simulation result chapter consists of:

- a brief overview of the algorithms
- a brief conclusion of the results
- more detailed graphs of the simulation results

1.1 Main results

- It seems clear that having PBRM type of network selection information server deployed in a heterogeneous network environment would improve the overall performance, when compared to scenarios where no such a server is used.
- In a simple heterogeneous network deployment (e.g. one WLAN network overlapping 3G coverage), a policy preferring WLAN when available may work well enough, especially in cases where overload of WLAN can be prevented by e.g. network planning.
- It seems that the use of load information would improve overall performance. However, the gain margin depends heavily on the scenario. When taking into account the added complexity for load information gathering, one has to carefully consider the pain versus the gain in order to estimate the usefulness of these load-based mechanisms in practice.

2. Simulation Cases

In this chapter, the simulation setup is described. First, some general background information and definitions are given. The actual simulated algorithms are described together with some remarks on simulator implementation and generated traffic.

2.1 General

Within this work, several different algorithms have been identified as research subjects for simulation work. The purpose of the simulation work is to compare these different algorithms against each other. The idea is to find and verify benefits or drawbacks of different algorithms compared to other algorithms. It should be noted that simulation results are just relative, not exact: for example, it is possible to say network selection algorithm 2 is better than 1 with rough estimates in this specific scenario, but exact figures for gains – e.g. algorithm 2 is 11,5% better than 1 – cannot be given.

In the simulations, it is assumed there are at least two different networks available (coverage may be different for different access networks). Also, there might be more than one access network supporting the same technology on a specific location (e.g. several WLAN networks).

It is assumed that UEs support necessary mobility protocols to enable inter-system handovers. In practice, this would mean Mobile IP support or equivalent.

In the simulations, the measurement criteria used is “satisfied user” (defined in chapter 2.2). All users are regarded equal, i.e. there are no different subscriber classes.

2.2 Definitions

Satisfied user: For TCP connections, the criterion is that mean bitrate in a certain time period is above a threshold. The period length and the threshold may depend on the application. A period length of 5 seconds may be appropriate for ftp and www transmissions. Also transmissions shorter than the period length do not make the user unsatisfied no matter what the bitrate is.

If there are streaming connections, the criterion is that the proportion of lost (due to transmission buffer overflow) or delayed (due to receive buffer underflow) packets is below a threshold. Again the threshold may depend on the application.

These should indicate the congestion of the access networks well enough although they do not measure the degree of satisfaction of individual users accurately.

Cell Load: At least the following load levels should be separated:

1. “Low”: Radio resources are underutilized. It is probable that a new user can be added so that all users stay satisfied.
2. “Medium”: Radio resources are (almost) in full use. Adding a new user degrades the service level of other users, but there is still good chance that all users will be satisfied.
3. “High”: Radio resources are in full use and there is a significant risk that adding a new user would make one or more users unsatisfied.

In 802.11 specification [2], there are 3 AP load related measurements: number of (associated) users, channel utilization and mean access delays. Corresponding measurements could be done in 3G. The first two are used in the cell load definition in the simulations (see also 2.4.5):

Channel utilization can be used to separate “low” and “medium” load levels. RAT specific threshold values are needed for this. Number of users could be used to separate “medium” and “high” load levels. Again, RAT specific threshold values are needed that depend on cell capacity and demands of the applications.

Reasons for network re-selection or handover: When UE selects a network it should stay on that network until one of the following handover/re-selection trigger is valid:

- Service level is degraded
    - bad signal quality
    - or congestion
- New application is launched

2.3 Network Selection Algorithms based on PBRM information

As described in [1], PBRM server can be used to deliver network selection information to the UE: first, UE requests network selection information from the network (i.e. PBRM server). Then the PBRM server constructs the network selection information for that specific user and provides it to the UE. The contents of the access network list may be affected e.g. by the load situation in the network.

This solution is UE-centric: UE has the final decision on what network to access based on the information provided from the network. UE shall take the information the network provides (prioritized access network list) into account when performing network selection.

In addition to selecting a network UE also selects a cell/access point from the chosen network. The cell is selected mostly using the same criteria that were used to select the network, except when more information is needed, which is stated in the description of each algorithm in 2.3.1.

There are a couple of cell selection principles that apply to all selection algorithms:

1. A WLAN AP can only be selected if the UE is in its coverage area and the AP does not block the UE.
2. A 3G cell can only be selected if the cell does not block the UE. There is no minimum requirement for signal strength or coverage.

If the UE is not in the coverage area of any existing network, UE will select a 3G cell that does not block the UE even if the signal strength is very poor. Cell selection can only fail if there are no 3G cells or all of them block the connection request. Contrary to the real-life implementations, admission control is separate for each cell (in live networks, 3G admission control is applied on network level, not on cell level).

2.3.1 Algorithms

1. Current status in today's networks: no network assistance or control for network selection.

   For example, 90 % of UEs/users choose 3G network, since accessing WLAN requires manual steps (in today's terminals). This scenario is used as a benchmark for the other scenarios/algorithms. Unsatisfied users (refer to 2.2 for definitions) do not change access network (connection terminated).

   In this case UE selects the network by using the probabilities above: if UE is in WLAN coverage, it is selected with 10 % probability. Otherwise, 3G network is selected.

   UE selects the cell with the strongest signal in the chosen network. In the current simulation implementation, signal strength is a strictly decreasing function of the distance between the terminal and cell/AP.

2. PBRM server provides static access network list, without any load information included. The access network list contains only network-level information (e.g. no info per AP). Unsatisfied users do not perform network re-selection, i.e. they are just marked as unsatisfied and the connection is terminated (i.e. resources freed for other users). Otherwise, similar to the third algorithm.

   This algorithm reflects the case what we have when only PBRM server is deployed without any mobility support (i.e. only network selection information provided, no handover support due to MIP/equivalent support missing). The access network list gives preference over different networks.

   UE selects the first network (from PBRM priority list, i.e. PBRM policy) whose area the UE is in.

   UE selects the cell with the strongest signal in the chosen network.

3. PBRM server provides static access network list, without any load information included. The access network list contains only network-level information (e.g. no info per AP). Unsatisfied users perform network re-selection (i.e. handover due to service degradation without mobility): if after trying all the available networks UE/user is still unsatisfied, the connection is terminated and user marked as unsatisfied. A new network selection is performed only when the quality of the current connection degrades (i.e. when the user becomes unsatisfied).

   This may be a realistic case when it can be assumed that WLAN will not get overloaded, e.g. home networks and enterprise networks where the WLAN network is planned so that it has enough capacity.

   Another reason for simulating this case is that it would provide a reference point to get comparisons for how much e.g. load info would improve compared to this algorithm.

   UE selects the first network (from PBRM priority list, i.e. PBRM policy) whose area the UE is in.

   UE selects the cell with the strongest signal in the chosen network.

4. PBRM provides static access network list, admission control implemented to prevent overload situations. The access network list contains only network-level information (e.g. no info per AP). Unsatisfied users and UEs with no access permission (due to AC) perform network re-selection (i.e. handover due to service degradation without mobility).

   In this case the signal strength is used in selecting the network as described in 2.4.1.

   UE selects a cell that accepts new connections and has the strongest signal in the chosen network.

   Note: in real life implementations, admission control for 3G covers all the cells, i.e. if access denied in one cell, it is not allowed for any other either. In the context of this simulation, the admission control functionality should be considered to be part of resource allocation: the access denied means no available resources on that cell, but there might be resources on some other cell(s).

5. PBRM provides load information to UEs together with the access network list. The access network list contains AP/cell-specific load information that PBRM server gathers from networks. Load information is network-specific; it is up to UE to “normalize” the load information in its access selection routine. Unsatisfied users perform network re-selection (i.e. handover due to service degradation without mobility).

   Load information is bound to time and location. Thus, UE is required to provide its location to PBRM server (e.g. on cell id level, or routing/tracking area). The information PBRM provides is required to be filtered by this location information (i.e. only limited number of entries included in the access network list).

Also, lifetime of this load-based information is limited. One validity criteria is the location where information is applicable (i.e. UE requests new information from PBRM when leaving the area where the previous information request was made). Also the time should be taken into account if UE can initiate a new connection on the same location. It should be simulated to see what lifetime value for load information would give the best overall result.

In this case UE combines load info with signal strength in network selection algorithm, as specified in chapter 2.4.1. In addition to the algorithm specified in 2.4.1, also other algorithms may be implemented for comparison (e.g. bitrate estimation based algorithm).

The above described algorithms are simulated for various scenarios and different setups. The results of those simulations are described on chapters 3 and 4. For the comparison of those results, a modified set of algorithms was defined and simulated. These modified algorithms and the corresponding further simulation results are described in chapter 5.

2.4 Simulation Implementation

The simulator is based on the IKR Simulation Library V2.6 [3] developed by Stuttgart University. The IKR simulator provides a well-working framework for further developments and additions. The IKR simulator does not contain radio models, so they were developed within the project. The theoretical maximum bitrate for WLAN in the simulations is 54 Mbit/s, i.e. the radio model corresponds 802.11g. Also, features defined in 802.11k are included. 3G radio model is based on HSDPA, and its theoretical maximum bitrate is 14.4 Mbit/s (downlink).

2.4.1 Algorithm Definitions

PBRM server provides the list of access networks (2.3.1 cases 2-5); UE goes through the list in the order defined by the PBRM server and chooses the first that fulfills UE’s criteria.

Criteria:

- signal strength above some minimum level (in algorithms 1-5)
- load level is “low” or “medium” if possible (in algorithm 5)
- criterion in access control (in algorithm 4): access is denied if load level is “high”.

If two or more access networks are ranked equal in the list, then choose the one that has the lowest load level provided that the signal strength is acceptable (algorithm 5). If also load levels are equal or load levels are not known, then the signal strengths decide.

For the algorithm 5 (load information provided from PBRM), the access networks are not ordered by priority: it seems it doesn’t make sense to prioritize the access when load information is provided. Instead, it is up to UE to decide what network to select based only on the load information.

2.4.2 Network Configuration

For the basic scenario, there is one WLAN network and one 3G network in the simulations (chapter 3). 3G has full coverage in the simulation area (300 x 300 meters), but WLAN only in certain spots. In the simulations, the number of WLAN APs was varied, but in the basic scenario there were four WLAN APs.

Another set of simulations contain three networks (chapter 4): two separate WLAN networks with either two or four APs, one 3G network.

Only NodeBs and WLAN APs are modeled in detail in the simulator, e.g. RNC functionality is not implemented as such. This means that some of the necessary RNC functionality (e.g. for network/cell (re-)selection) is divided to other network elements/functions in the simulator. The simulations concentrate on radio resource utilization in different scenarios, so this simplification is justified.

Results of the simulations may depend highly on the network configuration. Therefore several different configurations are needed.

2.4.3 Traffic

In order to make as clear difference as possible between the algorithms, the number of UEs should be high enough to get unsatisfied users, i.e. there should be enough load in the simulation scenarios. Applications should generate different amounts of traffic.

Applications used in the simulations:

- bursty web-page downloads over TCP
- FTP download over TCP
- YouTube download over TCP (higher bitrate requirement than for FTP download)
- constant bit rate video streaming over UDP (downlink)

It is assumed that downlink is a bottleneck in 3G model. However, WLAN channels are half-duplex and therefore uplink traffic is also significant.

Priorities: 4 access categories implemented in WLAN, but only best effort is used in the beginning. QoS classes implemented also for 3G model to support video streaming.

2.4.4 UE Mobility

UEs are stationary, i.e. there is no mobility support. When added into the simulation, a UE is placed in a random position in the simulation area. IKR supports a concept of hotspots: in practice, the probability for a UE to be positioned into a hotspot is higher than outside the hotspot. Hotspots may or may not be co-located with WLAN APs in the simulation scenarios.

Results of the simulations depend highly on the placing of the UEs. Therefore several alternatives need to be studied.

2.4.5 Remarks

Usually, there is application traffic load that adapts to the available capacity when RTT increases (TCP acknowledgement mechanism in the simulator). Therefore, although channel utilization and access delay measure the offered load they are not well suited to estimate user satisfaction alone. Channel utilization and access delay measure basically the same thing, but channel utilization is more illustrative. On the other hand we can separate priorities when measuring access delays and thus they are more informative if several priorities are used.

In real WLAN network the number of associated users may contain also idle users. In addition, the (unlicensed) channel may be utilized e.g. by interfering stations not belonging to the BSS. These aspects do not appear in the simulator. Therefore the number of users is not as good indicator of load in reality as it is in the simulator. Possibly one could count the number of active users instead of the number of associated users to solve these problems partly.

3. Basic simulations – 2 networks

In these simulations there is one WLAN network and one 3G (HSDPA) network. The simulated algorithms are defined in 2.3.1.

3.1 Notes on algorithms

3.1.1 Algorithm 1

Pros:

- No extra delays because of network selection and vertical HOs.

Cons:

- No load sharing between networks. In practice WLAN underutilized while 3G overloaded.

3.1.2 Algorithm 2

There are 2 alternatives, when scaling load higher:

- WLAN network becomes first overloaded
- 3G becomes first overloaded (by users with no WLAN coverage)

In the latter case this network selection algorithm is optimal.

Pros:

- WLAN is used efficiently to decrease 3G load. In practice algorithm 2 is always better than algorithm 1 measured in terms of satisfied users.

Cons:

- WLAN may get overloaded while 3G is underutilized.
- Initial network selection is slower than in algorithm 1.

3.1.3 Algorithm 3

Vertical HO can be used to balance load after static initial network selection.

Pros:

- If WLAN becomes overloaded while 3G is underutilized, then HO from WLAN to 3G improves results compared to algorithm 2 in terms of satisfied users.

Cons:

- Starting connections in a crowded network and performing then vertical HOs decreases connection quality.
- Users would get better service on average if the load was more balanced also when there is no overload (unsatisfied users) in the networks.

3.1.4 Algorithm 4

Pros:

- Less vertical HOs compared to algorithm 3 may be needed because of better initial network selection.
- In case of overloaded system admission control may block some users (in both networks), but the number of non-blocked unsatisfied users decreases. Then the number of satisfied users should increase compared to algorithm 3 as the unsatisfied users consume less bandwidth. Also from the user point of view it is better to block a connection in start-up than interrupt it later.

Cons:

- Finding decent AC parameters may be difficult in practice. If AC is too strict, then some users are blocked unnecessarily during load peaks and the system capacity may be even worse than in algorithm 3. If AC is too loose, then the gain compared to algorithm 3 reduces. This was also clearly visible in the simulations: AC parameters working really well in one scenario resulted often lousy performance in another.
- Users would get better service on average if the load was more balanced also when there is no overload (unsatisfied users) in the networks.

3.1.5 Algorithm 5

Admission control might bring the same benefits in this algorithm as in algorithm 4 when both networks are fully utilized.

Pros:

- Less vertical HOs compared to algorithm 3 may be needed because of more advanced initial network selection.

- More balanced load means that users experience better service quality on average also if there is no overload (unsatisfied users) in the networks.
- More efficient use of radio resources may improve system capacity (the cell with better signal strength is selected if the load levels are equal).

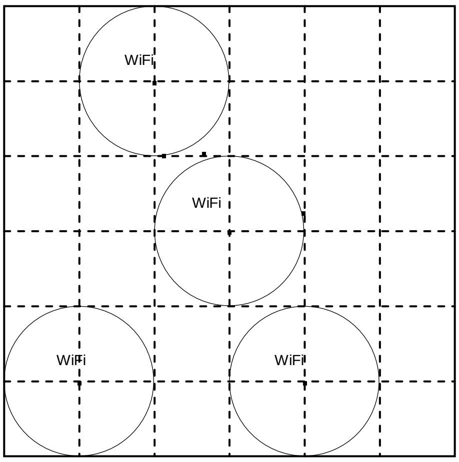
Cons:

- Possibly less satisfied users compared to algorithm 3 if 3G is selected more frequently, since it is always possible to make a handover from overloaded WLAN to 3G, but opposite is possible only for those terminals that are in WLAN coverage area.
- Possibly less satisfied users compared to algorithm 4 in heavily loaded system because of missing AC. Also useless HOs from overloaded WLAN to overloaded 3G and vice-versa may appear.
- Load estimation may work worse in reality than in simulations as in algorithm 4.

3.2 Assumptions

3.2.1 Network configuration and hotspots

Simulation area is 300 m x 300 m. There are 4 WLAN access points in the simulation area and one 3G cell in the center of the simulation area. 3G cell covers the whole simulation area.



3.2.2 Traffic mix

Three kinds of users:

FTP: 80% of users; file size random, mean 4 Mbit; minimum bitrate for satisfied user 100 kbit/s
YouTube: 10% of users; file size random, mean 24 Mbit, min. bitrate 290 kbit/s
Video: 10% of users; video length random, mean 600 s, bitrate 500 kbit/s

FTP and YouTube use TCP protocol, whereas video is UDP constant bitrate stream.

Mean file size is then 35.6 Mbit.

3.2.3 Remarks

WLAN coverage is in 35% of the simulation area. 93.5% of users are in WLAN coverage area. This is a large number, but it is selected to emphasize the differences of the algorithms. When there is only 3G network available, the network selection algorithm does not play any role.

In the simulations, 3G connections have CQI=30 in the whole simulation area and the cell throughput is 12 Mbit/s (downlink). WLAN coverage is 50 m, and an AP throughput is 4-16 Mbit/s (half-duplex, 100% utilization) depending on the distance. Using random distances the simulator gives 7.2 Mbit/s AP throughput.

If there were no deviation in the load, then 3G cell could have (12 Mbit/s) / (0.9*290 kbit/s + 0.1*500 kbit/s) = 39 simultaneous satisfied users, and arrival rate could be (12 Mbit/s) / 35.6 Mbit) = 0.34 users/s. WLAN network could serve 4*(7.2 Mbit/s) / (35.6 Mbit) = 0.81 users/s. Load deviation decreases utilization and consequently throughput, and is greater in WLAN APs.

Since WLAN proportion of the heterogeneous network capacity (4*7.2 / (12 + 4*7.2))*100% = 71% is less than the proportion of users in WLAN coverage area, some users in WLAN coverage area should use 3G in optimal network selection algorithm. This means that algorithm 3 is expected to give better system capacity than 2.

As the above calculation shows, the gains of the algorithms depend on the network configuration and load distribution.

3.3 Results

X-axis in the figures is offered load (mean user arrival rate multiplied by mean application size) [Mbit/s].

Algorithm 1 performance is bad in Figure 1. The reason seems to be low WLAN utilization as shown in Figure 4.

Algorithm 2 is worse than algorithm 3. Figure 5 shows that algorithm 3 can utilize 3G better (due to vertical HO) while WLAN utilizations are equal in Figure 4.

Algorithm 4 performance depends on the selected parameters. The benefit compared to algorithm 3 is smaller number of vertical HOs (Figure 2).

Algorithm 5 seems to give better capacity than others (Figure 1) and less handovers (Figure 2) than 3.

Figure 3 shows throughput. It should be equal to the offered load in x-axis when there is no overload.

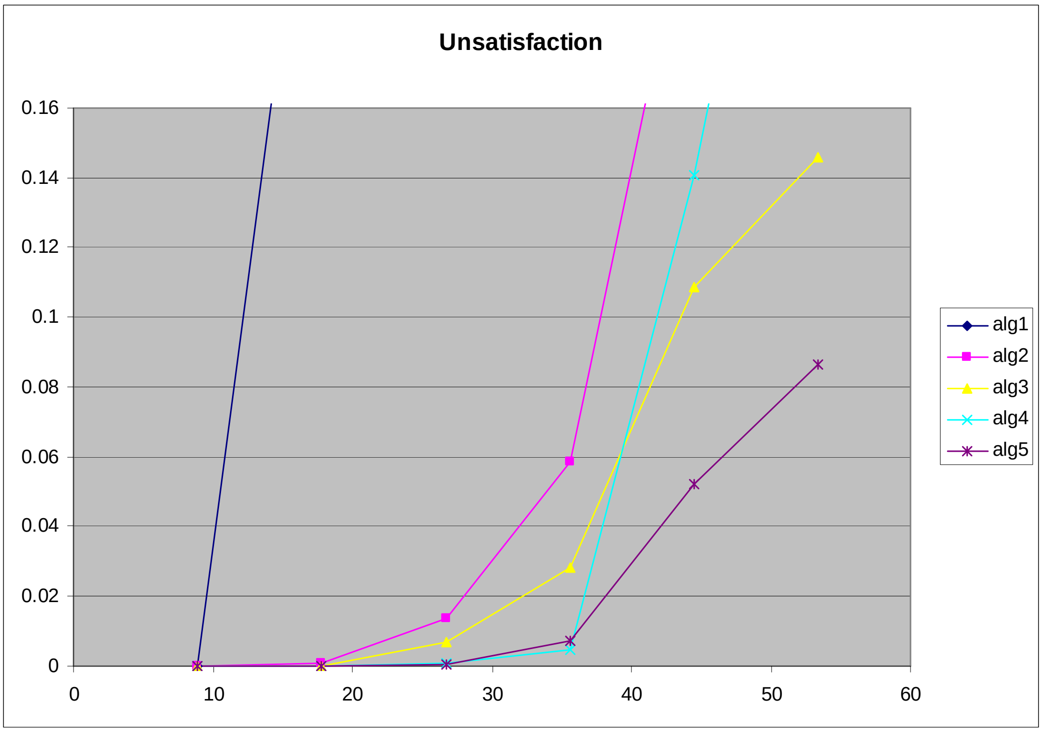


Figure 1: Proportion of unsatisfied users

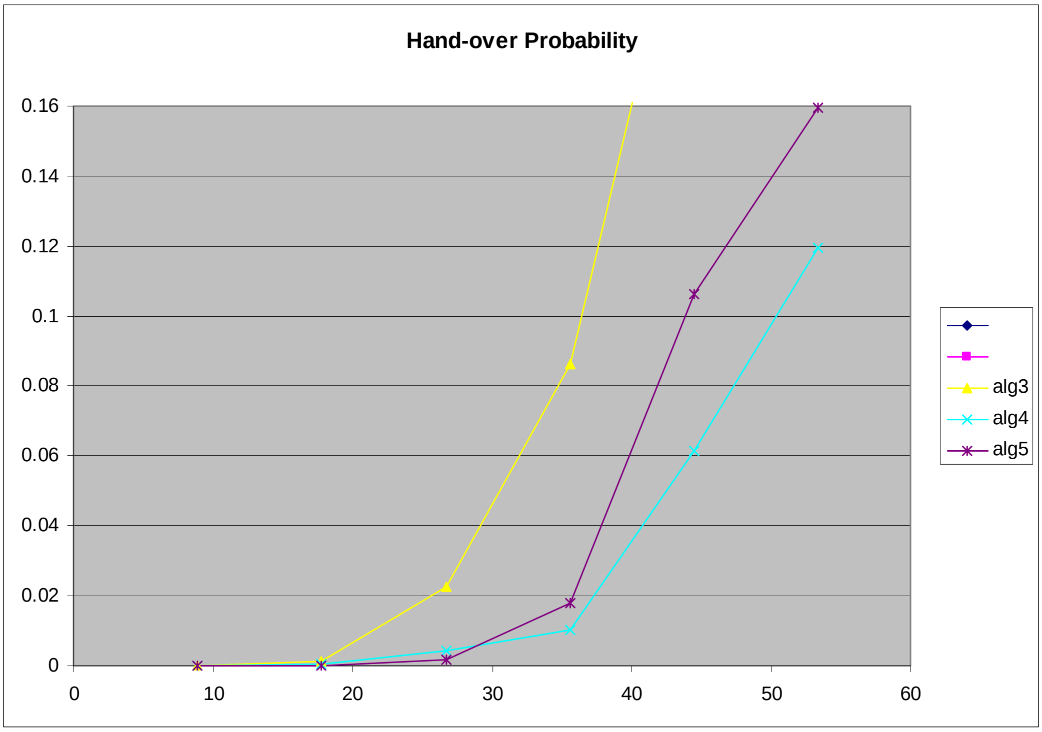


Figure 2: Proportion of users that made vertical HO

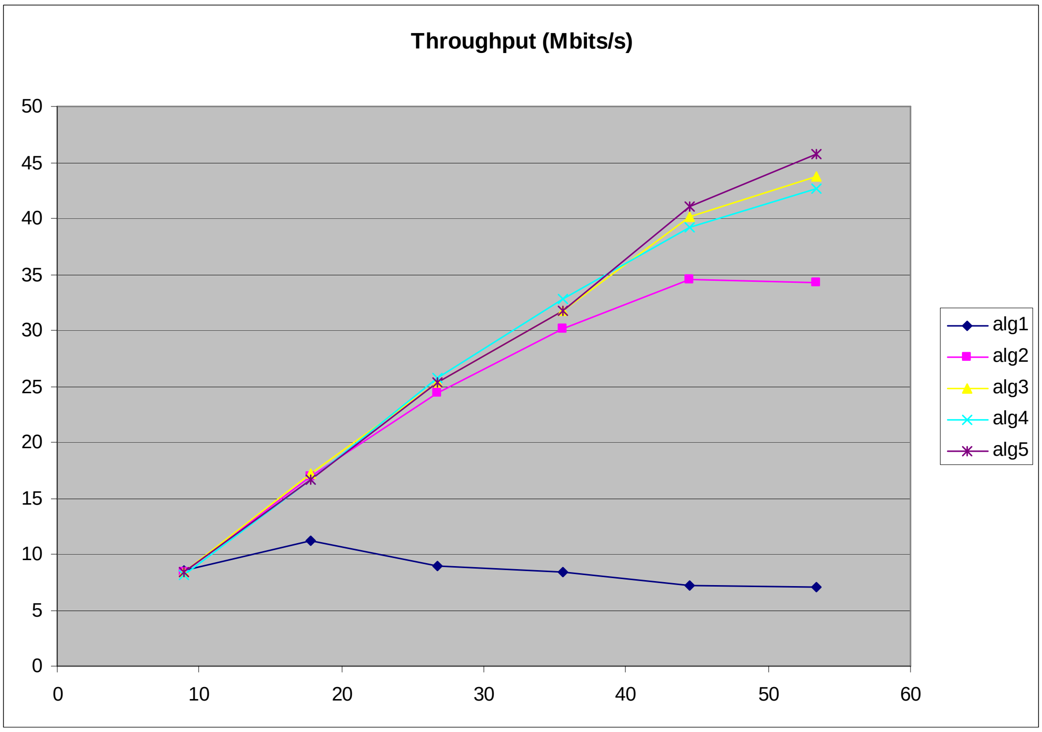


Figure 3: Mean system throughput [Mbit/s]

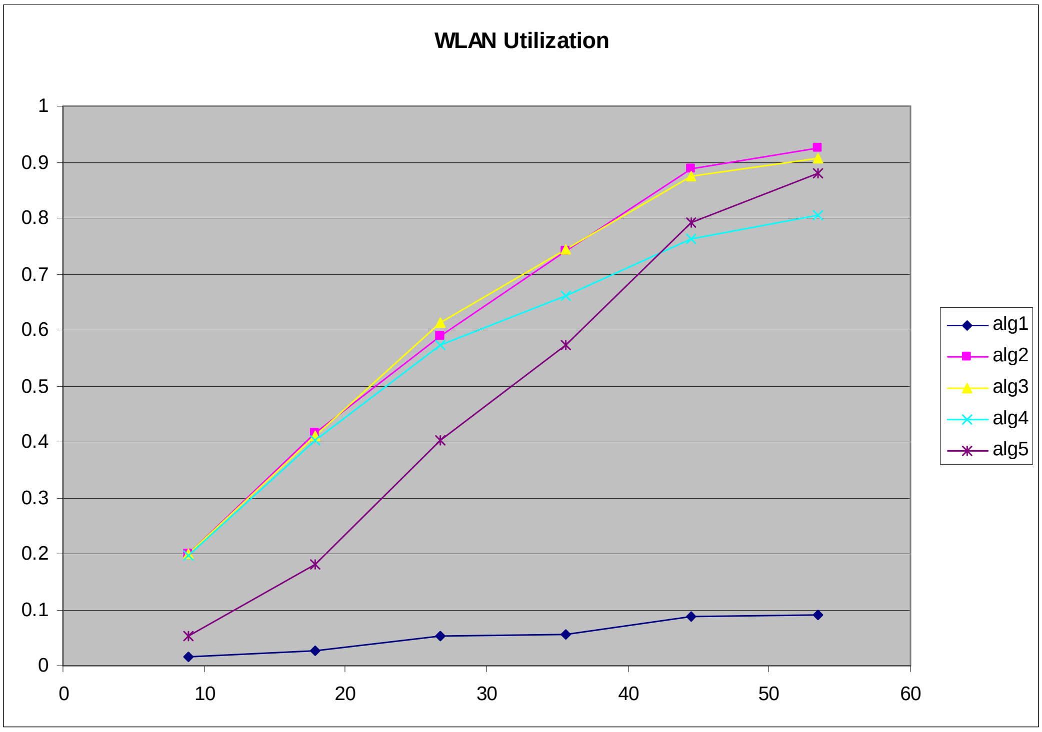


Figure 4: Mean WLAN channel utilization

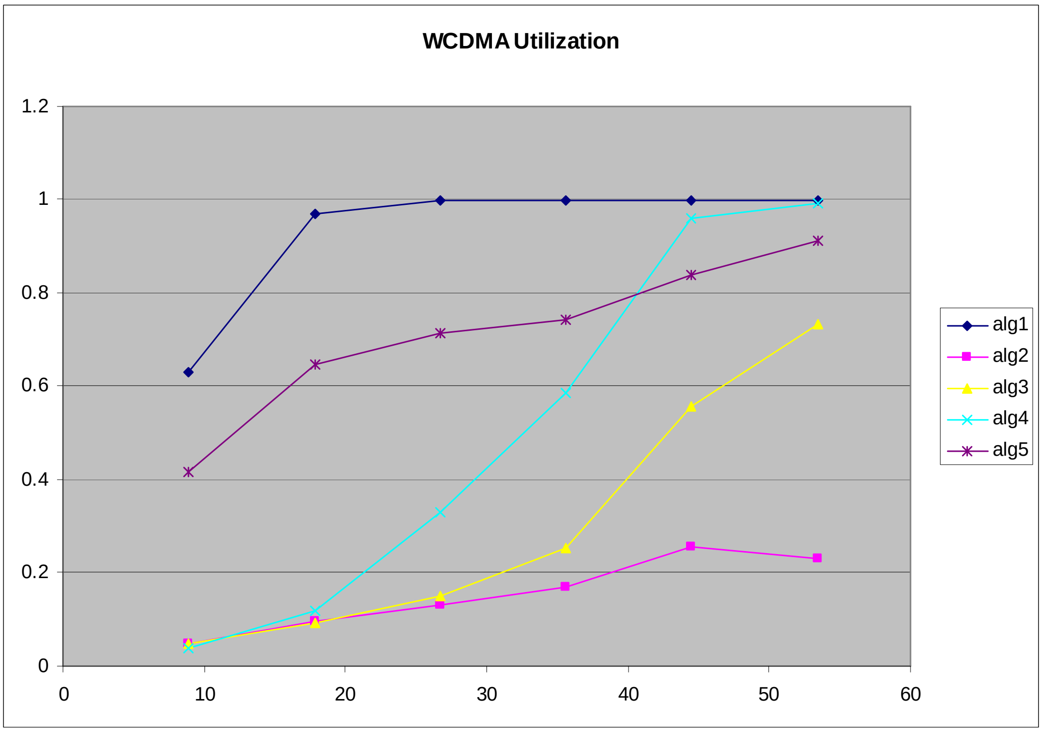


Figure 5: Mean 3G channel utilization

3.4 Conclusions

The algorithms 2 – 5 give improvements compared to algorithm 1. If hotspot load is low enough so that WLAN APs don't get overloaded, then algorithm 2 is adequate. Otherwise more advanced algorithm is needed. Note that algorithms 3-5 require mobile IP feature (or similar) for vertical HOs. In general, algorithm 5 performs best, but the margin depends heavily on initial simulation parameter settings. Also gathering load information may be difficult to implement. Algorithm 3 may be adequate when selecting between two networks only.

4. Basic simulations – 3 networks

In these simulations there are two fully overlapping WLAN networks and one 3G network.

4.1 Algorithms

Some changes to the algorithms are needed to the descriptions in previous section. The changes are summarized below:

1. 10 % of the users in WLAN coverage area select the WLAN with the strongest signal strength, others select 3G.

2. WLAN networks are preferred over 3G. The first WLAN is preferred over the second. This resembles the basic functionality provided by a PBRM server.

3. Initial selection as in algorithm 2. An unsatisfied user performs network re-selection to another network. The network earlier in the preference list is selected in re-selection.

4. No changes.

5. No changes.

Algorithm 2 has the problem that the load is not shared between the two overlapping WLANs, but the first one is always selected. In a modified algorithm 2’ WLANs are ranked equal by altering the order of the WLANs in the network list provided by the PBRM server, i.e. every second terminal gets the network list with the other WLAN network as top priority.

In algorithm 3 handovers can be done between the two overlapping WLANs in the head of the network list, but not from WLAN to 3G network, no matter what the load situation is. Therefore a modified version 3’ is added, where HO back to previously used network is not allowed.

4.2 Assumptions

Two out of the four hotspots in the previous section are used. Each hotspot has two WLAN access points – one from each network, so that total number of WLAN access points is four. The 3G network and traffic mix is similar as in the previous section. Then the number of users that the system is able to serve should not change much.

WLAN coverage is in 17% of the simulation area. 92% of users are in WLAN coverage area.

4.3 Results

Results are quite similar to the two network case (see Figure 8, Figure 9, Figure 10). From the figures Figure 8 and Figure 9, it is clearly visible that the basic algorithm 2 cannot take advantage of the two WLAN networks, i.e. the second WLAN network (in the prioritized network list) is not used at all. If an operator would have this kind of situation (more than two networks in the same area), the operator needed to take this into account e.g. by altering the prioritization for different users, as implemented in algorithm 2’ in the simulations.

Modifications in algorithm 2’ and 3’ improve 2 and 3 remarkably (Figure 6). Both 3 and 3’ have large number of vertical HOs (Figure 7).

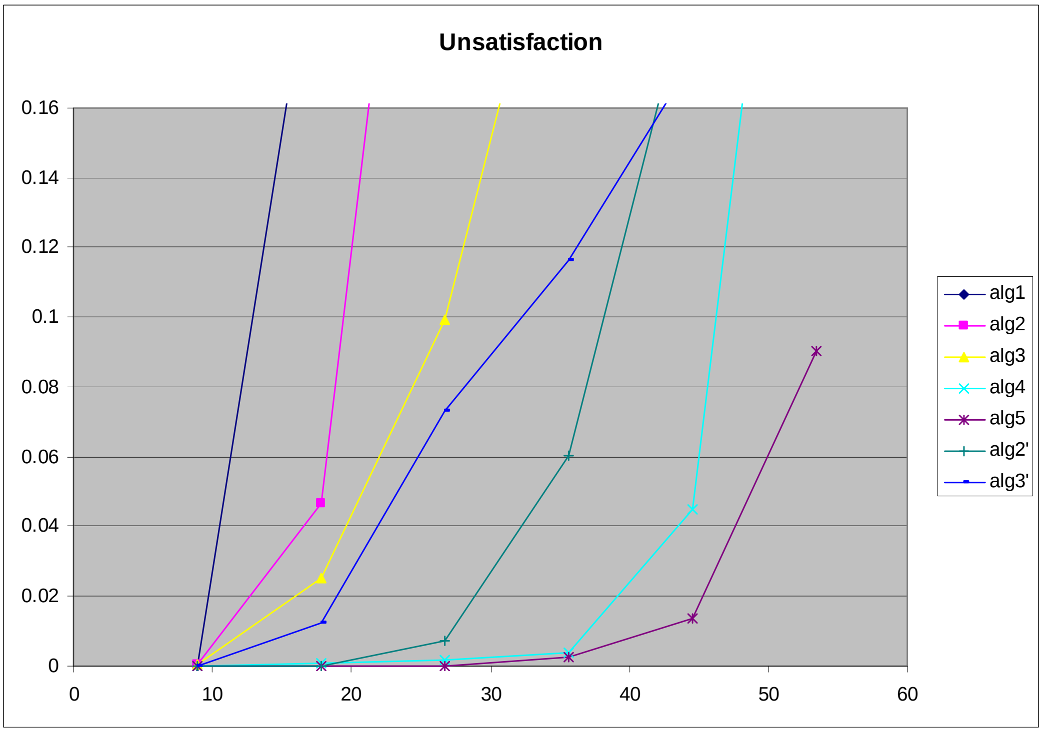


Figure 6: Proportion of unsatisfied users

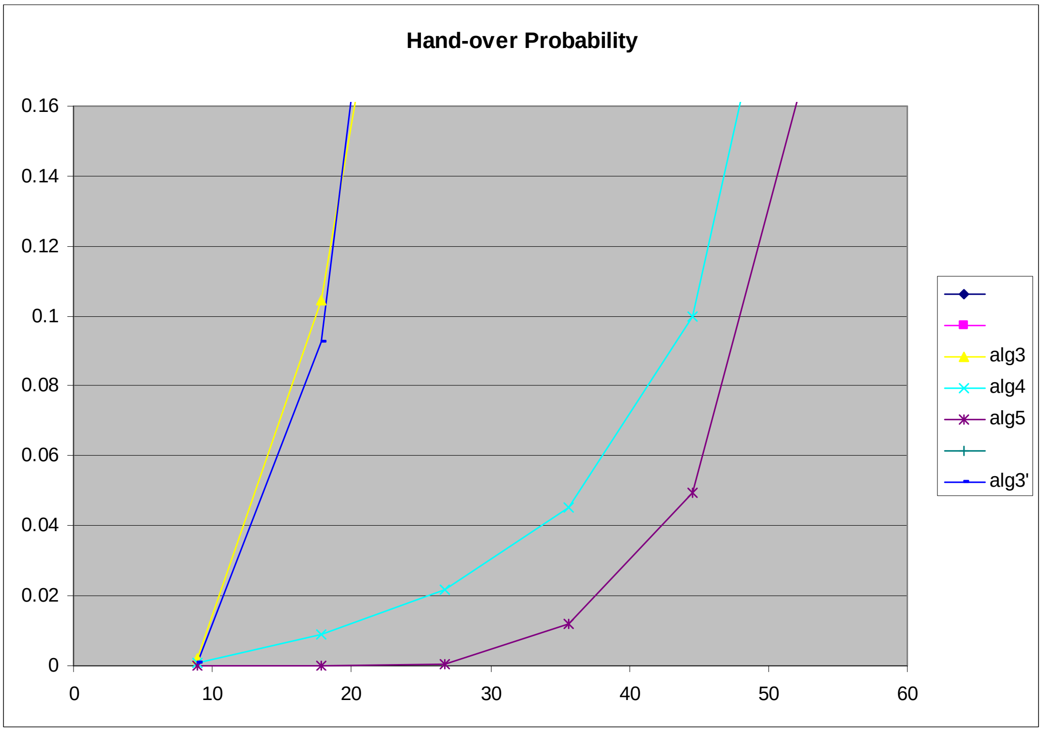


Figure 7: Proportion of users that made vertical HO

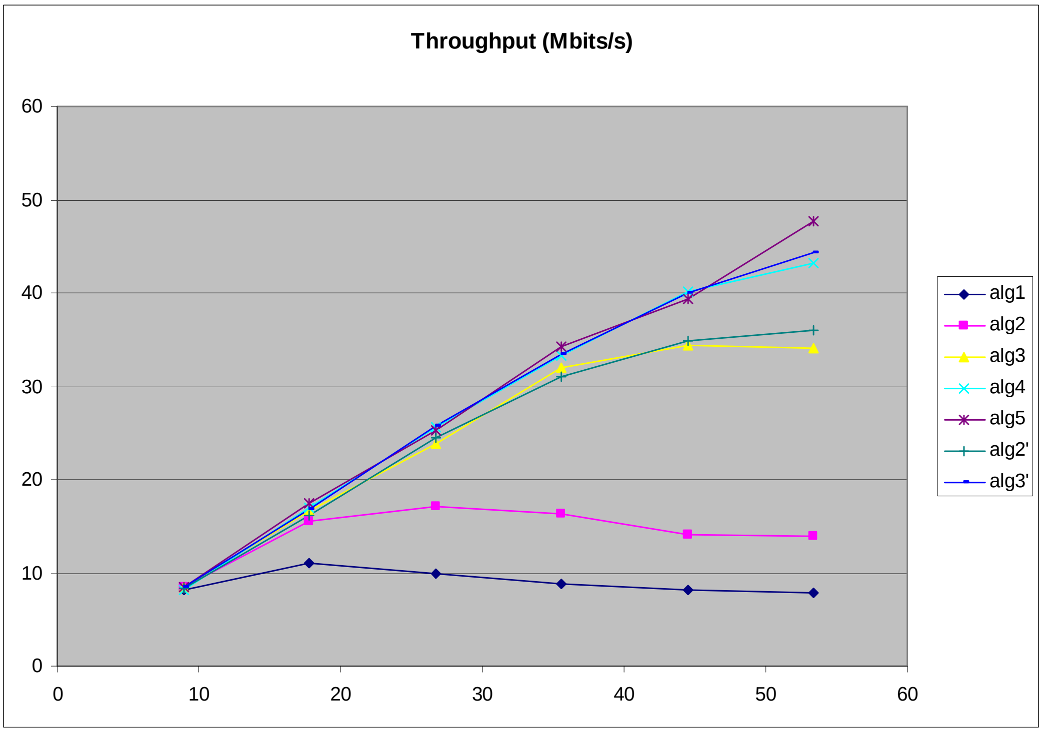


Figure 8: Mean system throughput [Mbit/s]

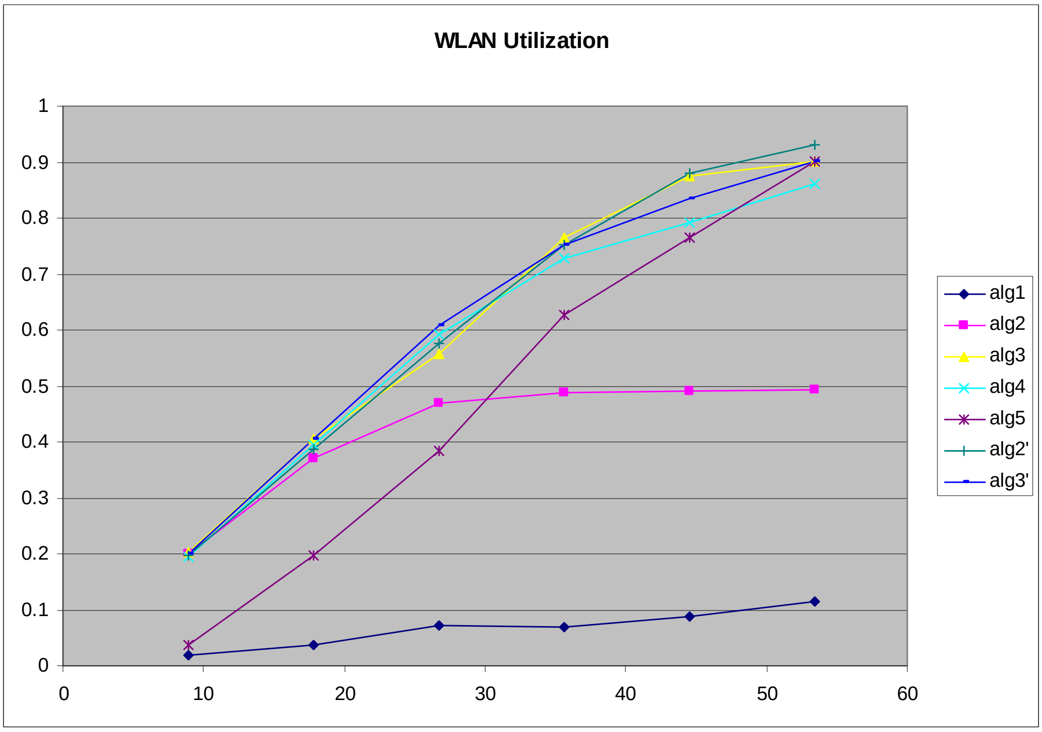


Figure 9: Mean WLAN channel utilization

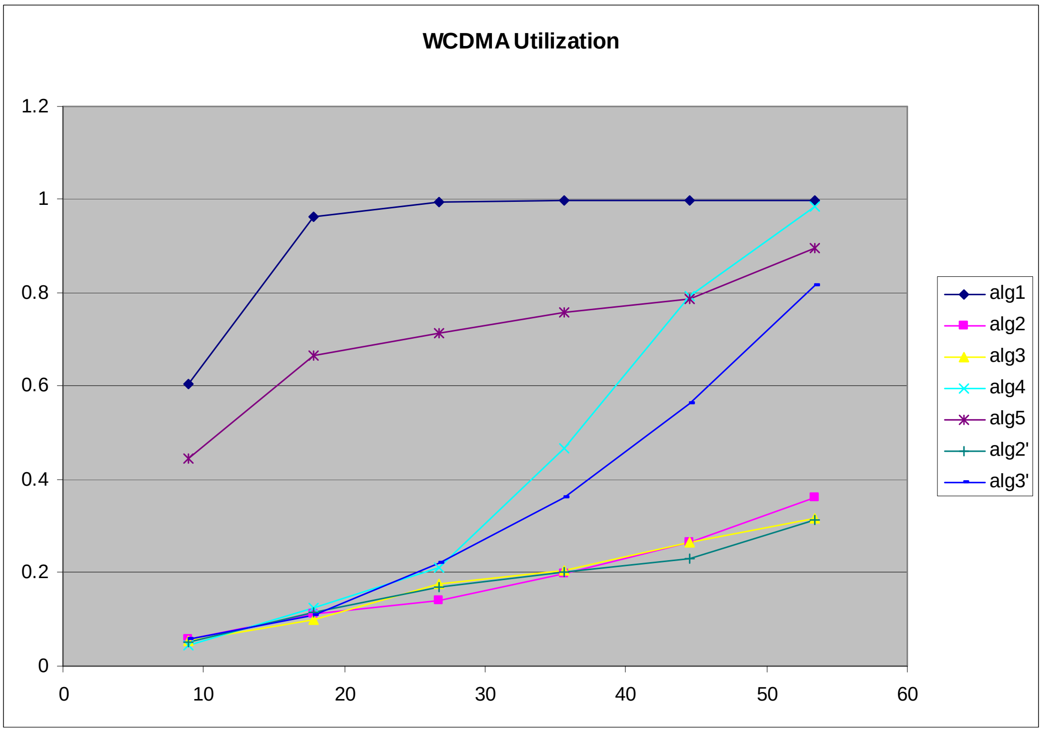


Figure 10: Mean 3G channel utilization

4.4 Conclusions

Modified algorithms 2' and 3' outdo 2 and 3. If hotspot load is low enough so that WLAN APs don't get overloaded, then algorithm 2' is adequate. Otherwise more advanced algorithm is needed. Number of vertical HOs may be a problem in algorithm 3'. Therefore algorithm that takes load information into account in network selection may be needed in three network case.

5. Modified simulations

5.1 Algorithms

In this chapter, different variations of the basic scenarios described in chapters 3 and 4 are studied. For example, the initial values of simulation scenarios are varied to estimate the impact of different scenario setups. Also, couple of new dynamic algorithms taking into account the load situation was more closely examined.

In the following simulations, for an algorithm from above chapters, there are 4 different variations: basic algorithm, basic+HO, basic+break, basic +HO+break. Also a “greedy” algorithm was added.

The algorithms used were as follows:

- Algorithm 1 (manual, 10% to WLAN when coverage) has been kept the same.

- Algorithm 2 and algorithm 3 (WLAN when coverage): algorithm 3 is in practice algorithm 2 combined with handover, so for this chapter algorithm 2 and algorithm 3 are combined into one entity. Thus, there are the following algorithms in this category: alg2, alg2+HO (= alg3), alg2+break, alg2 +HO+break. A handover is performed when the connection gets bad ("unsatisfied"), a new network is selected.

- Algorithm 4 (WLAN when coverage + WLAN admission control): 3 different cases where used for admission control as follows: Admission was denied if channel utilization was over 0.9, the number of active users is over 100, the number of active users is over 20. Admission control was used only for WLAN, the idea being to try to prevent WLAN from becoming overloaded. It also allows a mobile to connect at least to a 3G network. The scenario is that 3G provides coverage and that traffic is "offloaded" to WLAN when possible.

  It may be that admission control has to be done by the mobile: If it (somehow) detects that a cell/AP is overloaded it won't go there unless there are no other options. If admission control is done by the network (cell or AP) then it can happen that a mobile is denied admission even in the case when the cell/AP was the only possibility for the mobile.

- Algorithm 5 (separate coarse grained load & signal): alg5, alg5+HO, alg5+break, alg5 +HO+break. A handover is performed when the connection gets bad ("unsatisfied"), a new network is selected.

- Algorithm 7 ("greedy" algorithm): Network is selected according to an estimate of IP bitrate a mobile would get. The estimate is formed by estimating the maximum bitrate of a cell/AP, the number of active users, and by scaling it with coding and modulation. The latter is done by estimating the bitrate at physical level the mobile would get at its current location divided by the estimate near the cell/AP. In handovers the same criteria is used, i.e., the old network is selected (and no HO performed) in case its bitrate estimate is better than the estimate of other networks.

  In this scenario, the mobile informs the network – i.e. PBRM server – which cells/APs (of the networks defined by a policy it has) it can use. The network returns the load values of each cell/AP and the mobile calculates a bitrate estimate from it and the signal strength it has measured. Another option is that when the mobile informs the network of the available cells/APs, it includes corresponding signal strengths/quality (or location) and the network calculates the bitrate estimates.

- Algorithm 8. This is the same as above expect that bitrate estimates are checked at one second intervals and a HO is performed purely on its basis. In algorithm 7 a check is initiated only if the connection is bad ("unsatisfied"). This algorithm was tested only in some cases. The purpose was to get a "feeling" of network based load balancing algorithms. In the latter case, cell/AP "loads" could be checked e.g. at 1 second intervals and mobiles moved to other cells in case of "imbalance". An additional criterion to the latter would probably be that it should not degrade mobiles bitrate "too much". In algorithm 8 it is required that it should increase the bitrate through bitrate estimates. Otherwise it may be that algorithm 8 already does some load balancing.

The need for a HO was checked every 2 seconds; the need to break a connection due to bad service (i.e. bitrate requirement was not satisfied) was checked every 4.99 seconds.

In addition to the above algorithms, different kinds of traffic were tested. The idea was to get an understanding on the effect of traffic type. This may be of importance at least in the algorithms where load is used.

Also a network with only 2 WLAN APs was tested. Another addition was to check the effect of signal strength in 3G. In all the previous cases on chapters 3 and 4 the signal strength of 3G is constant and has the maximum value. In the “weak” 3G case, the power level of the 3G cell was lowered so that 3G’s signal strength would vary in the region.

The main variation between different simulation cases is the probability a mobile will appear in a hotspot. If it is low, WLAN will be underloaded and the “job” of the algorithm is to “minimize” 3G usage in WLAN coverage. In case it is high, WLAN will become overloaded and the “job” of the algorithm is to “maximize” 3G usage in WLAN coverage.

5.2 Conclusions

- It seems clear that the use load information would improve the overall performance, maybe up to 20% or more depending on how the comparison is done and on the simulation case i.e. network configuration and traffic.

  This is however a long term option, more work is required. In simulations the number of active users performed well, but for example, how dependent it is on traffic mix, can it be measured in practice. The simulations indicated that for load the number of active users may however be the “right” concept: The results when channel utility was used were worse; also using an average number of active users decreased performance. In principle long term averages of channel utility and of the number of active users gives the same sort of information, but the need for long term averages make it impossible to react to short term variations degrading performance.

- In a simple heterogeneous network deployment (e.g. WLAN network overlapping with 3G coverage), algorithm 2 could be reasonable. This would correspond to a case where a possible overload of a WLAN network is taken care by network planning. For example admission control to WLAN is done by access rights. In this case the users are known and therefore WLAN can be planned so that overload would be an exception.

  Vertical handovers would probably improve performance and lessen the need for a full proof network planning. It should however be noted that this may depend on e.g. round trip times, the longer it is, the more time it takes for TCP to recover from packet losses and transmission breaks. Another issue is when a HO should be performed. In the simulations, E2E bitrate criteria were used. In practice E2E bitrate may not be available; also it depends on other factors than radio interface, e.g. congestion in core network. Related to this is the stability of the load based HO, e.g. in principle it can happen that all connections in a cell decide to make a HO to another network. More work is required to define the actual criteria used for a load based HO. Also if the use of load information is the long term goal, load based HO option may not be in the upgrade path unless it can be seen as a part of mobility based HOs.

- Algorithms based on admission control may be difficult to implement. The ones tried in the simulations did not give the expected performance.

5.2.1 Notes on the greedy algorithm

It seems that the greedy algorithm could be suitable and hence it may be worth while to document it in more detail:

The basis is that load is measured by the number of (active) connections of different QoS classes.

For real-time connections this load measurement amounts to resources reserved for them.

For non-real-time connections the basis for this load measurement is that radio resources reserved for them are (more or less) shared. In case scheduling is done on a round robin basis and the QoS classes of the connections are equal, then they get approximately the same throughput at radio interface provided they have the same signal quality. In addition the protocols used for non-real-time data transmission, e.g. TCP, are typically adaptive meaning that on the average the connections share their common transmission capacity. An additional point is that when allocating a connection request, the only thing known about the request is its QoS class so also from this point of view the only reasonable point is to consider connection requests of the same QoS class as equal. The term active is needed to mark those non-real-time connections which are not in an idle state.

In non-real-time case signal quality is combined with the above load information as follows: The network estimates what would be the L3 level bitrate the connection would get if the signal quality is the best possible. This is done as follows: First the capacity left for non-real-time connections is calculated by subtracting from the total capacity the reserved capacity. Then the amount of resources to be used by the new connection is calculated by taking into account the relative amounts each existing connection and the new connection would take. This depends on QoS classes and how it affects scheduling. This gives an estimate for a maximum L3 level bitrate the connection would get provided that the signal quality is the best possible. This estimate is then scaled down by signal quality to get an approximate estimate for the L3 bitrate the terminal would get. These bitrate estimates are formed for each cell / AP the terminal hears. Depending on the operator specific network selection policies (see below), the access technology which has the best bitrate estimate is selected. It is then up to the access technology to select the actual cell / access point.

In each access technology there is some bookkeeping on existing connections. The above requires that this is extended by adding a state to a connection: active or in-active. This could be implemented by e.g. a timer: If nothing has been sent through a connection for e.g. 1 second, then it is marked as inactive.

There are two main approaches in obtaining and combining the needed load and signal quality information into one place so that the above bitrate estimate can be made: Either the (final) bitrate estimate is made at the network or at the terminal.

Non-real-time load information can be passed in at least two formats: One consisting of load information of different QoS classes and one consisting of the above mentioned maximum bitrate estimate formed by the corresponding network. The latter may be more maintainable as the estimate may depend on scheduling; also all the relevant information needed for forming the maximum bitrate estimate is at the corresponding network.

As for the signaling procedures there are at least the following possibilities:

- When a terminal needs a connection, it informs the heterogeneous network entity (e.g. PBRM server) through IP signaling of the cells and APs it hears. The entity would then request the maximum bitrate estimate from the corresponding networks for the corresponding cells / access points. QoS parameters of the request would be given as input. After receiving the bitrate estimates, the entity would send them back to the terminal. The terminal then combines this information with signal quality as explained above and selects the network.

- When a terminal needs a connection, it informs the heterogeneous network entity (e.g. PBRM server) through IP signaling of the cells and APs it hears and the corresponding signal qualities. The entity would then request the actual bitrate estimate from the corresponding networks for the corresponding cells / access points. QoS parameters of the request and the signal qualities would be given as input. This case may be more suitable when also the network knows the signal quality, e.g. when it is required by the access technology that the terminal reports it to the network. It may also be that the network has a good enough guess on the location of the terminal for estimating the needed signal qualities. In addition, in this case either the heterogeneous network entity can make the network selection or it can pass on the actual bitrate estimates for the terminal so that the terminal can make the selection.

Another option would be to broadcast the load information. This may however be difficult to implement: bitrate estimate for the new connection depends also on the QoS requirements of the connection. This would mean that the broadcasted load information should be such that the QoS parameters of the connection could be somehow combined with it to form the bitrate estimate for the connection. This may mean that some scheduling information describing how radios resources are to be used should be included into load information.

Note that the above defines what would be the best access technology in terms of bitrate. A network may be selected also by other criteria, e.g. an enterprise network is preferred irrespective of bitrate estimates. More generally, there probably will be (operator specific) policies which define what kind of information is to be used and how it is to be used in network selection; bitrate estimate may be only a part of it.

5.3 Simulation results

In the simulations, the offered load was increased until the network becomes overloaded (meaning that a cell or an AP becomes overloaded). In the mean bitrate graphs, network has become overloaded when a graph “turns back”. In the unsatisfaction graphs overload starts when the graphs turn abruptly upwards.

In comparing different algorithms with different options the following two points should be considered:

- At what point overload starts, i.e., how much data can be sent through the whole heterogeneous network. The more an algorithm is able to use the resources in the network, the more data can be sent through the whole network.
- How well the whole network operates when not overloaded, e.g., what is the probability or the relative amount of unsatisfied users, what is the mean bitrate a user gets. (Note: Bitrate depends also on the amount of data sent due to TCP ack mechanism.)

Both mean user bitrate and unsatisfaction are displayed as functions of the datarate through the whole heterogeneous network.

In other words, X-axis in the following graphs always represents the data that is going through the whole network, including both WLAN and 3G networks. When the networks are getting loaded, the curves seem to “turn back” due to this representation.

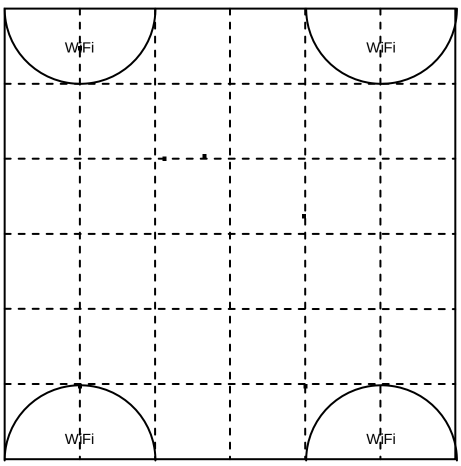
The unit used in bitrates is Mbits/s in both X- and Y-axis. For graphs representing unsatisfaction, the Y-axis shows proportion of unsatisfied users compared to number of all users (represented as a value between 0 and 1, i.e. percentage).

5.3.1 Settings

5.3.1.1 Network configuration and hotspots

In simulation cases 1 and 2 of this chapter, the network configuration has been the same as above (3.2.1).

In simulation cases 3, 4 and 5, network configuration consisted of 2 WLAN APs located in the region as follows:



Hotspots have been located centered at WLAN APs with the same area.

In all cases there is one 3G cell located in the center of the above region. Except in case 3, signal strength of the 3G cell is at maximum value (i.e. CQI = 30) in the whole region. In case 3 path loss variables have been modified so that CQI varies between 16 and 30 (depending on the distance). Note however that within WLAN coverage, variation of CQI is only between 16 and 17.

5.3.1.2 Traffic mix

In simulation case 1, traffic mix consisted of two types of users:

- FTP: 90% of users; exponentially distributed file size with mean 1 Mbyte, minimum bitrate requirement 100 kbit/s for both triggering of HOs and for breaking a connection

- YouTube: 10% of users; Weibull distributed file size with alpha=1.144 and beta=11.193 scaled by 281162 so that the mean file size is 24 Mbit, minimum bitrate requirement 500 kbit/s for both triggering of HOs and for breaking a connection

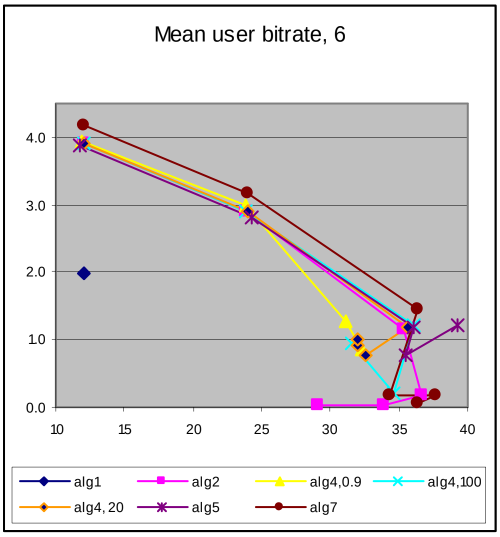
In simulation cases 2, 3 and 4 traffic mix consisted of one type: The amount of data downloaded is exponentially distributed with 1 Mbit mean. In simulation case 5 the mean was 0.1 Mbits. Bitrate requirements were 100 kbits/s.

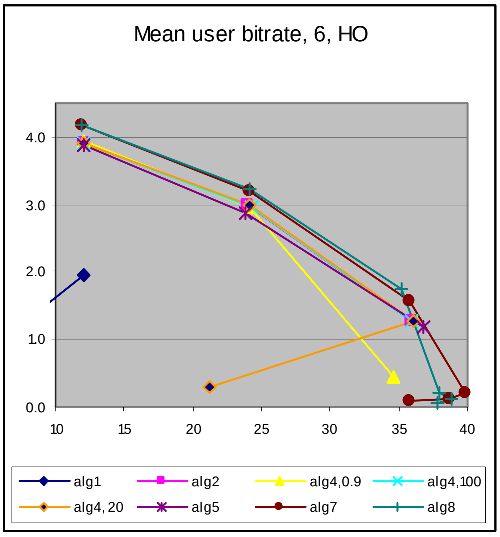
5.3.2 Case 1

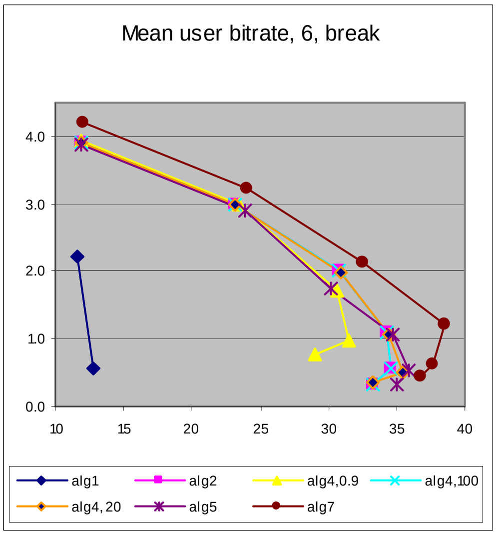
Network configuration: 4 WLAN APs, 1 3G BT

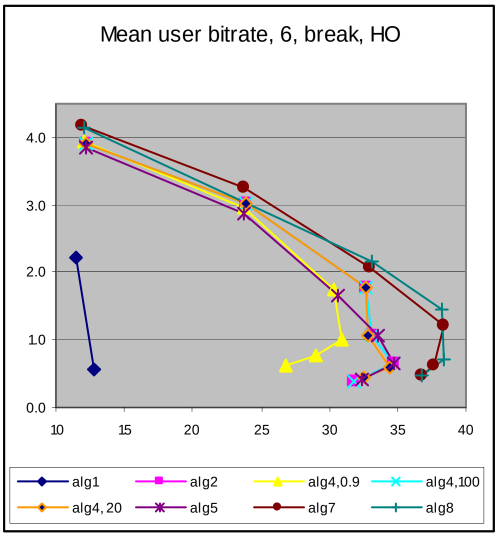
Traffic: Mixed, YouTube and 1 Mbyte

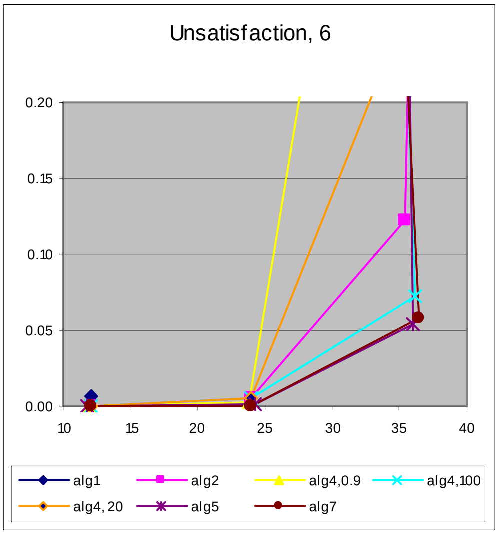
5.3.2.1 Hotspot probability 0.6

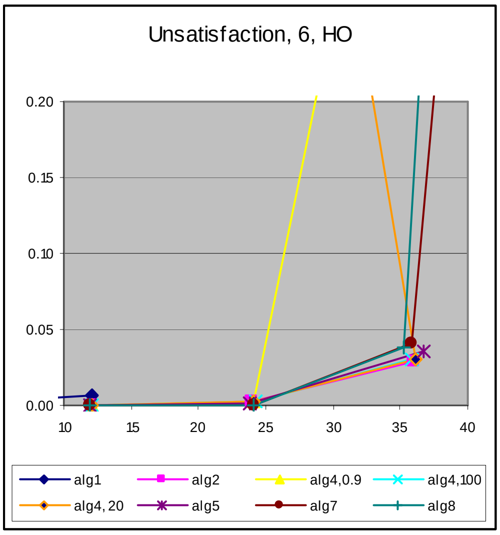


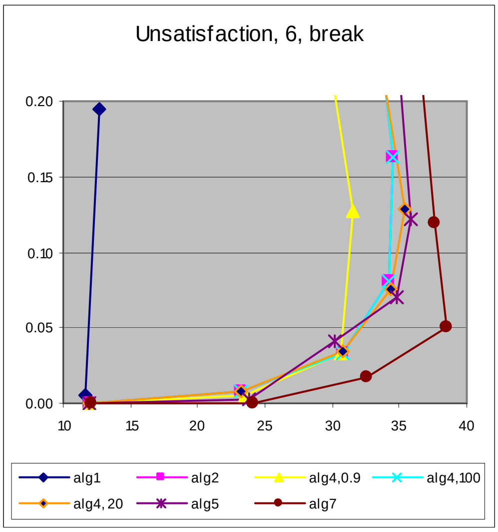


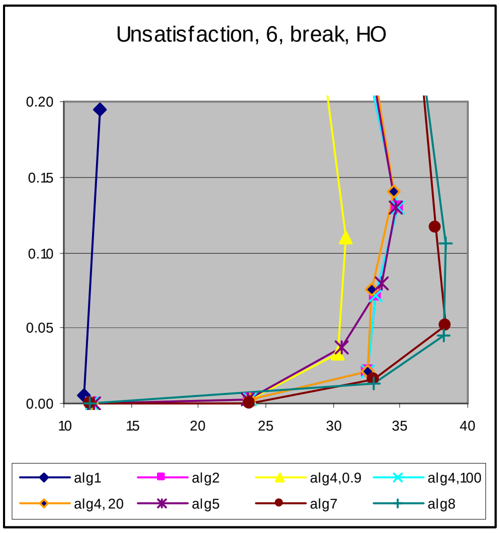

Mean user bitrate, 6, break
4.0
3.0
2.0
1.0
0.0
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg4, 20
alg5
alg7


Mean user bitrate, 6, break, HO
4.0
3.0
2.0
1.0
0.0
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg4, 20
alg5
alg7
alg8

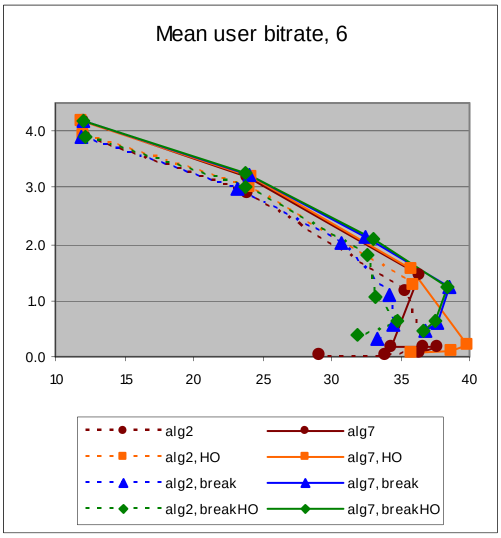

Unsatisfaction, 6
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg4, 20
alg5
alg7

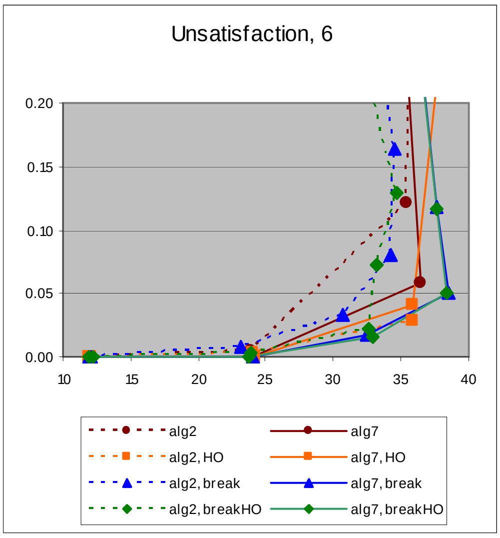

Unsatisfaction, 6, HO
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg4, 20
alg5
alg7
alg8


Unsatisfaction, 6, break
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg4, 20
alg5
alg7

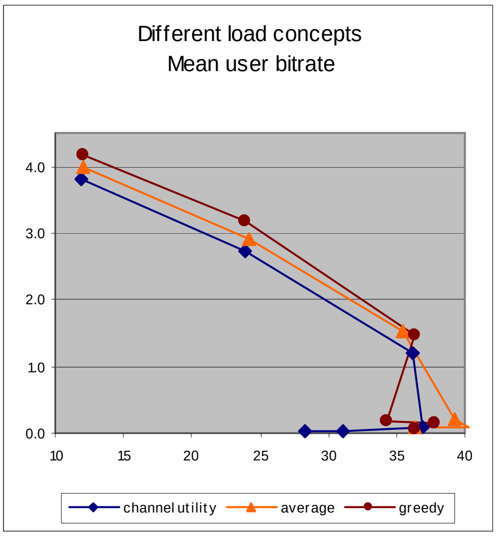

Unsatisfaction, 6, break, HO
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg4, 20
alg5
alg7
alg8

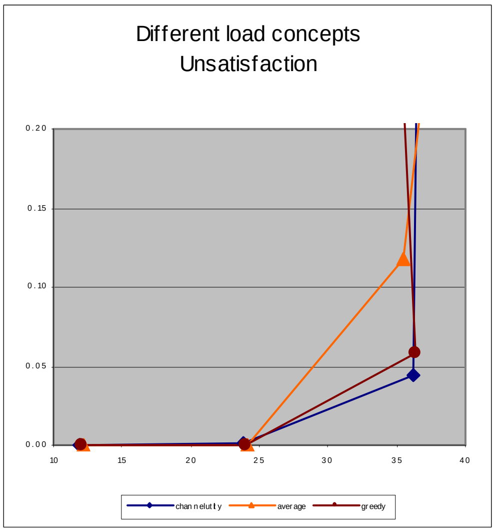
5.3.2.1.1 Algorithms 2 and 7





5.3.2.1.2 Different load concepts





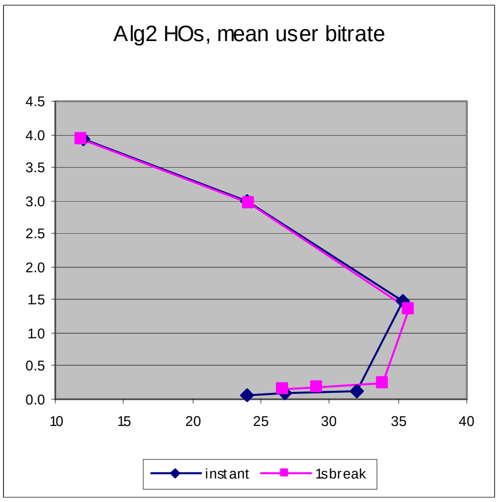
Channel utility: 10 second average of channel utility.

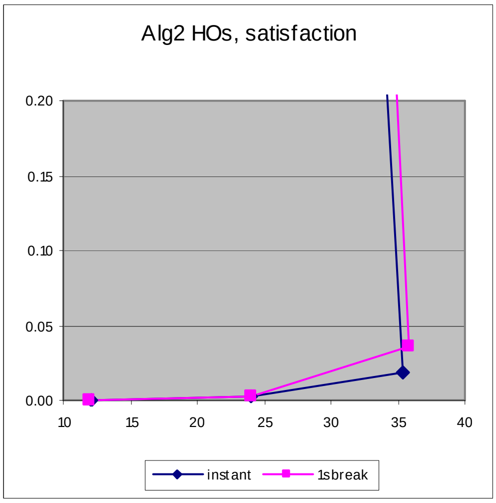
Average: Sliding average consisting of 15 samples at 1 second intervals of the number of active users.

It seems that the use of averaging decreases performance. The longer the time over which the average is taken, the more it resembles a static allocation. It does adapt to slow changes, e.g. to daily load fluctuations. However a greedy type of algorithm which uses current instead of average values can react also to short term changes.

It seems that channel utility does not give a good enough load measure. The reason for this may be that it does not take into account the shared nature of common resources, for example one connection can utilize all the available radio resources. It should be noted that long range averages of the number of active users and channel utility give the same information. However it seems that it has to be a long range average and hence channel utility is not as useful measure for non-real-time load as is the number of active users.

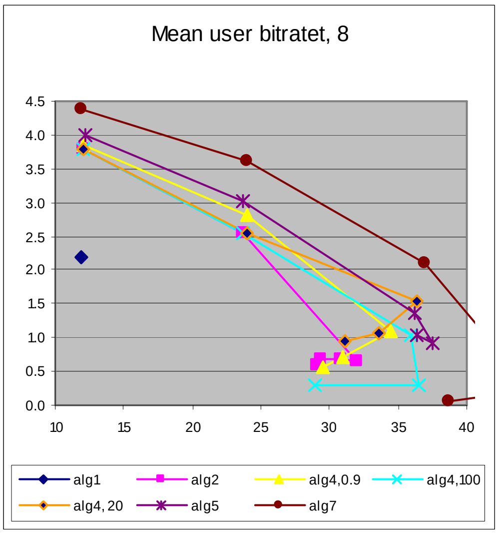
5.3.2.1.3 Different HO times

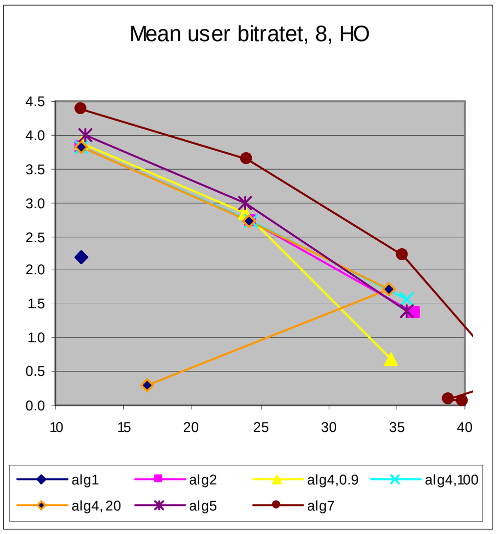


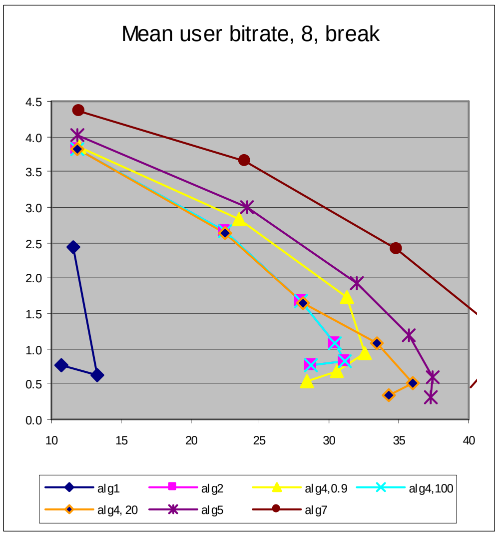


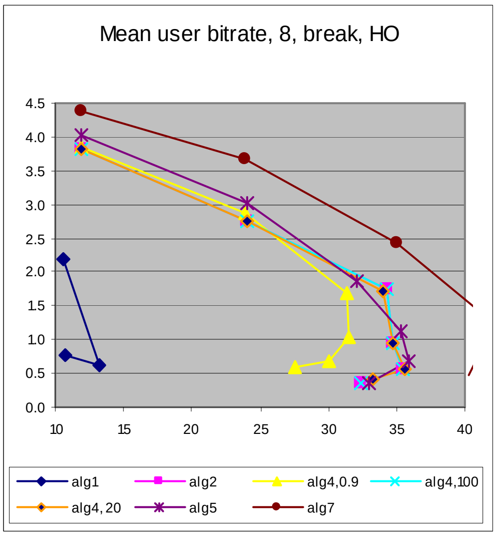
In the simulations, two different handover schemes were tried: immediate HO (break 0 s) and HO with 1 s break in data transmission. It seems that performance of HO does not depend on the length of possible break in the data transmission. The reason for this may be the following: Handovers are executed only when the load is high. In this case packets are most likely buffered at radio networks and these are lost irrespective of how long a handover takes. Having in addition a break in transmission may not worsen the situation from TCPs point of view that much. Also a handover is triggered only when the bitrate is low; the total time for a data transfer may be e.g. 10 seconds. In this case an increase of a second may not have a noticeable effect. It should also be noted that the longer the round trip time (RTT) is, the longer it takes for TCP to recover from lost packets.

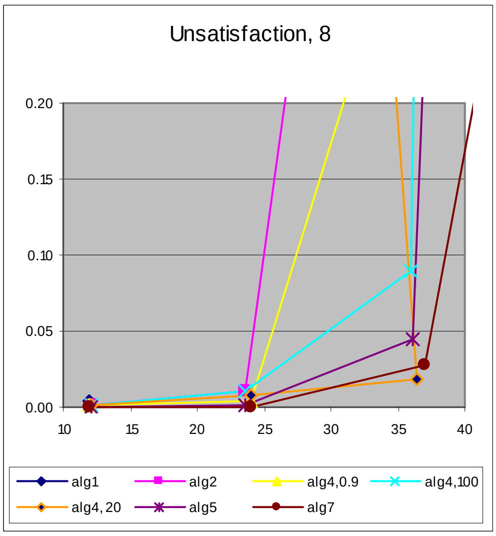
5.3.2.2 Hotspot probability 0.8

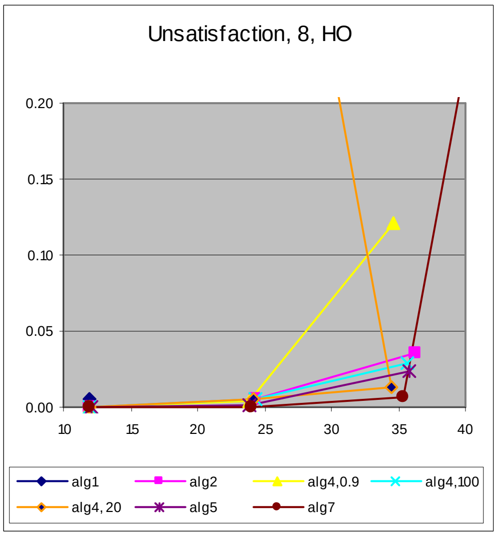

Mean user bitratet, 8
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
10 15 20 25 30 35 40
alg1 alg2 alg4,0.9 alg4,100
alg4, 20 alg5 alg7

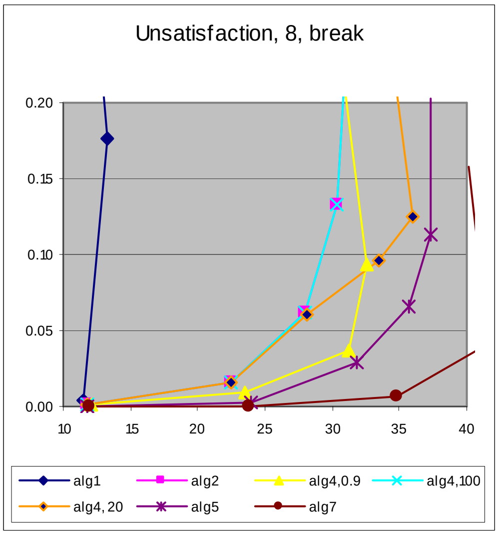

Mean user bitratet, 8, HO
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
10 15 20 25 30 35 40
alg1 alg2 alg4,0.9 alg4,100
alg4, 20 alg5 alg7

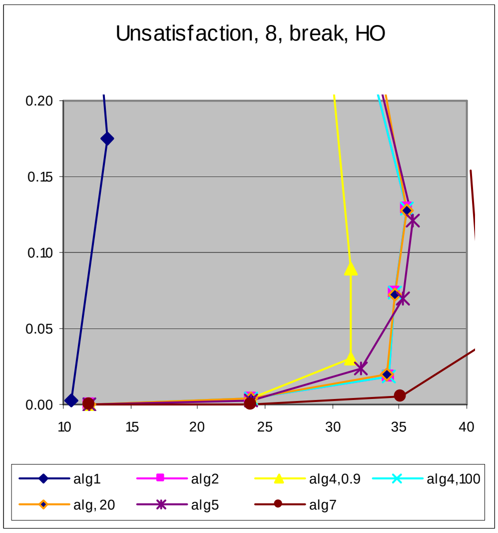

Mean user bitrate, 8, break
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
10 15 20 25 30 35 40
alg1 alg2 alg4,0.9 alg4,100
alg4, 20 alg5 alg7


Mean user bitrate, 8, break, HO
4.5
4.0
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
10 15 20 25 30 35 40
alg1 alg2 alg4,0.9 alg4,100
alg4, 20 alg5 alg7

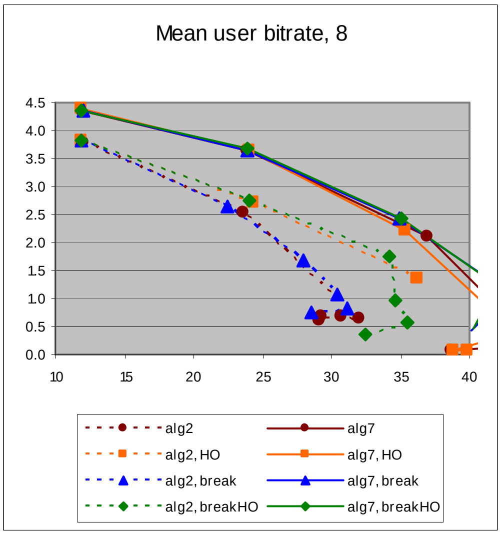

Unsatisfaction, 8
0.20
0.15
0.10
0.05
0.00
10 15 20 25 30 35 40
alg1 alg2 alg4,0.9 alg4,100
alg4, 20 alg5 alg7

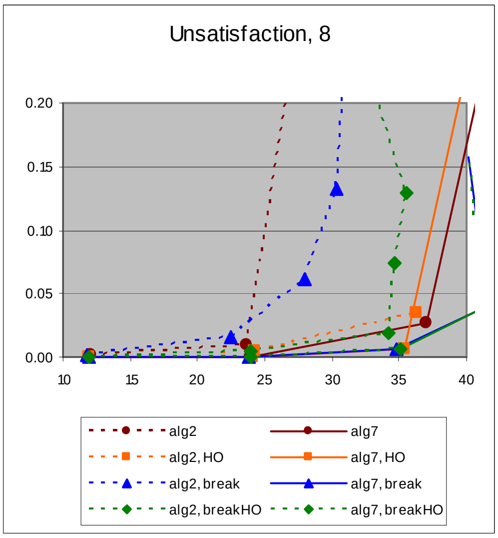

Unsatisfaction, 8, HO
0.20
0.15
0.10
0.05
0.00
10 15 20 25 30 35 40
alg1 alg2 alg4,0.9 alg4,100
alg4, 20 alg5 alg7



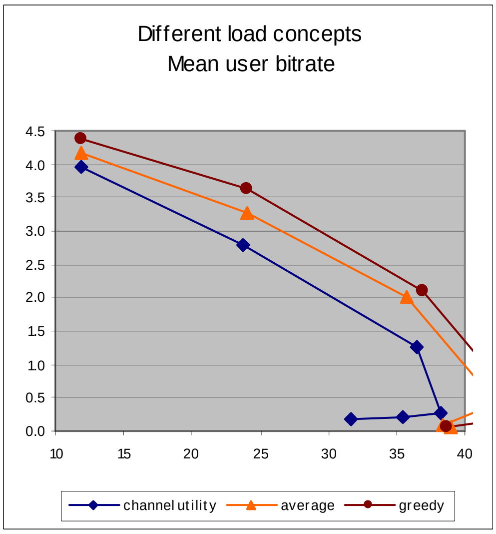


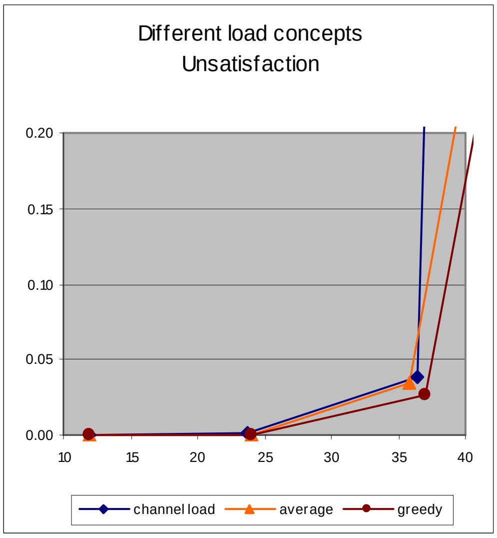
5.3.2.2.1 Algorithms 2 and 7





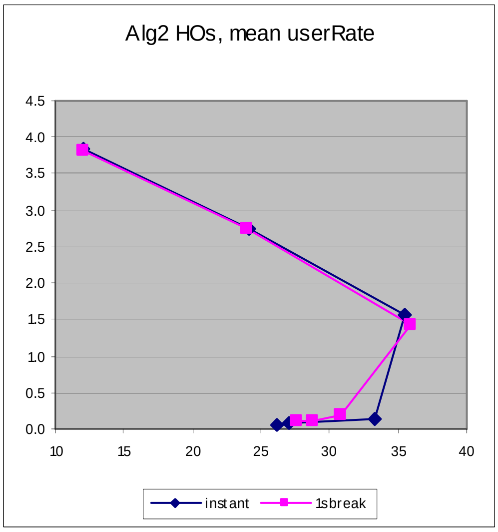
5.3.2.2.2 Different load concepts

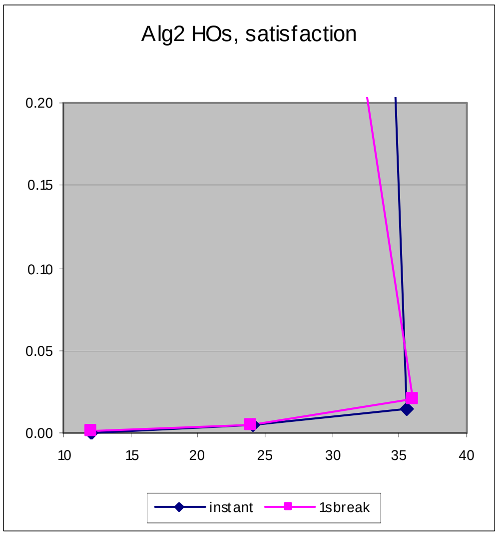




For comments on these results, see above 5.3.2.1.2

5.3.2.2.3 Different HO times





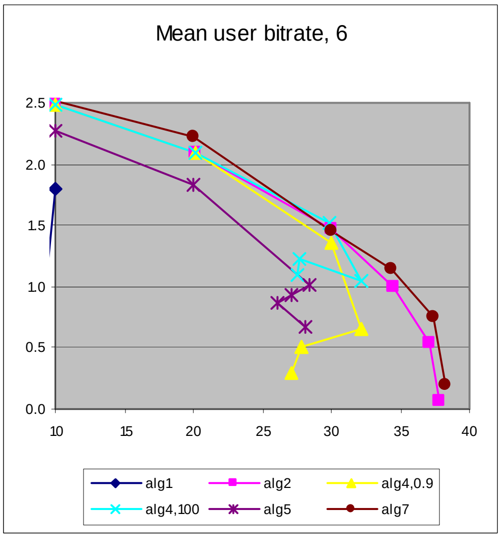
For comments on these results, see above 5.3.2.1.3

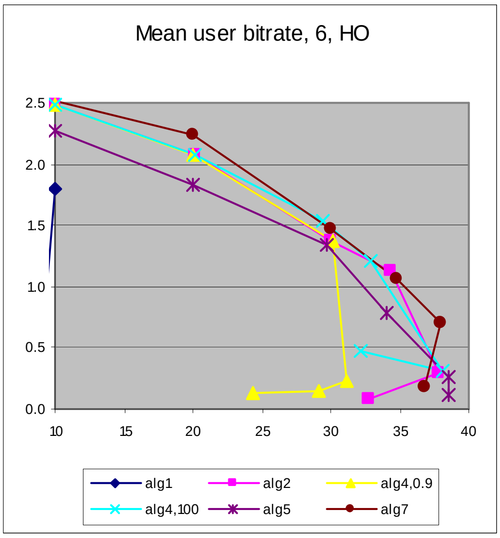
5.3.3 Case 2

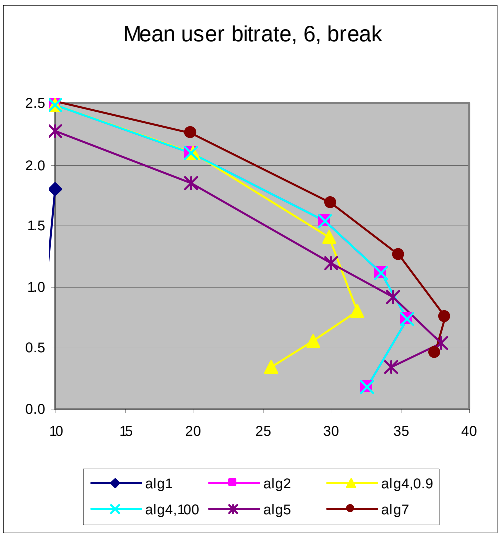
Network configuration: four WLAN APs, one 3G base station.

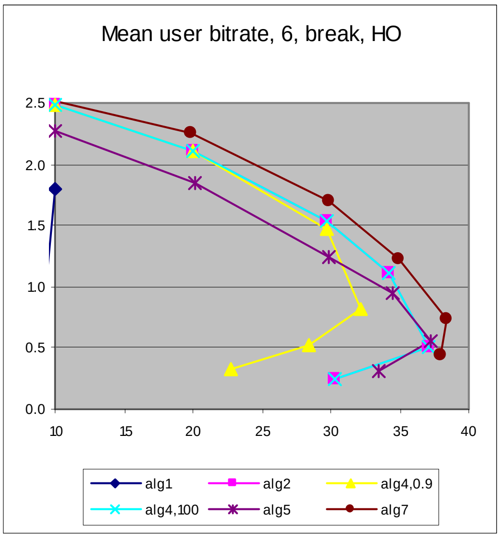
Traffic: 1 Mbit exp

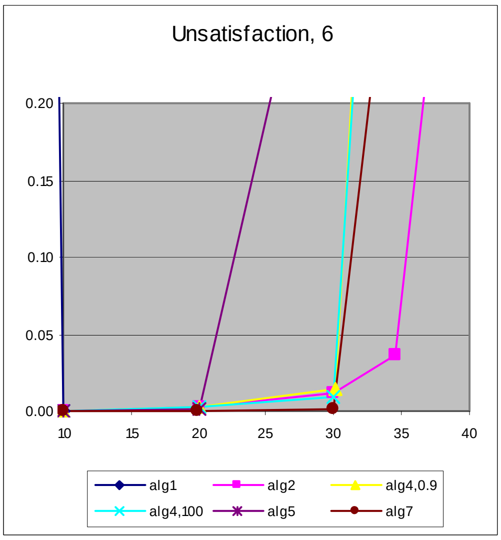
5.3.3.1 Hotspot probability 0.6

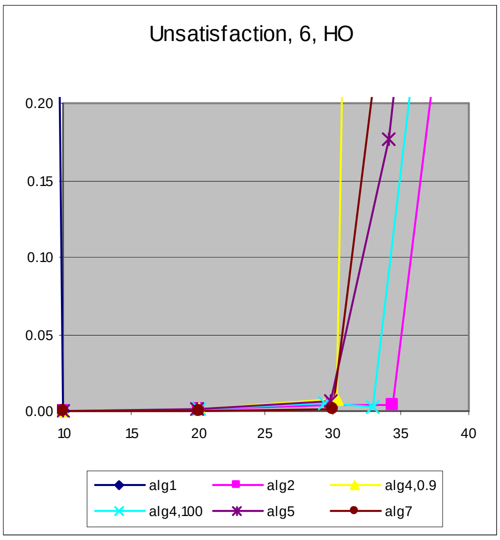

Mean user bitrate, 6
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

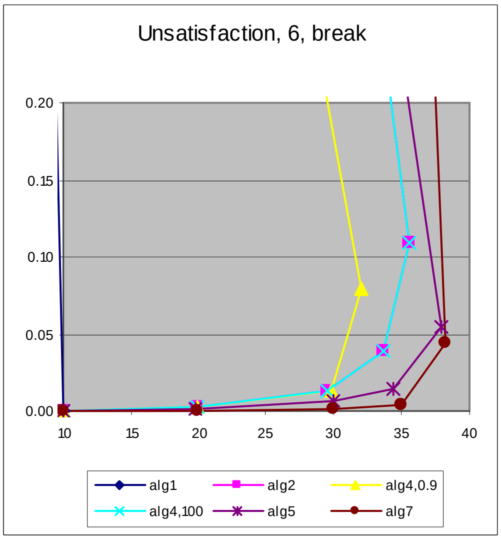

Mean user bitrate, 6, HO
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

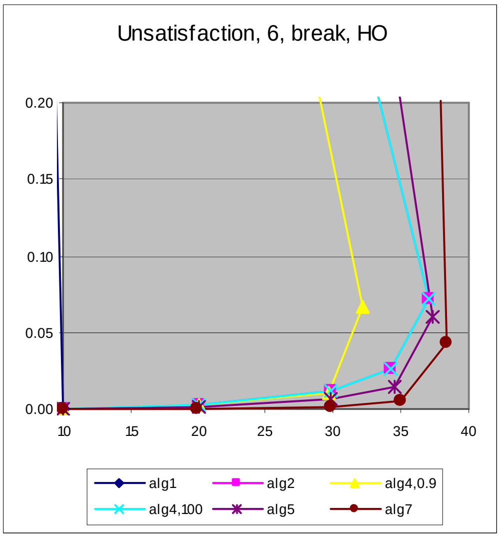

Mean user bitrate, 6, break
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7


Mean user bitrate, 6, break, HO
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

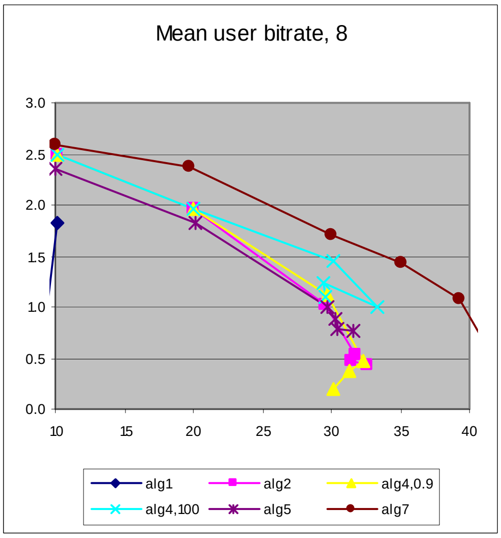

Unsatisfaction, 6
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

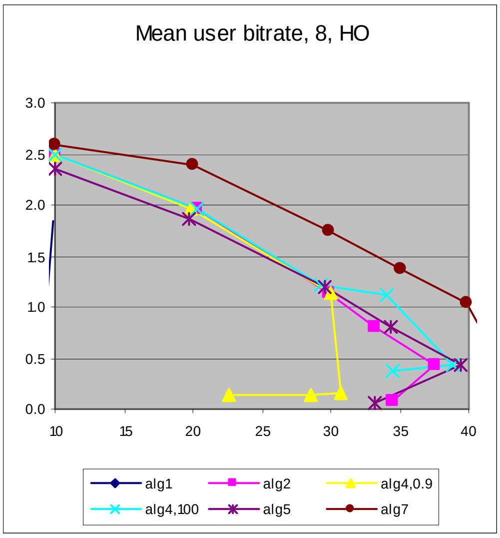

Unsatisfaction, 6, HO
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
35
40
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

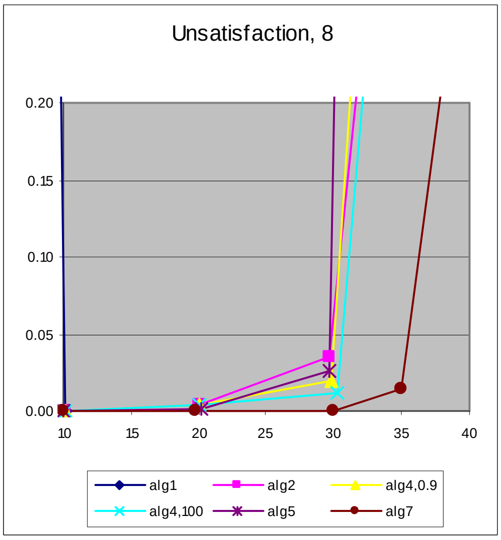


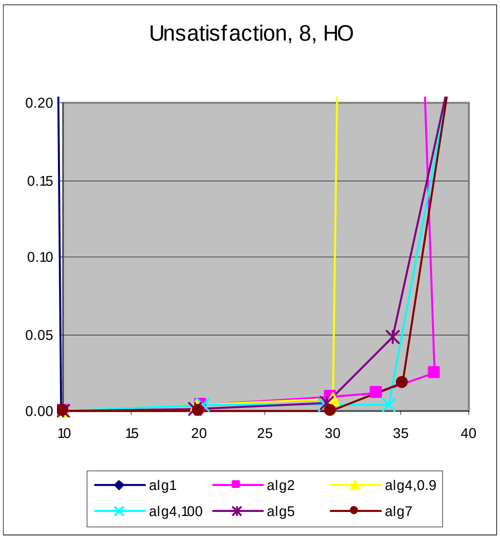


5.3.3.2 Hotspot probability 0.8

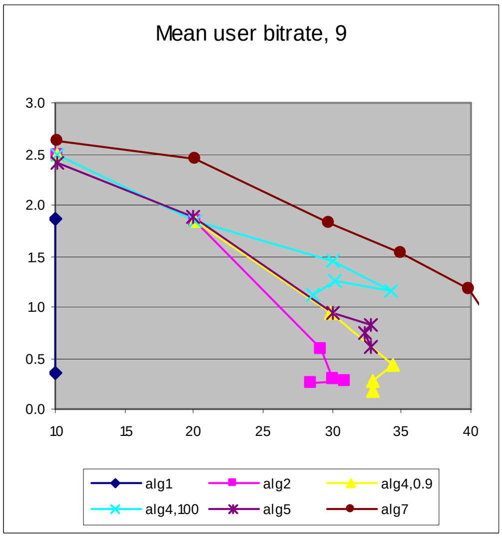


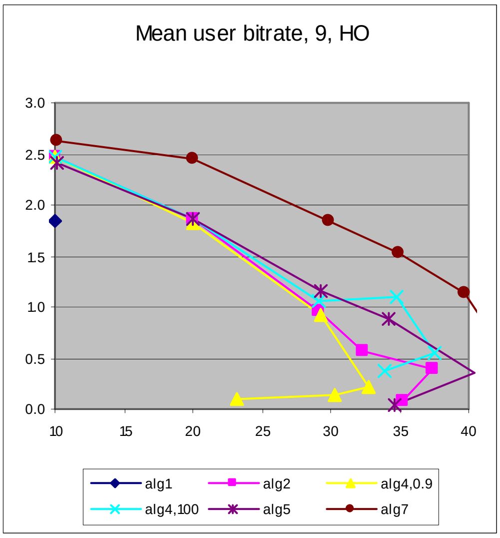


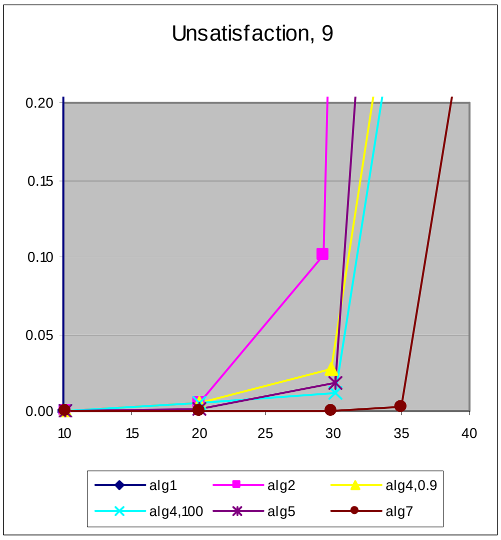


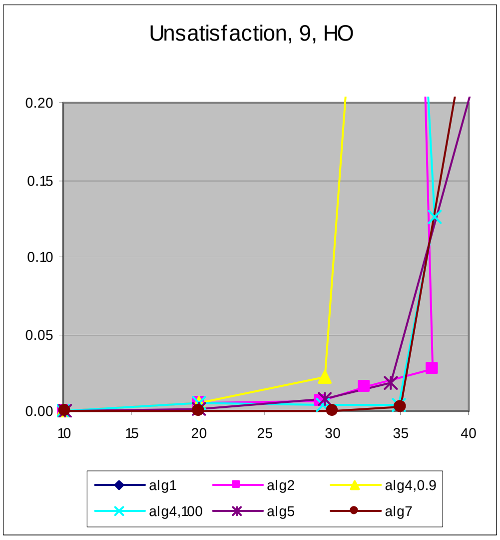


5.3.3.3 Hotspot probability 0.9







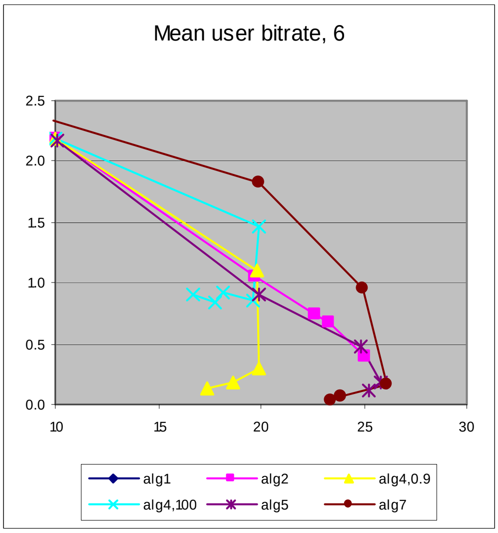


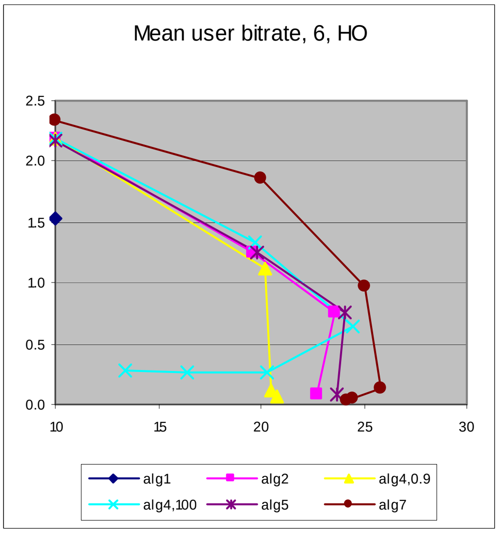
5.3.4 Case 3

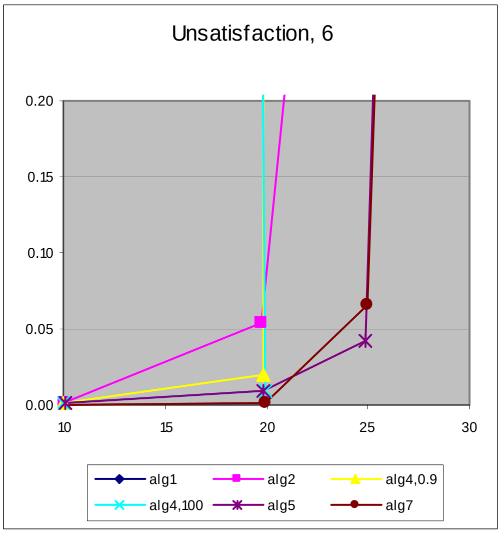
Network configuration: two WLAN APs, one 3G base station, weak 3G.

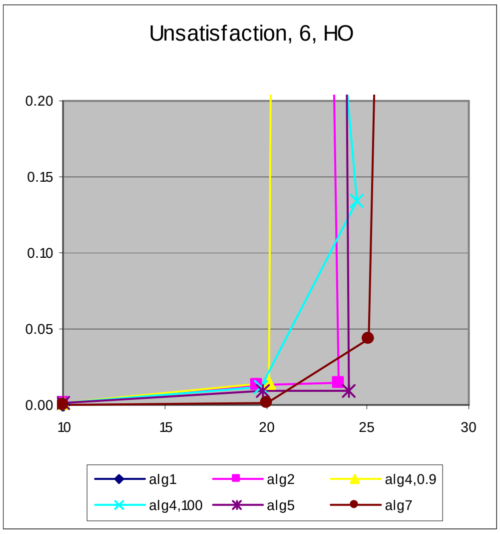
Traffic: 1 Mbit exp

5.3.4.1 Hotspot probability 0.6







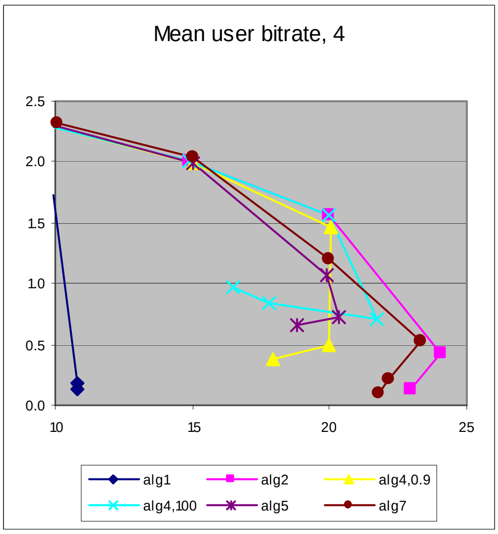


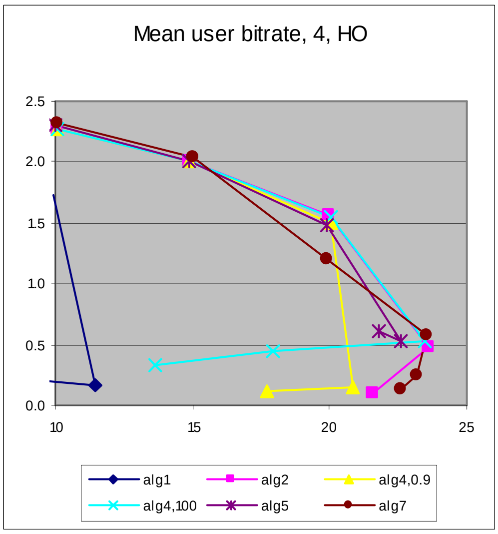
5.3.5 Case 4

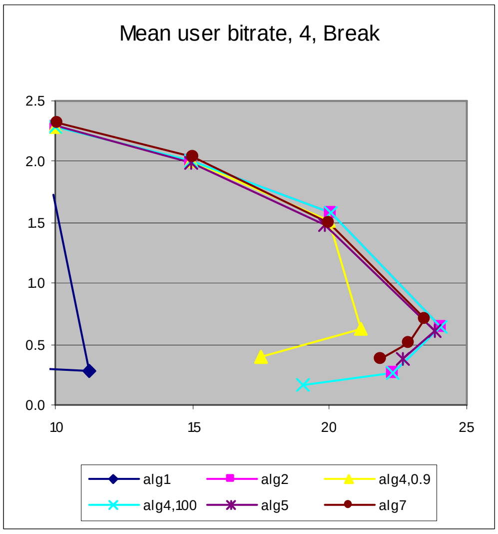
Network configuration: two WLAN APs, one 3G base station.

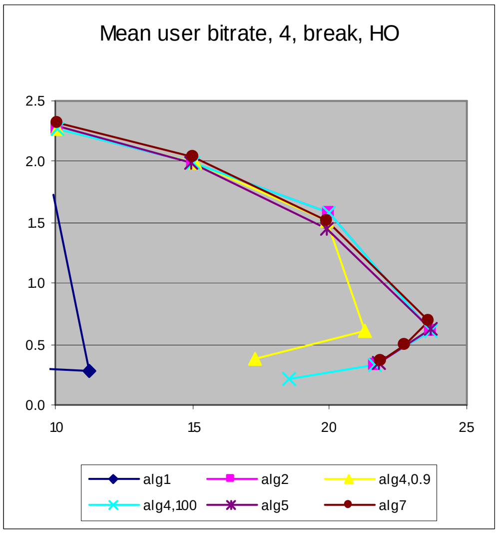
Traffic: 1 Mbit exp

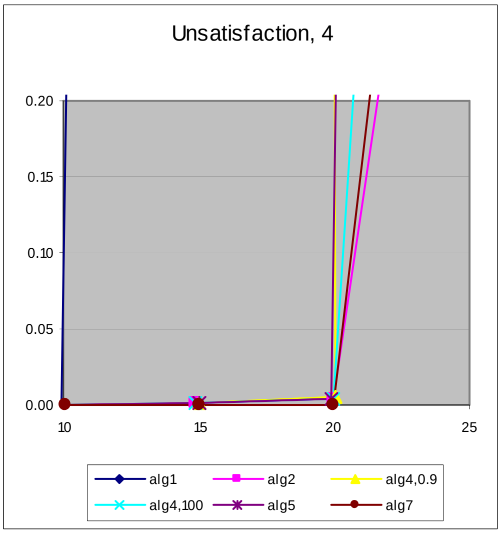
5.3.5.1 Hotspot probability 0.4

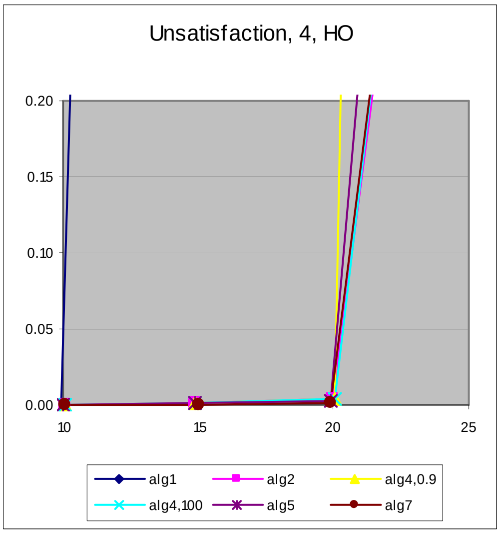


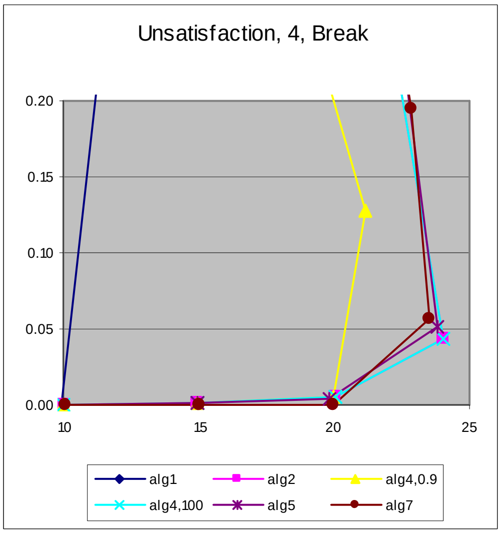


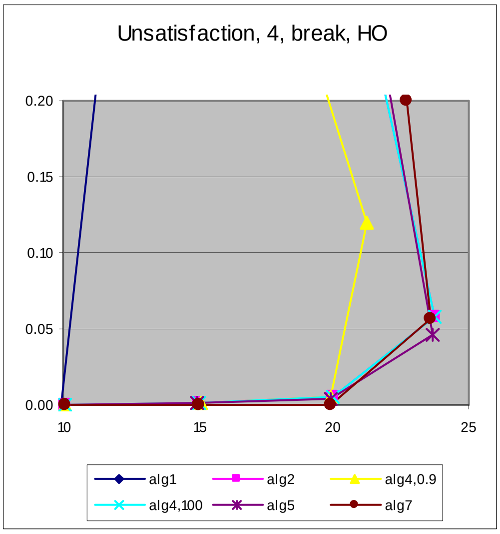

Mean user bitrate, 4, Break
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7


Mean user bitrate, 4, break, HO
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

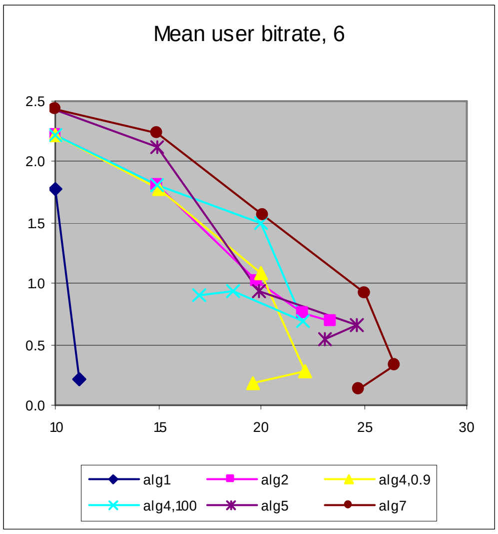

Unsatisfaction, 4
0.20
0.15
0.10
0.05
0.00
10
15
20
25
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

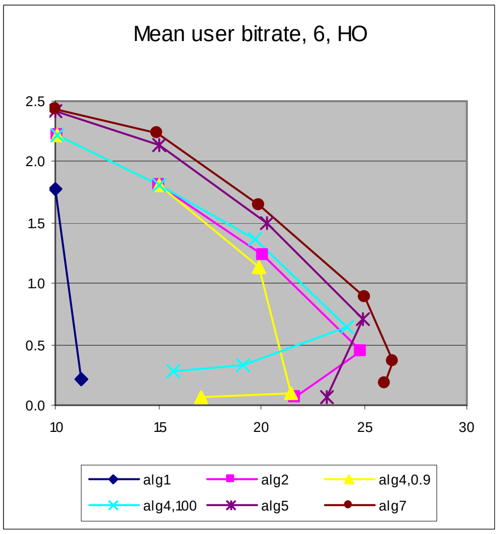

Unsatisfaction, 4, HO
0.20
0.15
0.10
0.05
0.00
10
15
20
25
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

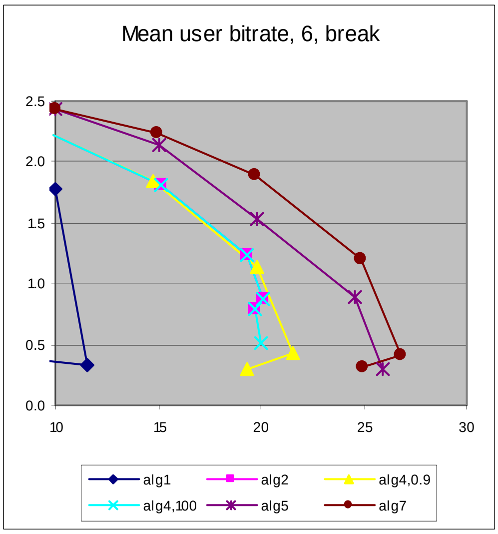

Unsatisfaction, 4, Break
0.20
0.15
0.10
0.05
0.00
10
15
20
25
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

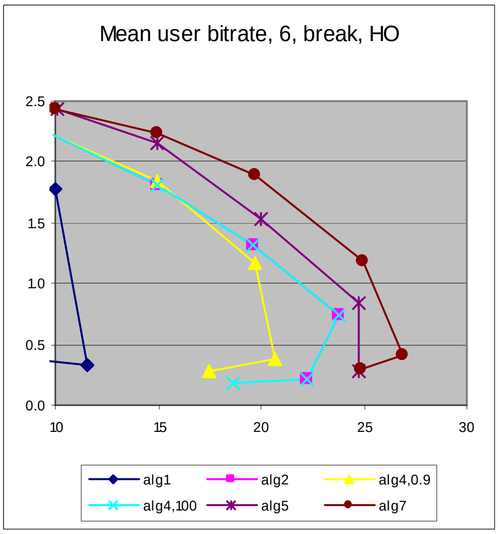

Unsatisfaction, 4, break, HO
0.20
0.15
0.10
0.05
0.00
10
15
20
25
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

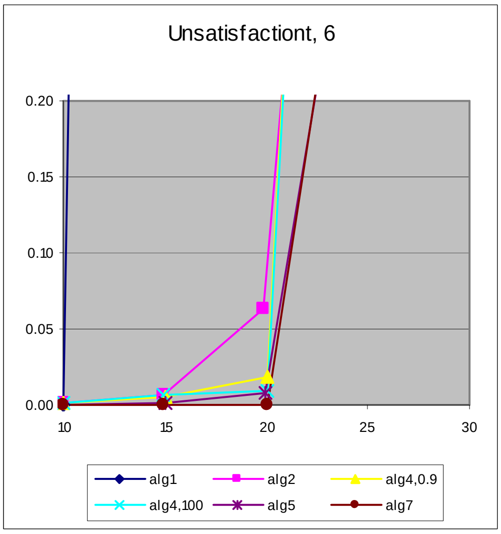
5.3.5.2 Hotspot probability 0.6

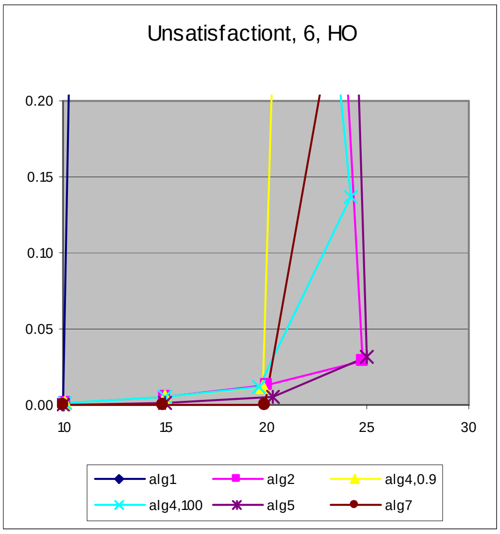


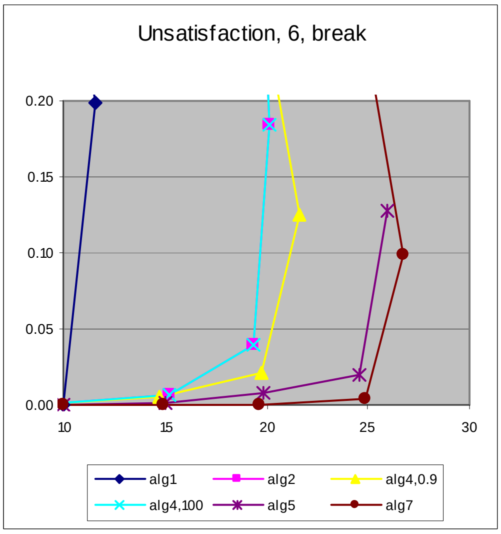


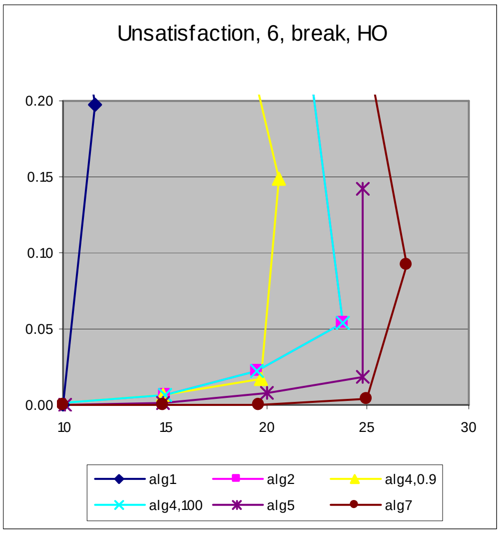




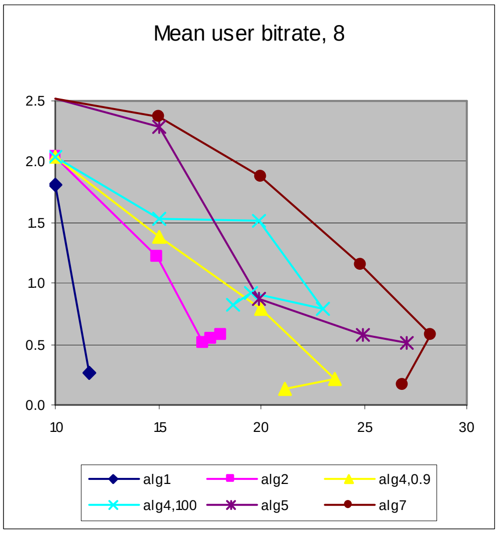


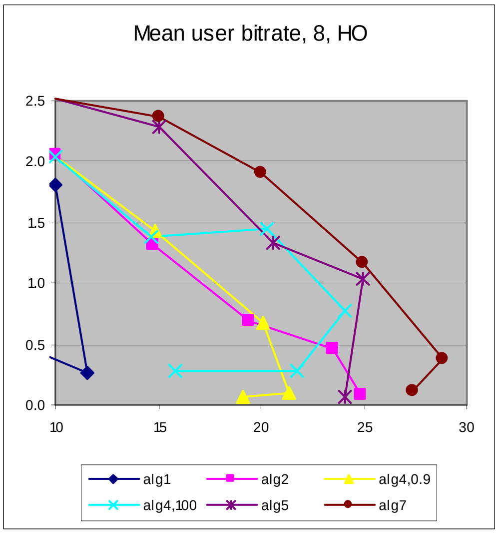


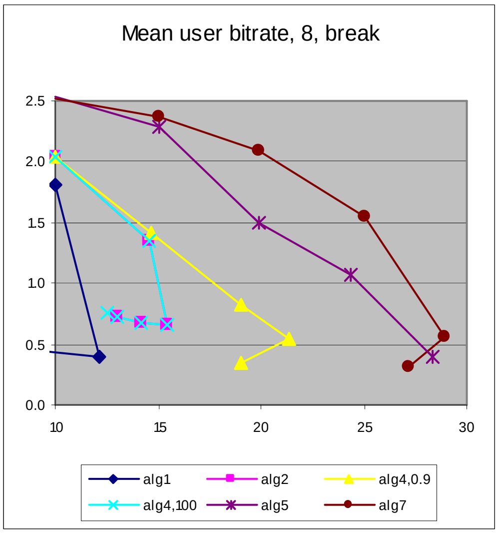


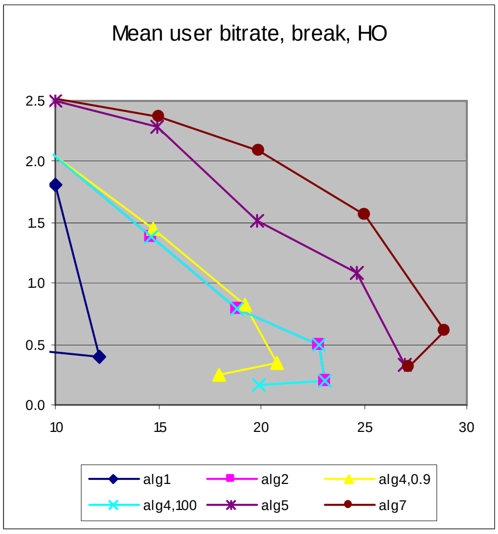


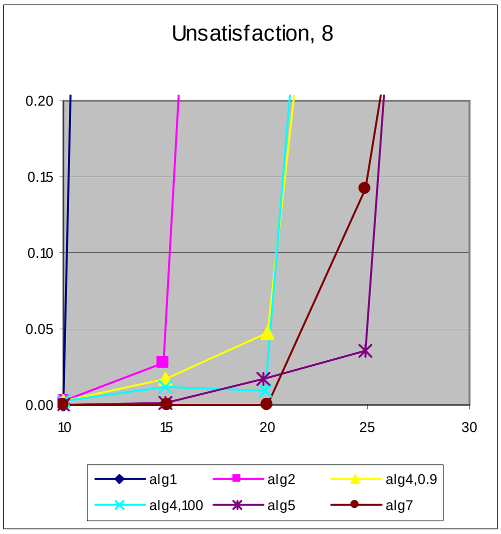
5.3.5.3 Hotspot probability 0.8

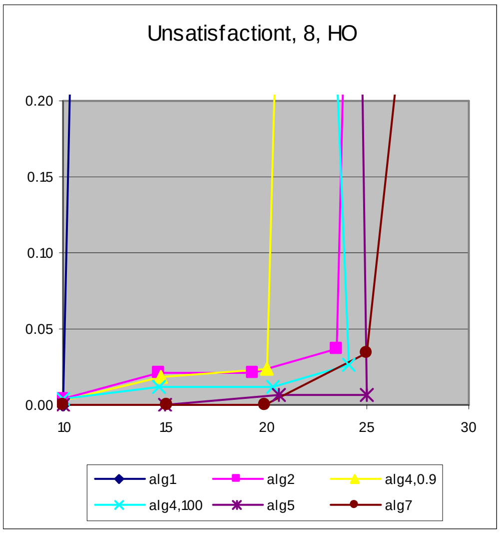

Mean user bitrate, 8
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

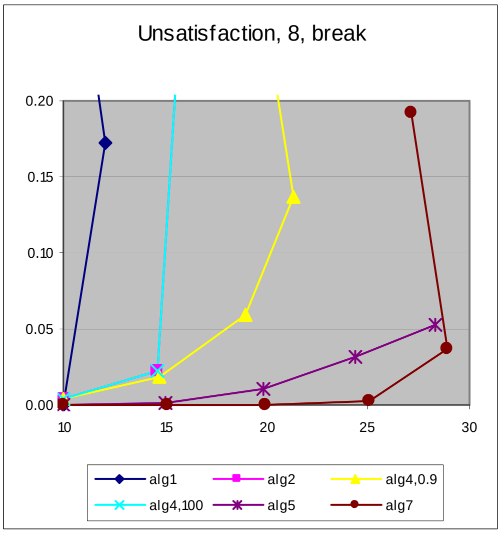

Mean user bitrate, 8, HO
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7

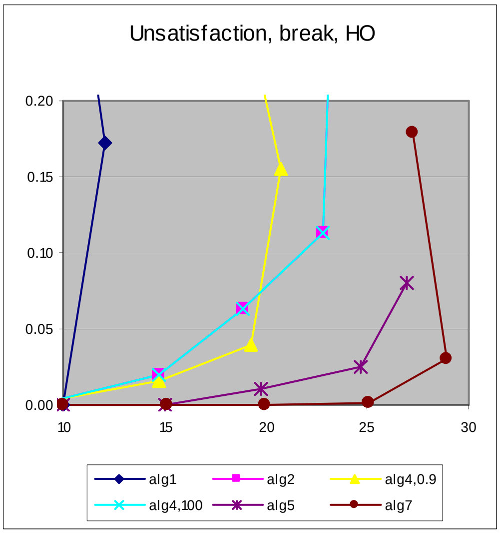

Mean user bitrate, 8, break
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7


Mean user bitrate, break, HO
2.5
2.0
1.5
1.0
0.5
0.0
10
15
20
25
30
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7


Unsatisfaction, 8
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7


Unsatisfactiont, 8, HO
0.20
0.15
0.10
0.05
0.00
10
15
20
25
30
alg1
alg2
alg4,0.9
alg4,100
alg5
alg7



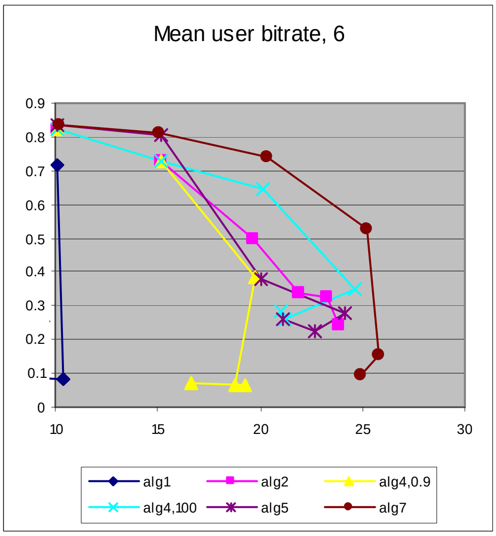


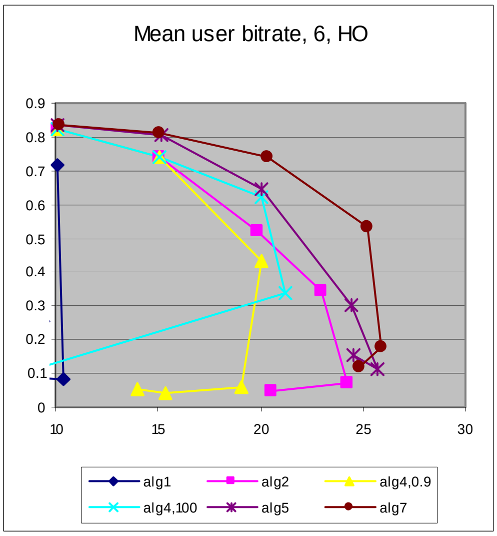
5.3.6 Case 5

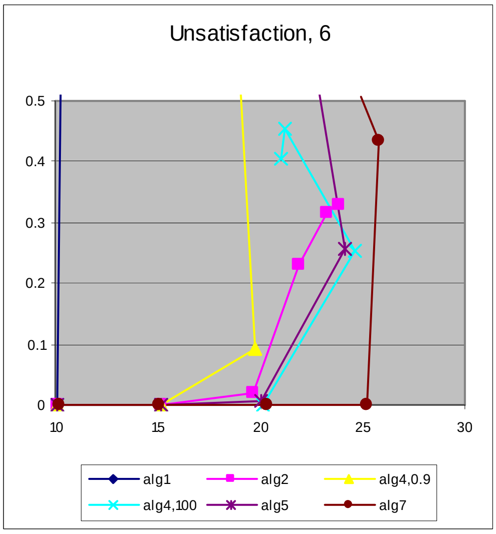
Network configuration: two WLAN APs, one 3G base station

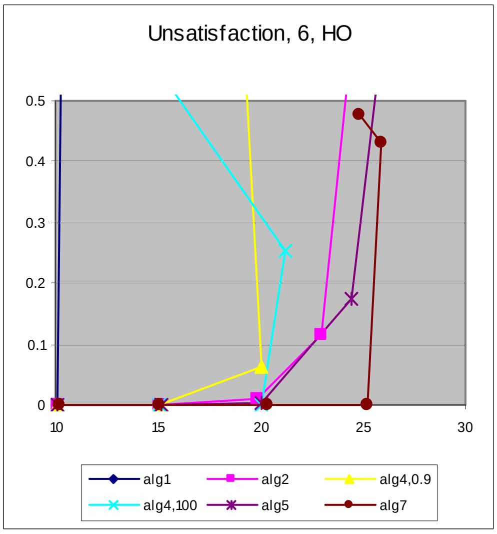
Traffic: 0.1 Mbit exp

5.3.6.1 Hotspot probability 0.6

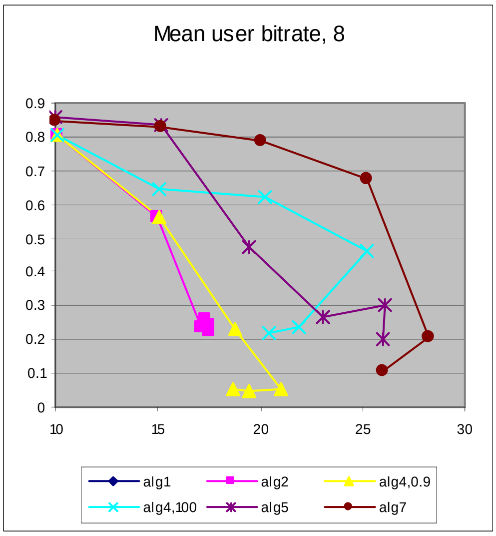


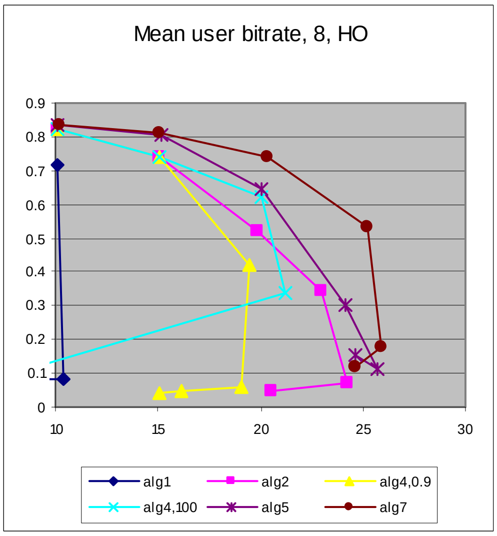


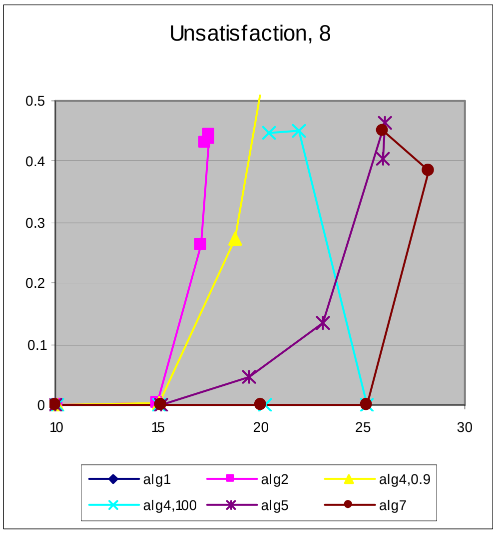


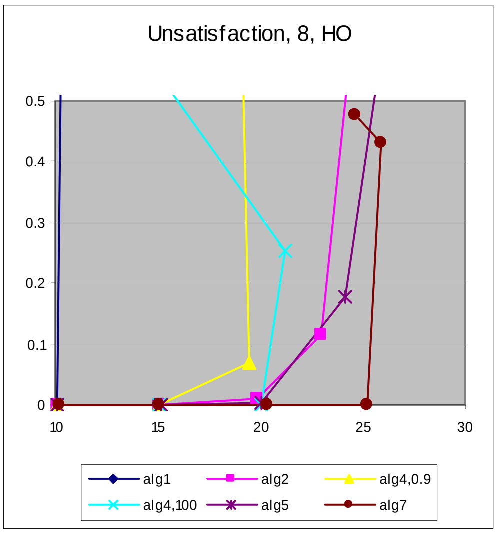


5.3.6.2 Hotspot probability 0.8









6. References

[1] Policy-Based Resource Management, Future Internet Programme, Activity 2.4 Deliverable DA42.4.1, Janne Tervonen (ed.)

[2] 802.11k-2008 Radio Resource Measurement of Wireless LANs, IEEE, June 2008

[3] IKR Simulation Library V2.6, http://www.ikr.uni-stuttgart.de/INDSimLib/.

7. Acknowledgements

This work was supported by TEKES as part of the Future Internet program of TIVIT (Finnish Strategic Centre for Science, Technology and Innovation in the field of ICT).