Master Thesis

User Satisfaction Oriented Vertical Handover Decision Enhancement by means of Dynamic Network Information

ユーザ満足度向上を目指した動的ネットワーク情報を用いた垂直ハンドオーバー決定の改善

Thesis Supervisor

Professor Hitoshi Aida

The University of Tokyo
Graduate School of Engineering
Department of Electrical Engineering and Information Systems

Amani AbuQdais

37-085951

August 18th , 2010

ABSTRACT

Next generation wireless networks will involve many different Radio Access Technologies (RAT) forming the so called Heterogeneous Networks. Devices equipped with different Radio Access Technology Network Interface Cards (NICs) are capable of performing handovers across these Heterogeneous Networks in a process called the Vertical Handover.

In Vertical Handovers, Network Information is a very important factor in the target network decision process and cannot be obtained using traditional network discovery methods.

One of the ways to provide network information is through Information Servers as a location-based service. However, the server-based method only provides static information about networks. If users perform their handover decisions based on such static information, the handover target might not be as "good" as predicted.

In this thesis, a method to provide dynamic network information, especially information about the current network load, is proposed in order to enable smarter handover decisions and increase the "satisfaction" level of users. In this method, user devices are used as Information Agents to report network information to an information server.

It is shown by means of simulation that the user satisfaction level can be increased to a higher level compared with methods that only use static information. An evaluation of the system performance is also provided showing how the proposed method not only elevate the user satisfaction but also result in a less handover blocking rate.

ACKNOWLEDGEMENTS

Work on this thesis would not have been possible without the encouragement and support from many people.

First of all I would like to express my deepest gratitude and appreciation for my supervisor, Professor Hitoshi Aida, for his continuous guidance, support and for always being an unlimited source of ideas and knowledge. I was fortunate to be under his supervision for two years.

I would like to thank the lab secretary, Mrs. Sayuri Nakayama, who besides her work as a secretary, was always there to give us, students, her warm encouraging words and advice in rough times which made the lab life feel more comfortable.

Many thanks to the technical staff member, Mr. Shingo Chiba, for his continuous technical support.

I would also like to thank all the Aida lab members with whom I shared many tough and fun times for 2 years of my life.

All appreciation for the Japanese Government for giving me this extremely precious chance of continuing my study in Japan for 2.5 years under the most comfortable conditions.

And last but not least, I would like to express my deepest gratitude to my wonderful parents and family who always provided me with everything necessary to live a comfortable life and for all their support.

August 18th, 2010

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

1

1.1 Background and Problem Definition

Due to the recent advances in wireless communications, a lot of different wireless technologies have emerged creating network environments where different types of Radio Access Technologies (RATs) can coexist. Future fourth generation (4G) systems are considered to be heterogeneous networks composed from multiple differing RATs as well as more capable user terminals that are equipped with multiple Network Interface Cards (NICs) allowing users to connect to and handover between networks of different RATs in a process called the “Vertical Handover”. For example a device might have both WiFi and WiMAX NICs allowing it to connect and handover to either of the two network technologies wherever available and whenever needed.

Compared to traditional handovers, which are mainly based on the received radio signal levels, vertical handovers are more complex and require taking more parameters and criteria into consideration when deciding the need for handovers and selecting the target network. This is mainly due to the large differences in network characteristics of different RAT networks. Thus traditional handover mechanisms are no longer sufficient to meet the requirements of future networks and if such traditional methods were going to be applied in the case of vertical handovers, it will cost user devices high power consumption and long delays to perform the handover process. This is mainly because heterogeneous devices have multiple NICs and turning them all will be very energy consuming.

In the case of vertical handovers, devices need more information about surrounding networks as the Received Signal Strength indicator alone is not enough in this case.

Moreover, if devices could obtain more detailed information about the networks available in its current location, such as bandwidth, used radio channels and cost, it will be able to perform handovers in a shorter time, with less power consumption and can choose the most suitable network for the user’s current situation. This will also enable it to predict potential handovers before a need for one actually occurs. In this thesis, the term “smart handovers” is used to refer to these handovers that are situation-aware.

1.2 Objective and Scope of this thesis

The final goal of this work is to enable users to choose the most suitable network for the handover target and thus increasing the satisfaction level of these users with less power consumption and delay. This is achieved by allowing user devices to inquire about available networks in their current location and obtain information about these networks. Then they can utilize this information to evaluate the different network options (handover target candidates) and choose the most suitable candidate as the handover target.

Any type of handover can be divided into three stages, namely, Network Discovery, Handover Decision and Handover Execution. Research efforts on the field of enhancing handover in heterogeneous networks can be classified by the stage of handover they consider.

This thesis is mostly concerned with the first two steps of the handover process. The proposed method helps in reducing the delay and energy consumption contributed by the initial stage. It also helps in achieving smarter decisions for handovers. The remaining stage, Handover Execution, with the signaling involved with it is out of the scope of this thesis.

1.3 The solution approach

Network information is offered to devices by means of network information servers that manage network information on a regional basis. Moreover, these information servers are enhanced by making them capable of providing dynamic network information, which is a point that is still not yet adequately considered in the context of heterogeneous networks.

Network information servers will hold, in addition to the static network information, dynamic information about the load of each network. Such dynamic information is collected by allowing devices to work as agents and enabling them to send periodic updates to the information servers in which they report the current conditions of the network they are currently using. This way, the proposed system can be thought of as a third party service that can provide information about networks regardless of the network operators.

To divide the burden of sending updates among the agent devices, servers will choose a different device at each time an update was needed. Also, this rate at which devices send updates to the server is variable over time, meaning it will be adaptive to the network activity.

To enable such network information system to be applicable on a global scale, we suggest that information servers be organized in a hierarchical manner while each server in the lowest level is responsible of managing network information about networks in a certain geographical area.

.

1.4 Chapter Overview

The rest of this thesis is organized as follows. In chapter 2, the issues related to handovers are discussed while pointing out the difficulties faced in the case of vertical handovers. Also, recent research efforts trying to tackle the problems of vertical handovers are reviewed.

In chapter 3, the proposed system is explained in detail and the method of dynamic information collection on the server side is presented.

In chapter 4, the simulation experiments used to evaluate the proposed system and their results are presented. The chapter is finished by a discussion of these results.

Finally, this thesis is concluded in chapter 5 and future directions are stated.

Chapter 2

Vertical Handover Issues and Related Research

In this chapter the issues related to vertical handovers are discussed in more depth and recent research efforts that tried to tackle these problems are presented.

2.1 The Importance of Vertical Handovers in Future Networks

As was stated in the introduction, vertical handovers refer to handovers that take place between two networks of different radio technologies. Devices that are able to perform such handovers need to have multiple Radio Interface Cards, each for one type of the radio technologies. Fig.2.1 shows an example on such a device. Fig.2.2 shows 3 devices performing different types of handovers. Device1 is performing a traditional WiFi-to-WiFi handover while devices 2 and 3 are performing vertical handovers (WiFi to WiMAX and vice versa).

Fig. 2.1 A device that has multiple Network Interface Cards (NICs)

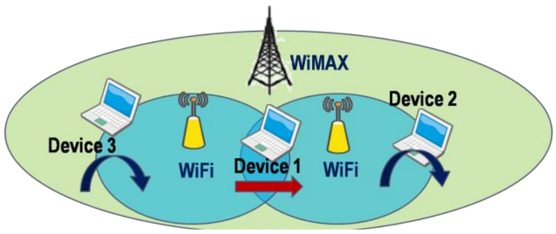

Fig. 2.2 Three devices performing different types of Handovers

In traditional networks, handovers mostly occur for mobility reasons, i.e., when devices move out of the coverage area of the currently connected network, it needs to transfer its connection to a new access point or base station to preserve its connectivity to the network. If devices are able to use networks of many different radio technologies, as the device in Fig.2.1, handovers can occur for other reasons, such as

QoS or performance reasons. This means that the device will perform handovers to the networks that are "better" from the point of view of the user. Some users might prefer networks with higher data rates, others can cope with low data rate networks as long as the service is provided with low costs.

Heterogeneous networks consist of multiple radio access networks of different technologies. These technologies are different on the physical layer and the data link layer (layers 1 and 2) of the OSI layered model.

The motivation for having heterogeneous networks comes from the fact that no single radio technology on its own can satisfy all the requirements of today's users among the currently available technologies [1], which makes heterogeneity a requirement rather than an exception in the next generation networks. Instead of replacing the current systems with a new uniform standard for all wireless communication systems, the trend is to integrate the various available wireless technologies to form the future systems[2]. Table 2.1 shows a number of radio technologies and their different characteristics.

Also, having a heterogeneous network environment gives many advantages for the network users. This is because users can select the most appropriate network that best fits their needs. For example, if the user was looking for a network that provides higher data rates, the device can prefer that network and perform a handover to switch its connection to the higher data rate network, as in Fig. 2.3.

Technology	Coverage	Data Rate	Mobility
UMTS	20Km	~2Mbps	High
WiMAX (IEEE802.16a)	Around 30Km	Max 70Mbps	Medium
WiFi (IEEE802.11a)	50~300m	54Mbps	Low
WiFi (IEEE802.11b)	50~300m	11Mbps	Low
GSM/GPRS	35Km	~144kbps	High

Table 2.1 Different wireless technologies with their characteristics

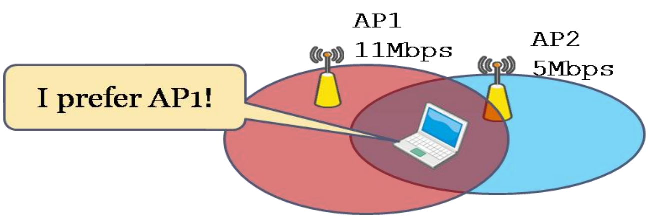

Fig. 2.3 A handover that occurs for "preferences" reasons rather than mobility reasons. In this example, the user prefers higher data rate networks

As a result, vertical handovers will occur more often in future networks and reducing the problems and the inefficiencies associated with them becomes a vital issue.

2.2 Handover Stages

There are generally three main steps in the handover process: *Discovering* candidate networks, *Deciding* a handover and *Executing* the handover, as shown in Fig.2.4.

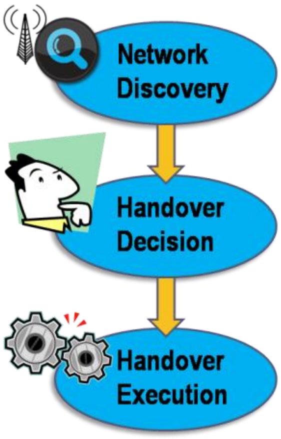

Fig.2.4 The stages involved in any handover process

Following is a discussion of these different stages and of the challenges faced at each stage.

2.2.1 Candidate Network Discovery Stage

Discovering currently available radio access networks is one of the most challenging issues in the heterogeneous wireless network environment [3]. In the discovery step, a device looks for potential target networks to handover to.

Traditionally, the discovery of candidate networks is done by turning on the NIC and listening to broadcast frames that announce the existence of such networks.

For example, in WiFi networks, this is merely done by listening to beacons of available APs (Access Points) and constructing a list of available networks. This listening based network discovery is inefficient in the case of a heterogeneous environment. This is due to the fact that heterogeneous devices are equipped with many NICs; one for each supported radio technology. If the listening based discovery was to be used in heterogeneous networks, the device would have to turn on its NICs one by one and try to discover available networks by scanning many radio channels on each NIC. This has three main disadvantages.

Disadvantages of using traditional discovery methods in Heterogeneous Networks

First, turning on all NICs for networks discovery is very power consuming, where power consumption is a very important issue for mobile devices.

Second, it is time consuming to wait until the scan on all NICs is performed and then decide on the target networks. This time will also translate to power consumption. Fig. 2.5 shows a comparison between the discovery time for a single NIC device versus a heterogeneous device in case the traditional candidate network discovery method was used. Handover latency needs to be smallest in order to ensure no disconnections or packet losses occur.

Third, listening based network discovery provides us with networks that are only available in the current location of the device and with very small information about each network that doesn't exceed the needed connection parameters in most cases.

In order to perform smart handovers, more network information is needed. Also, for highly mobile devices, information about networks that are located ahead of the device can help in better handover decisions and possibly to an ahead of time prediction of the need of handovers.

For the three reasons mentioned above, research is heading towards finding more efficient alternatives for network discovery methods in heterogeneous networks, as will be explained in the coming sections.

In WLANs, Selective Scanning methods are used to reduce the number of channel to be scanned from the full number of channels to only those channels that the system judges that they are used by neighboring APs. Fig. 2.6 shows the results of the selective scanning method proposed by Hye-Soo Kim et. al. [4]. Reducing channel scanning time also results in less power consumption since the device radio needs to be on for less time.

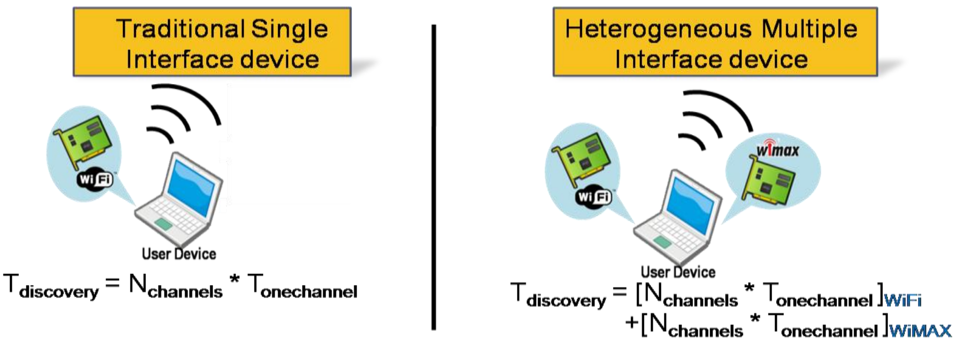

Fig.2.5 Comparison between the discovery time for a single NIC device versus a heterogeneous device in case the traditional candidate network discovery method was used

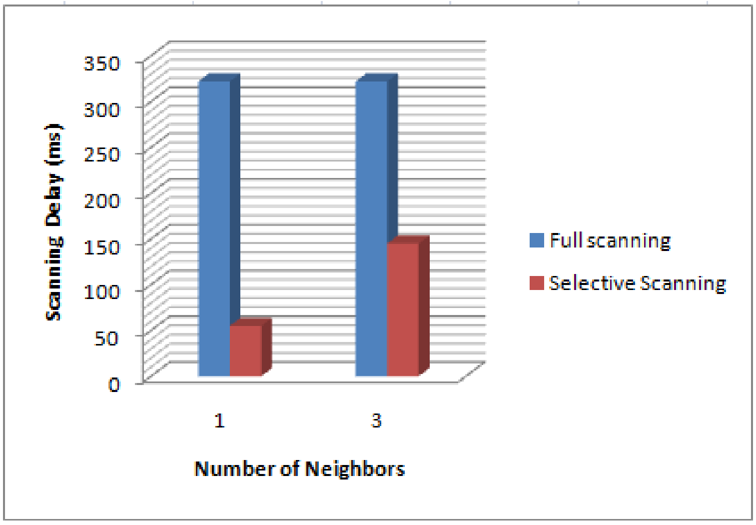

Fig. 2.6 A comparison between the time required to discover WiFi APs in the basic full scanning method versus the selective scanning method

2.2.2 Handover Decision Stage

In the decision stage, the device decides, according to a certain policy or settings, whether to perform a handover or not, and if a handover was found necessary, it decides the target network to handover to.

In traditional networks, handover decision is mainly based on RSS (Received Signal Strength) level. When the current serving network RSS goes down under a certain threshold, the device will measure RSS from neighboring networks and choose the network with the highest RSS. However, RSS based decision is not valid in the case of heterogeneous networks and more criteria needs to be considered for a smarter decision.

Modern approaches deal with network selection as a multi-criteria decision problem. The user is assumed to have an interface allowing him to define and configure different profiles that correspond to different situations.

Fig. 2.7 User Preferences Settings Screen Example

Each profile has a number of networks selection criteria, and the user can specify the priority of each criterion. A network evaluator or selector takes network information and profile parameters as inputs and gives the most suitable target network for handover. Fig.2.7 shows an example on a preferences and profile setting interface.

The handover decision stage can be further divided into two substages. In the first stage, user devices need to determine whether there was a need for a handover or not. A need for a handover occurs when the device is out of coverage of the currently serving network, or when a better performance network is in range. In the second stage the device needs to choose the target network from a number of candidate network options that are available in the device's current location.

A lot of network selection algorithms are being proposed and studied in the literature. In [5] authors considered a wide range of criteria for networks selection (cost, link quality, device velocity, battery life time...etc.) and used the utility concept for multi-criteria network selection. Each potential network is compared against the user preferences and the utility of each is calculated. The device will then choose the network with highest utility to handover to.

Lin C. Jian et. al. [6] propose a fuzzy based algorithm for handover decision in heterogeneous networks. The algorithm determine the possibility of a handover at a given instance of time and can adapt to the dynamic conditions in the network.

2.2.3 Handover Execution Stage

The last step involves signaling on the data link layer and possibly the network layer in order to reserve resources on the target network and complete the handover process.

In the context of WLANs, pre-authentication methods are used to enable the device to authenticate with the target access point before the handover actually occurs in a method called Fast Handover. This way, the delay that is caused by the exchange of authentication messages is reduced at the time of handover.

When Network Layer signaling is needed (switching between subnets or completely different networks) a network layer mobility protocol needs to be used.

MIPv6 (Mobile IP v6) is a strong candidate for a Layer 3 protocol in future networks[2]. It helps both devices and network entities in locating each other after handovers take place. As an enhancement to reduce the handover delays in MIPv6, Fast MIPv6 (FMIPv6) [7] was proposed in the literature.

FMIPv6 comes to address the following problem: how to allow a device to send packets as soon as it detects a new subnet link, and how to deliver packets to a device as soon as its attachment is detected by the new Access Router. In other words, FMIP's primary aim is to eliminate the factors of delay introduced by the address auto configuration procedure.

Research on handover execution has been extensively conducted and thus current research is focusing on the first two steps; how to perform network discovery in less delay and less consumption of power, and how to achieve smarter decisions through algorithms that take into consideration the user preferences and the current situation of the communication environment [8].

2.3 The Focus of this work

In this paper, we believe that the handover decision step is crucial in the overall handover process and that a smart handover decision depends heavily on the information available to the device at the time of handover decision. Thus, the main focus of this paper is the network discovery stage of the handover process and how to enhance the accuracy and timeliness of the information provided for devices in order to achieve better decisions and accordingly enhance the satisfaction level of users.

Next section describes recent research efforts on the network discovery step which is the main focus of this thesis..

2.4 Research Efforts on Network Discovery

2.4.1 Broadcasting Network Information

In this method, heterogeneous network information about networks that are located within the coverage area of a bigger network is broadcasted in that big network and user devices connected to the big network can receive this information. As an example of this method, Choi et al. proposed that 802.16e (WiMAX) base stations (BSs) periodically broadcast the information on the density of WLAN APs within their cell coverage [8]. By using the density information, a mobile device decides the scan interval of the WLAN interface by considering the energy consumption. But such approach is probabilistic and discovery can be done more efficiently if more accurate information is available.

Yaqub et. al.[9] propose a network entity called a Cell Broadcast Center (CBC) to broadcast partial network information in a selected number of networks called the Cell Broadcast Area (CBA). This partial information is mostly information about the existence of that network in that certain area. Then, if devices need more detailed information, they contact a server, called Network Information Repository (NIR), to obtain the detailed information about these networks. This method has three major drawbacks. First, there should be a method to define whether one network is located in one CBA or not, which makes the method more complex.

Second, if we assume that the device is connected to a large coverage network, small networks like WLANs might exist in that network but will not be necessarily available in the device’s current location. Yet, the device will still receive information about the existence of these WLANs while they cannot actually access them.

Third, the traffic of the information broadcast will consume part of the network bandwidth which otherwise can be used for actual communication needed by user applications. This is a drawback that is common to all broadcast based methods.

2.4.2 Utilization of an Information Server

The most efficient way for network discovery is to turn on the device NICs only when there are networks available. This requires the device to have information about available networks and their location. In such schemes a server that stores network information is used.

Providing Network Information by means of Information Servers is one of the network discovery methods that depend heavily on the device location. In such Network discovery schemes, devices contact Information servers to obtain information about candidate networks covering the device’s current location without performing a channel scan as in the traditional way. The concept is shown in Fig. 2.8. Thus, this scheme can be thought of as a location-based service and each server is responsible for a certain region as depicted in Fig. 2.9.

This method is also utilized in the new IEEE 802.21 standard dubbed as the Media Independent Handover (MIH) [10]. In MIH, devices obtain network information by sending queries to an information server that stores detailed network information. The query message includes the current device location, the radius of the area around the device of which network information should be returned, as shown in Fig.

2.10, and some other parameters that specify the required network type and the required information. This way, mobile devices can specify what network information it needs without wasting bandwidth with a response that contains many information that might not be of use for the device. For example, a device with only WiFi and WiMAX interfaces will not be interested in receiving information about available cellular networks so it specifies that it only needs information about surrounding WiFi and WiMAX network only.

Server based schemes overcome the previously mentioned limitations of the scanning based methods. The advantages can be explained in the three following points:

1) The device can obtain information about differing technology networks by switching on only one technology NIC while other NICs can be completely switched off.
2) The device can get much more detailed information about the available networks resulting in smarter handover decisions. It can also get the operating channels information which will eliminate the need for a full scan when performing a handover.
3) Highly mobile devices can effectively request information about networks that will come ahead in the direction of its movement. Thus it can plan ahead its handovers reducing the latency and packet losses incurred by the handover guaranteeing users a better service continuity.

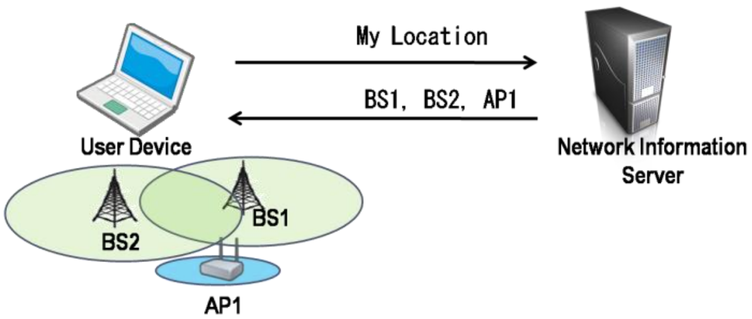

Fig. 2.8 The Concept of Server-based Network Discovery

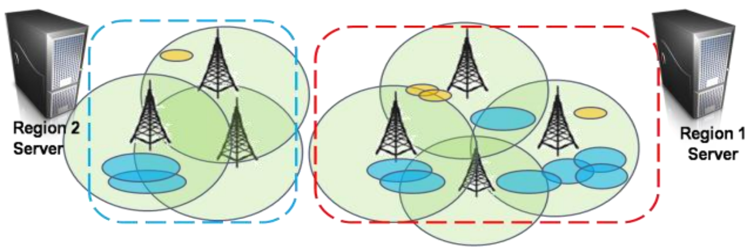

Fig. 2.9 Managing Network Information by Region.

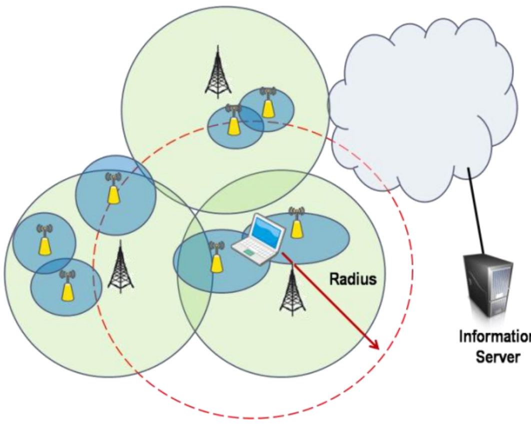

Fig. 2.10: Information Server-based network discovery in MIH

As another example of a server-based method, Chen et al. [11], proposed the Location Service Server (LSS) which stores the geographical information of WLAN AP locations. With the help of the LSS, a device can turn on its WLAN interface only if the WLAN is available.

Inoue et. al. [12] proposed the Basic Access Signaling (BAS) for RAN discovery in the MIRAI system . In this method, devices send a packet including its location information to a Resource Manager in the network to get the RANs available for the user in its location. However, in this method, there is no mechanism to update the Resource Manager in case network conditions change.

Information server-based methods have generally 2 main issues: First, devices need to discover these information servers. Second, there should be a way for information collection and update on these servers. One more point is that there are no deployment specifications or a supporting architecture that can enable management of such servers on larger scales. Most researches assume that the information server is already available for the device and that the information is already configured on the server.

As for information server discovery, S. Park et al. [13] proposed a new mechanism in conjunction with DHCP. It defines a new option for DHCPv4 for information server discovery. The address of the information server will be received by the mobile device in the typical DHCP offer message. However this method is only effective in small-scale deployment scenarios like a campus or enterprise network. Furthermore, it assumes that the information server address is already known and configured in the DHCP sever.

For information collection, authors in [3] propose a WLAN-only information server and consider a case where some APs are not managed by an information server. If the mobile device finds an AP that is not registered in the information server, it sends the server an update message.

A weak point about this approach is that we do not need to have all APs registered in an information server. This is the case in private WLANs where owners are not interested in announcing AP existence to the public.

To summarize, server-based methods need a way to enable devices to discover information servers, a way for these servers to collect network information and keep them up to date and a supporting architecture that is flexible enough to be applied on a global scale. These issues are still not adequately addressed in the literature and are the focus of this paper.

2.4.3 Methods Considering Dynamic Load Information

Balachandran et al. [14] presented an adaptive load balancing solution for WLANs where a centralized admission control server contains load information of all access points. Part of the information in the AP state table (e.g. AP name, IP/hardware address, and location) is static. In addition, the state table also maintains dynamically varying information like the number of associated mobile nodes in every cell and the aggregate throughput at each AP. This information needs to be updated at regular intervals of few tens of seconds by sending periodic requests to the APs in the network in order to keep the dynamic state updated.

One weak point about this method is that APs should be configured to cooperate with the server in order to respond to the requests of dynamic information update. If such method was to be applied in a heterogeneous network environment, APs and Base stations of all networks which belong to different operators are required to cooperate with the server. This might not be possible due to the fact that different network operators might not be willing to share network information about their own networks for competency reasons. In other words, network operators might not be willing to offer network information that might be used by devices to decide handover targets which might happen to be a network from a competing operator.

In the light of the information presented so far about the recent research efforts, the goal of this thesis is stated in the next section..

2.5 The Goal of this Research

In the current network-centric approaches for network and handover management, operators keep tight control over users so that their network is used to its greatest potential. But the telecom market is facing a migration from network centricity towards user-centricity since they prove to be the most efficient ways to guarantee the Always Best Connected (ABC) principle for users [5].

Our proposed approach is a user-centric method and aims to increase the users satisfaction level by enabling them to perform better handover decisions that take into consideration both static and dynamic information. It differs from the other efforts in this field in the fact that the information collection and offering to devices is performed in an operator independent way and without needing any infra-structure changes to existing networks or any kind of cooperation from the operators.

Chapter 3

The Proposed Network Information Server System

This thesis aims to increase the user satisfaction level by enabling them to perform more accurate handover decisions through providing not only static information but also dynamic information about surrounding networks. The proposed system can be thought of as a third party service that can provide network information in a network operator independent way.

This work differs from other efforts in the network discovery field by the following points:

1) Devices are used as agents to report network information to the server
2) Provide network information for devices in an operator independent way, hence no operator cooperation is needed.
3) No infra structure changes to the existing networks are needed to implement the proposed system
4) Network information broadcasts are not needed since a pure server method is utilized

First the basic server system that we previously proposed in [19] is briefly explained.

.

3.1 The Basic Network Information Server System

The server based network discovery methods were extended to enable them to be used on a wider scale by specifying the way in which servers should be deployed and the information that must be contained in these servers data bases. This will also make it easier for the device to find the server.

3.1.1 Basic Components

In the proposed network information architecture we define two main entities as shown in Fig.3.1.

1) Top-level Directory (TLD)
2) Regional Information Servers (RIS)

One Regional Information Server contains detailed information about networks in a predefined area. User devices contact these servers to get location-based Network Information, which means that the device sends its current location information to the servers and information about networks available in that location are returned. Fig. 3.2 shows example entries in an RIS.

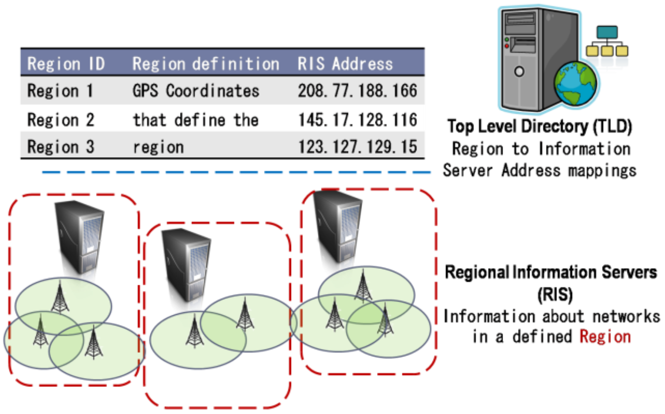

Fig. 3.1 The basic Network Information System. RISs contain Network Information. TLDs contain RIS addresses.

Network ID	Type	Location coordinates	Coverage radius	Data rate	Channels
1	3G UMTS	35.70043, 139.718628	30 Km	1Mbps	2100MHz
2	WiMAX	34.839356, 138.949585	5Km	2Mbps	30GHz
3	WiMAX	35.771772, 140.180054	2Km	5Mbps	10GHz
4	WiFi	36.21621, 139.32312	100m	11Mbps	6, 11

Fig. 3.2 RIS Network Information entries example

Top-level directories are used by user devices to retrieve the address of the RIS responsible for the area where the device currently is. This is important for devices that are highly mobile and move out of the area that is managed by the current RIS. In this case, devices will contact the TLD to obtain the address of the RIS responsible for the new area. Thus, TLDs contain data entries of RISs, their IP addresses and the region they are responsible for.

In addition to Network Information, RISs also include the address of the TLD they belong to. This way, devices will know how to contact the TLD in case they move out of the area of the current RIS.

It's important to mention that the strongest point about server based method, is the fact that devices can request network information much before a need for a handover occurs. This way, the handover latency contributed by the network discovery stage can be reduced to zero.

3.2 System Enhancement by the Addition of Dynamic Information Support

The system in the previous section is further enhanced by the addition of the dynamic information support capability on the server side. This means that the server will host dynamic network information in addition to the static network information considered so far.

Dynamic information is defined as those information that keep changing overtime. Network load and the per user throughput are examples on such information because they vary with the number of users in the network. On the other hand, static network information is information that is almost constant and does not change over short periods of time. For example, network coverage and location are considered static information.

In this thesis, one type of dynamic information is considered. That is, the network load information. Network load is defined as the number of users connected to the network at a certain point of time.
The next subsection explains the motivation behind considering dynamic network information.

3.2.1 The Importance of Providing Dynamic Information

Even with smart handover decision algorithms, having a service that can only provide static network information results in a number of problems. The importance of having dynamic information comes from:

1) To avoid misleading handover decisions

Handover decision algorithms try to choose networks that provide the best performance for users. For example, an algorithm would choose the network with the highest data rate. But in this case, the decision algorithms running on all user devices would choose that high data rate network resulting in an increased load on that network. New coming users running the same algorithm would assume that this network will provide them with high data rate, but because many users performed handovers to that “Good network”, the amount of actual data rate available for that user will be much less than the static value obtained from the server, or in the worst case the user won’t be able to handover to that network.

This happened because the decision algorithm lacks dynamic information about the network loads, so users were misled and accordingly their satisfaction level will decrease. This idea is shown in Fig 3.3.

2) Load balancing

When load information is not considered as criteria in the handover process, this will lead to problems regarding load balancing in networks. Users will prefer one network and overload it while the other is not well utilized.

Most Cellular networks employ some kind of network controlled load balancing techniques and so it might not be as severe as in the case with WLANs. This is mostly important in handovers that are user controlled.

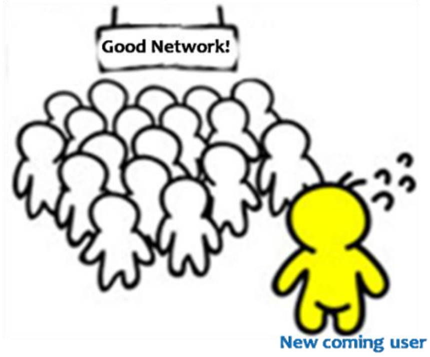

Fig.3.3 The problem with handover decisions not taking network load information into consideration

Therefore, by providing dynamic network information, network load information in our case, we can avoid the above mentioned situations and also achieve some kind of load balancing without having to implement special mechanisms on the network side for this purpose.

3.2.2 Dynamic Information Collection

The proposed method is based on the principle of making use of information currently available for one device in order to benefit other devices. An example on a service that makes use of this principle is PlaceEngine[18]. PlaceEngine is a service that makes it possible for users to easily estimate current location by utilizing WiFi devices, so that they can find out more about their surroundings and gain access to local information. The PlaceEngine server estimates the device's location (latitude/longitude, address, etc) from the WiFi information sent from the Client software and Wi-Fi information database maintained at the server. This server information is actually collected by devices. Users can register information about their current location along with the WiFi AP names available in that location regardless of whether these APs belong to the user or not. This way, other users who are located in that same location can make use of such information and get their location information. Fig. 3.4 shows how a user can input AP information even ones not belonging to that user, and register these information in the PlaceEngine database.

Fig. 3.4 PlaceEngine method of collection location information based on information provided by users.

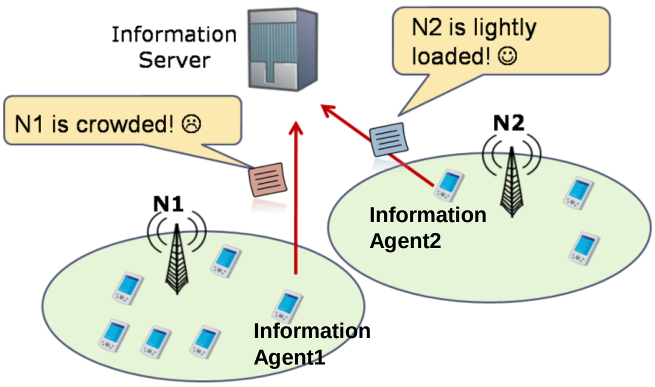

Fig. 3.5 The main concept used in load information collection. Information is reported by devices.

In the same way, in our proposed system , network information is collected by user devices. We call these devices that report network information as “Information Agents”. This is shown in Fig. 3.5.

Information Agents need to send load information updates for the server in a periodic manner about the networks they are currently connected to. There are many ways in which Information Agents can obtain the load information of the network they are currently connected to. One of these methods is from the beacons or the frames that announce the network existence.

For example, vendors incorporate certain load balancing features in their device drivers, AP firmwares, and WLAN cards. In these solutions, APs broadcast their load levels to users via modified beacon messages[15]. Also, different AP commercial products such as Cisco aironet access points [16] announce their utilization in beacon frames. Information Agents make use of such information and report it to the server.

In case the access point or base station does not provide capabilities for announcing the network load, load estimation methods, such as in [17] can be used to get a rough value estimating the network load.

We employ a polling method that is implemented on the server side in order to periodically choose Information Agents and request them to send network load updates about the network they are currently connected to.

3.3 System Model and Operation

Messages exchanged between the server and the devices are listed in Table 3.1.
User devices are assumed to be running a client program for interacting with the information servers. This program will be also responsible for managing user preferences and triggering the handover decision function.

When devices decide to contact the information server for information requests, they first need to register themselves as users of the system using the `register_device` message.

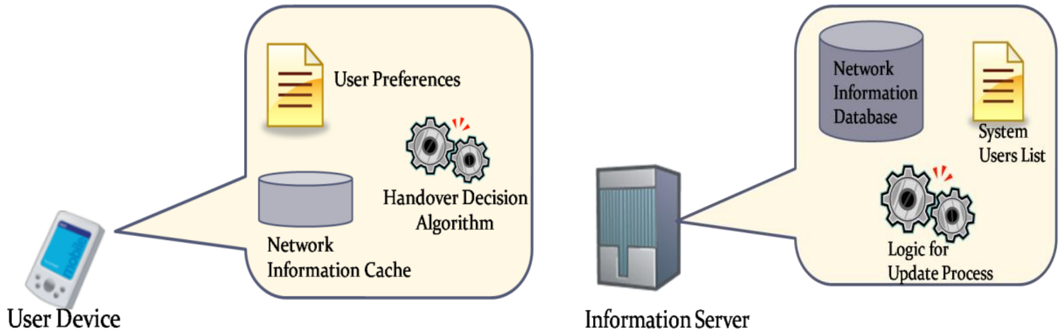

Fig. 3.6 The components on both the device and the server side of the proposed system

After registration, devices will be able to request information from the server using the `information_request` message.

Servers keep a list of devices registered as users of the service. This list is indexed by networks and each network entry contains a list of devices that are currently connected to that network. The server uses this list to choose devices to be Information Agents and poll these devices for load information updates using the `report_request` message. Fig. 3.6 shows the main components on the device and the server side of the system.

Information Agents will be polled for load information updates at a certain rate and in order to divide this update burden between all Agents, the server chooses a different device for each `report_request`.

For example, if we assume that the server sends a `report_request` each 30 seconds, first it sends the request to device1, then after 30seconds it chooses another device from the system users list, say device2, for the poll and so on. In the system implementation, devices are chosen in a round robin fashion for simplicity.

It's important to mention that devices can provide load information only about the network that they are connected to at the time of the poll. This means that the system will lose a device temporarily if it handovers to another network. For this reason, if devices perform a handover, they need to re-register with the server from the new network.

To summarize, devices and servers operate as in Fig.3.7 and Fig.3.8 respectively.

Message Name	**Message Direction**	**Contents**	**Effect**
`register_device`	Device→Sever	Device identification Currently connected network identification	Server will add the device to the devices list. Device can request information from the server and can participate in the dynamic information collection process
`information_request`	Device→Server	Request parameters such as minimum required bandwidth and maximum cost	Server will respond with an `information_response` message
`information_response`	Server→Device	Network information that match the parameters in the request message	Device receives network information and use them for handover decision
`load_report`	Device→ Server	Load information of the currently connected network	Server updates the load parameters of the network the sending device is connected to
`report_request`	Server→Device	Optionally include a value for polling rate	Device responds with a `load_report` message

Table 3.1 A list of the main messages exchanged between the server and devices

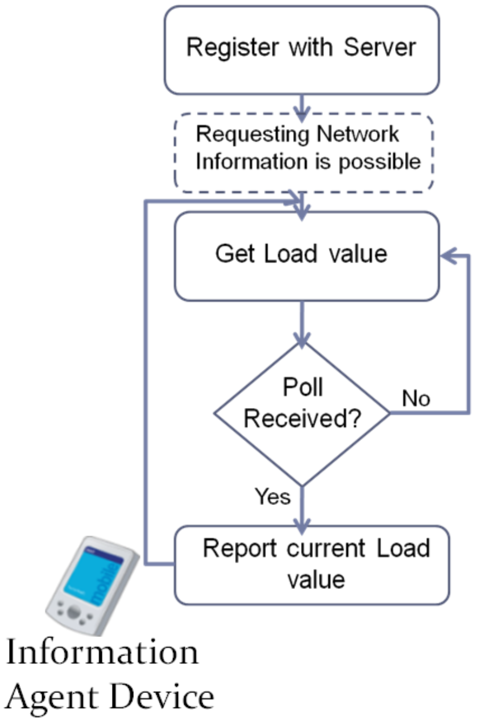

Fig. 3.7 Agent Devices operation flow chart

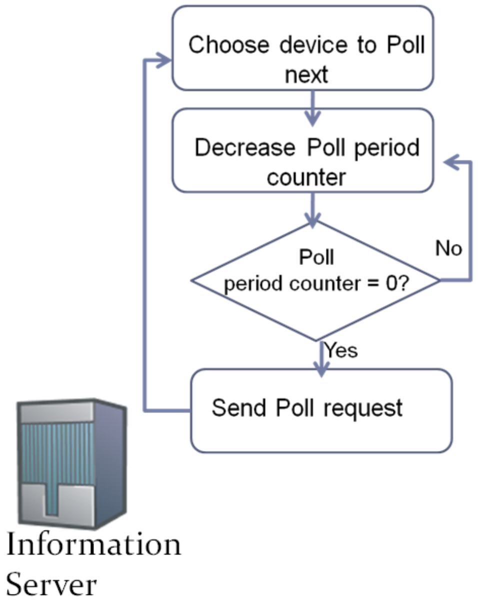

Fig. 3.8 Server operation flow chart

3.4 Adaptation to network activity level - Update Rate Tuning

The system is further enhanced by making it capable of adjusting the rate at which devices are polled for each individual network. This means that instead of polling the devices of the same network at the same rate all the time, the system should be able to adjust the poll rate to meet the characteristics of different networks. For example, some networks might be less active than other networks, meaning that the rate at which users enter or leave that network is less than some other network where users are constantly changing in number. Such networks with low activity don't need polling devices for load updates at the same rate as the networks with higher activity because this means that the server will need to deal with frequent updates that don't have much differences in the reported value.

For example assume the WiFi hotspots in some airport. These hotspots will have more activity (users entering or exiting the network) during periods when flight arrivals are high and will be less active in less flight arrivals conditions.

To reduce the amount of unnecessary load updates, a method is proposed to enable servers to monitor the changes of loads of the different networks and accordingly adjust the polling rate value.

3.4.1 The Adaptation Algorithm

Servers will be monitoring load updates from devices on each network. A time window of size N is defined to slide over the past updates and calculate the differences between each two consecutive updates. The time average of these differences is calculated for each window. If this average value exceeds a certain threshold, the update rate will be increased to adapt to the increased activity of the network. In the same way, if that average falls under another threshold, the rate will be decreased to adapt to the reduced level of activity of the network. Fig. 3.9 shows an example on how poll period changes in response to changed activity of the network.

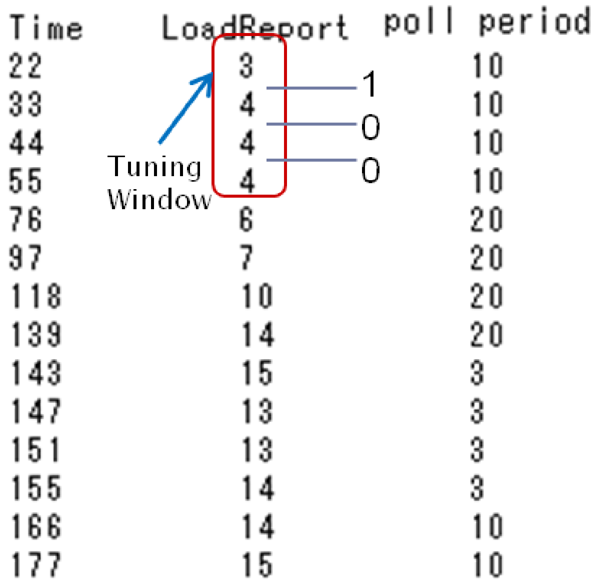

Fig. 3.9 Poll rate tuning example with window of 40 seconds

For the current implementation, 4 thresholds were defined at which polling period will be increased or decreased. These are shown in table 3.2.

Parameter	**Value**	**Action**
Increase threshold	<0.15	Increase poll period by 5 seconds
Decrease range 1	0.15 ~ 0.45	Decrease poll period by 3 seconds
Decrease range 2	0.45 ~ 0.65	Decrease poll period by 7 seconds
Decrease range 3	> 0.65	Decrease poll period by 10 seconds

Table 3.2 Main parameters of the tuning algorithm

The parameters in the table above were chosen such that the algorithm will react quickly to an increased user arrival rate with three different adaptation ranges. The poll period will be decreased when the average difference of consecutive readings in one window decrease below 0.15. This means that the user arrival rate decreased and thus the server can reduce the polling rate of devices.

3.5 Assumptions

Devices are assumed to agree to perform as Information Agents in return for allowing them to access the Information Server Database. They are also assumed to have the capability of determining their current location.

This system proves to be best used when users are using foreign networks; i.e. networks that they don't usually use. For example, business men that travel very often and need to keep connected to their email accounts or perform web conferencing can make the best use of such system especially when they have the capabilities of roaming and accessing networks of many providers.

3.6 Summary

This chapter is concluded by a summary of the advantages of the proposed method. The first three points are unique to this work while the remaining last two points are common for server based methods.

Advantages of the proposed method

1) Devices can get information about the load of heterogeneous networks while connected through one interface.
2) Since it's a cooperative method depending on devices for information collection, devices can obtain load information about networks of different providers
3) Devices do not need to listen to potential handover targets to get load information as in traditional methods. Devices can obtain load information of surrounding networks before a need for a handover rises.
4) Devices can retrieve information about different RAT networks by using only one NIC. This greatly saves power because devices do not need to turn all its NICs.
5) Devices that obtained network information can directly perform handover to the new network without the need for performing a scan on all channels, since the obtained information include the operating channels of the different networks. This reduces the time for network discovery.

Chapter 4

System Evaluation

This chapter provides an explanation for the system evaluation metrics and the simulation experiments that were used to produce the results. Among the number of metrics that are used for evaluation is the *User Satisfaction Level* which is derived from the *Utility* of the network. First, the concept of “Utility” is briefly explained in the next section.

4.1 The Utility Concept

The Utility value is one of the many mathematical solutions for multi-criteria selection problems. Below is an explanation of how this concept is employed in network selection.

4.1.1 Application of Utility in the Network Selection field

Utility is a method to find an optimum solution from a set that contains many options each with a number of criteria. As an example for the application of Utility-based selection in the network selection field, consider the following example.

For a network set of S={Net1,Net2,Net3} each with their own values of data rate, cost, and coverage as shown in Fig.4.1 , we need to find the best network for each of user1(a user preferring high data rate networks) and user2(a user that prefers the cheapest possible network).

Network Options

Name	Data Rate	Coverage	Cost
Net1	2Mbps	Small	5
Net2	300Kbps	Large	20
Net3	1Mbps	Small	0

User preferences

Fig.4.1 Example of utility application to find the most suitable network for different users

Each of the two users will calculate a value, called the "Network Utility", for each one of the 3 network options by inputting the options' parameter values (in our example, data rate, cost and coverage) to a function called "the Utility Function". Each of the 3 criteria have their own utility function that defines the behavior of the user's satisfaction curve as the criteria values change.

It will assign a value to each of these networks that shows how desirable these networks are. For example, the utility function for the data rate criterion should be an increasing function since users are more satisfied with higher bandwidth networks. In the other hand, the utility function for the cost should be normally a decreasing function, i.e. users become less satisfied with networks of high cost. Fig. 4.2 shows how utility values are derived from criterion values using the utility function.

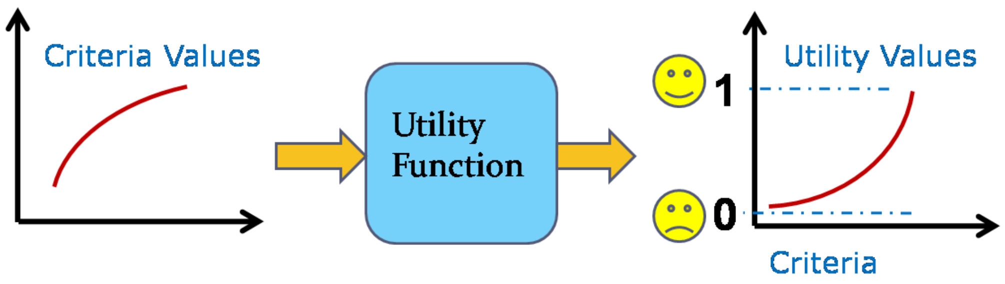

Fig.4.2 Utility functions map the values of a certain criteria to values that represent user satisfaction.

In this thesis, we employ this utility method in order to find out the best network option that gives the user the highest level of satisfaction. Utility values are an indicator on how satisfied the user is expected to be at any point of time in the simulations.

4.1.2 Utility Calculation

In this work, utility functions and parameters defined by Q.-T. Nguyen-Vuong et al. [5] are used by the decision functions running inside the devices to choose the most suitable network. The utility value for each network can be calculated as follows:

Step 1 : Defining the criteria type

Decision functions need to know whether a criterion is an upward criterion or downward criterion. In upward criteria, the value of the utility increase as the value of the criterion increase (like data rate). Downward criterion have decreasing utilities as the criterion value increases (the cost criterion is an example on this).

Step 2 : Preferences weight setting

Users set the weight values for each of the network properties in a way that reflect their preferences. For example, user1 in the example of Fig.4.1 will give the highest weight for the data rate property since he cares the most about the data rate of the networks. User2 in the same example will assign the highest value for the cost property since he is a cost sensitive user, meaning that a network with low data rate is acceptable for him as long as the cost is lowest.

For functions with 3 levels of preferences (ignored, low, high) numbers such as (0,1,2) can be used to represent these preference levels.

Step 3 : Calculation of elementary utility value u_m for each single criterion

Nguyen-Vuong et al. [5] use the following function to find elementary utility values for each single criterion of all networks. They performed a survey on different utility function forms and proposed a new utility function that can better model the utilities of different networks.

The elementary utility for an upward criterion m that has a value x is calculated by

$$u_m(x) = \begin{cases} 0, & x < x_\alpha, \\ \frac{\left(\frac{x-x_\alpha}{x_0-x_\alpha}\right)^\zeta}{1+\left(\frac{x-x_\alpha}{x_0-x_\alpha}\right)^\zeta}, & x_\alpha \leqslant x \leqslant x_o, \\ 1-\frac{\left(\frac{x_\beta-x}{x_\beta-x_0}\right)^\gamma}{1+\left(\frac{x_\beta-x}{x_\beta-x_0}\right)^\gamma}, & x_o < x \leqslant x_\beta, \\ 1, & x > x_\beta, \end{cases} \tag{4.1}$$

where χ_α is the lower boundary of the criterion value χ. χ_β is the upper boundary of χ and χ_0 is the value at which the utility function will return a utility value of 0.5. ζ defines the steepness of the function.

The function for a downward criterion is **1-$u_m(\chi)$**. Fig. 4.3 shows an example of some downward criterion utility functions with χ_α=0, χ_β=80 and with different parameter settings for χ_0 and ζ.

Step4: Calculation of the aggregate utility U for each one of the Networks

The final step is to calculate the aggregate utility of the network. This can be obtained by the following equation where α_m is the weight assigned to criterion m.

$$U(\mathbf{x}_i) = \prod_{m=1}^{M} [u_m(x_m)]^{\alpha_m} \tag{4.2}$$

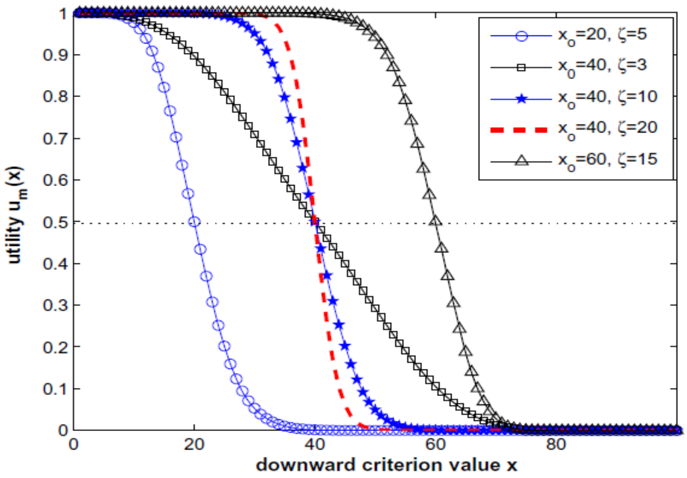

Fig 4.3 Elementary utility function curves for a downward criterion (χ_α=0, χ_β=80)

The value obtained in step4 is the final utility value for network i and is the value compared between different networks in order to decide the best network. The network that has the highest value of U among a number of network options is the network that best suits the user.

4.1.3 User Satisfaction Level

One of the most important evaluation metrics is the user satisfaction level. The user satisfaction function is a Utility-like function but takes into consideration the actual parameters of the network.

In other words, the utility function is used to evaluate the handover candidates and choose the best network that is expected to be the best matching for the user. Its input parameters are the network information obtained from the information servers. On the other hand, the satisfaction level is the utility of the network after performing a handover and is dependent on the conditions of that network at the time of handover.

4.2 Evaluation Metrics

The proposed system is evaluated through two main categories of metrics; one related to the main concept of having dynamic network information and the other one is related to the system performance.

4.2.1 Concept Evaluation

From the simulation study, we aim at exploring the effect of having dynamic network information on:

1) The overall satisfaction levels of users

Having dynamic network information helps in achieving more accurate handover decisions and thus user satisfaction levels are expected to increase.

The average of satisfaction levels of all users is calculated over time in a scenario that makes use of the proposed system and is compared to a scenario that doesn't make use of dynamic information.

Also, the increase in satisfaction is explored. This is defined as the average of the difference between the user satisfaction levels over time.

2) The handover blocking ratio

Handover blocking occurs when a device tries to handover to a heavily loaded network. The resources of that network can't handle additional users so the device is not allowed to handover to that network.

Handover blocking ratio is defined as the ratio between the number of times a handover was blocked and the total number of times a handover attempt took place.

With dynamic information telling users about the load levels of networks, users will avoid handovers to heavily loaded networks reducing the handover blocking ratio.

A comparison of the overall handover blocking rate is performed between scenarios with dynamic information support and others without dynamic information support.

4.2.2 System Performance Evaluation

The effect of Information Update Tuning enhancement is investigated over:

1) Amount of error in server data

The difference between the actual network loads and the load information stored in the server is investigated for different update rates and for the case that uses rate tuning. Update rate tuning is expected to reduce the load error. The time average of the server error is calculated and compared.

2) Update demand on devices and traffic arriving on the server

Devices are responsible for sending information updates and thus a portion of the bandwidth is used for this purpose. The effect of this process on the demand on device reports and server traffic is studied.

4.3 Simulation Model

4.3.1 Environment

Parameter settings used in simulations are as shown in Table 4.1, unless otherwise stated. Network selection is performed based on three criteria, the Network Datarate(static), cost(static) and Network Load(dynamic). Their elementary utility curves are shown in Fig.4.4,Fig.4.5 and Fig. 4.6.

User interarrival time was chosen to be a poisson distribution with a lambda of 7 seconds. The distribution of user location over the simulation area is a normal distribution to give more density of users near the center of the area (where more networks are available) and less number of users near the edges of the coverage of the networks. Devices are mobile with a probability of 70% and have velocities ranging from stationary to medium mobility (0~10m/s).

Parameter	Value
Simulation Area	800m ×600m
Simulation run time×run count	600seconds×5 runs
Maximum number of users	100 users
User inter arrival time	Poisson distribution λ = 7 Seconds
User location distribution	Normal distribution with X: μ = 400, σ^2=200 Y: μ=300 , σ^2=180
Probability of a user being Mobile	70%
Mobility Model	Random Way Point Mobility
Movement direction change probability	25% for X, 15% for Y
Device velocity	0~10m/s
RTT to server	40ms from WiFi/ 180ms from Cellular
Tuning window size, (max,min) poll period	30 seconds , (34 , 3)
Utility criterion parameters (χ_α , χ_β , χ_0 , ζ)	Data rate(100,3000,600,3) Cost (5,70,35,2) Load (0, 100, 47, 2)

Table 4.1 Simulation parameters

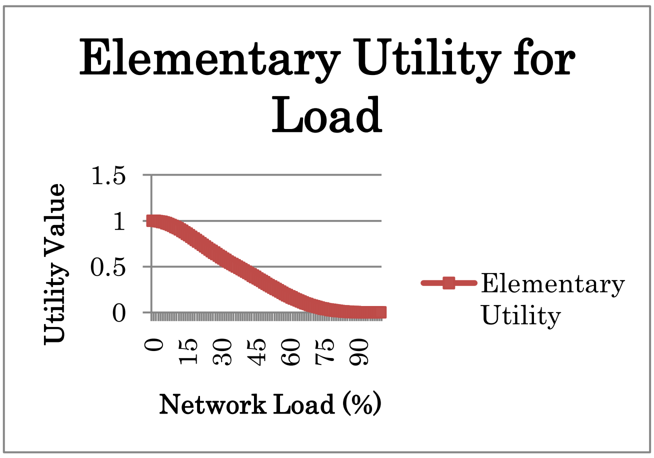

Fig. 4.4 Curve for elementary utility of load criterion

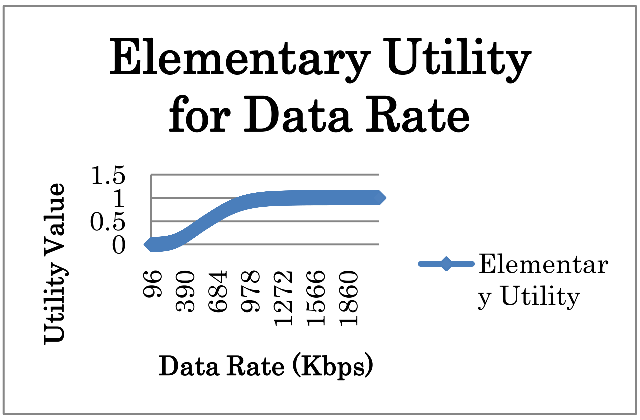

Fig. 4.5 Curve for elementary utility of the data rate criterion

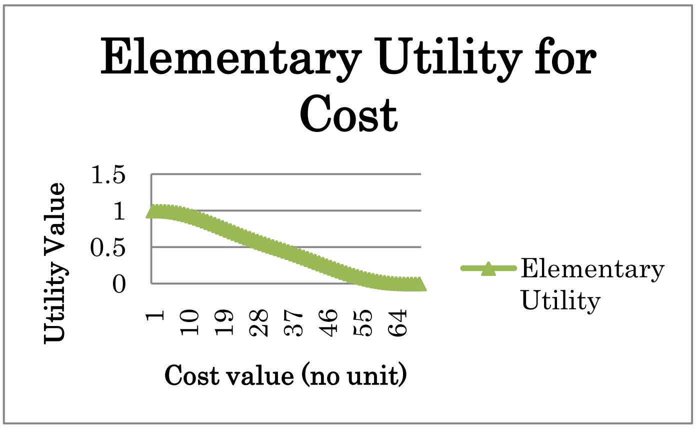

Fig. 4.6 Curve for elementary utility of the cost criterion

Simulations are mainly performed considering two network scenarios.

1) Scenario1

In this scenario handovers due to devices mobility and handovers due to the user satisfaction level are both possible.

In this scenario there are 6 networks with the following parameters:

#	Type	Location (X,Y)	Coverage	Max users	Theoretical data rate	Cost
1	Cellular	200, 290	300m	30	500Kbps	20
2	WiFi	250, 150	100m	15	2Mbps	10
3	WiFi	250, 350	80m	25	1.5Mbps	15
4	WiFi	300, 250	80m	15	2Mbps	Free
5	Cellular	500, 290	300m	35	500Kbps	20
6	Cellular	450, 290	All over the simulated area	15	300Kbps	15

Table 4.2 Network parameters of scenario 1

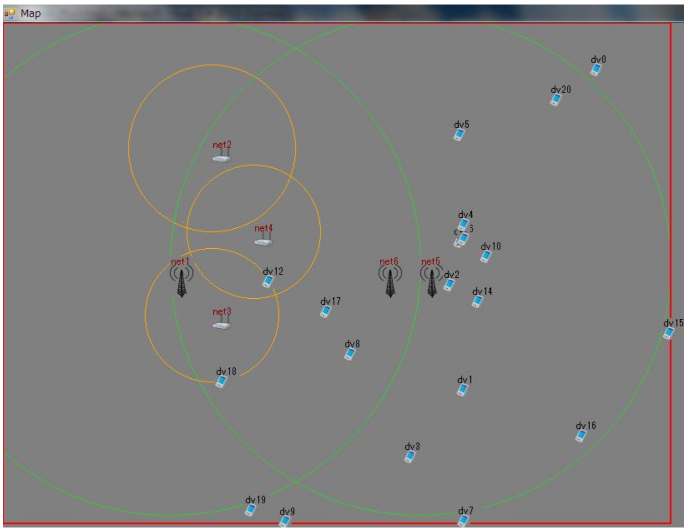

Fig. 4.7 Snapshot of scenario 1 showing relative locations of the different networks

2)Scenario 2

In this case, there are 3 networks that are available all over the simulation area. This means that no mobility-type handovers occur in the simulation.

#	Type	Location (X,Y)	Coverage	Max users	Theoretical data rate	Cost
1	Cellular	380, 320	All over the simulated area	60	500Kbps	10
2	WiFi	440, 320		10	2Mbps	Free
3	WiFi	380, 390		35	1.5Mbps	Free

Table 4.3 Network parameters of scenario 2

Fig 4.8 Snapshot of scenario 2

4.3.2 Users

In the simulations three categories of users are defined. These are shown in Fig. 4.9 and their weight settings is shown in Table 4.4.

1) Normal users

These represent users with highest percentage. They give the highest importance to the cost of the network meaning that they prefer networks with low cost as a first priority before considering the data rate or the load. Web browsing and mail checking are the main activities of these users.

2) Business users

Represent 20% of users and care the most about having a good performance network to complete their tasks thus giving the highest weight to the network load criterion with a zero weight for cost.

3) Load-sensitive users

Represent the rest of users and give the highest priority to the load parameter.

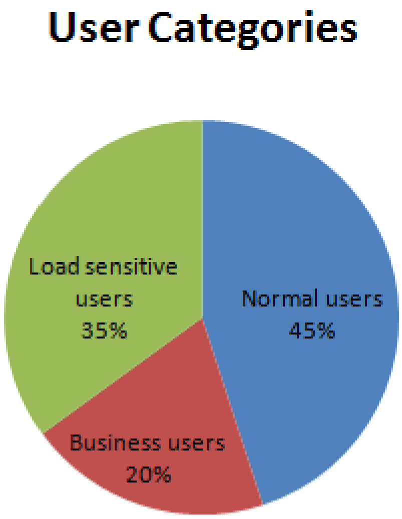

Fig. 4.9 User category percentages

User Category	**Weight Settings (Datarate, Cost, Load)**
Normal users	(1, 2, 1)
Business users	(1, 0, 2)
Load sensitive users	(1, 1, 3)

Table 4.4 Utility function weight settings for the different user categories

4.3.3 Handover Triggers

In the simulations, devices would try to perform a handover in one of the following 2 cases:

1) Going out of coverage of the current network

2) If the current satisfaction level decreases such that there exists another network that can provide better satisfaction for the user.

As for the 2nd point, we define a value called the handover gain,

$$\text{Handover Gain} = \frac{\mathrm{U}(\mathrm{N}_{\mathrm{Best}})}{\mathrm{Sat}(\text{Current Net})} \tag{4.3}$$

Where $\mathrm{U}(\mathrm{N}_{\mathrm{Best}})$ is the Utility value of the network with the highest utiliy, and $\mathrm{Sat}(\text{Current Net})$ is the satisfaction level achieved by the currently connected network.

The handover will occur according to the following condition,

if (Handover Gain > gainThreshold)

then Perform handover

Handover Gain was set to 1.3 for the simulations in all cases. This value was chosen such that no frequent handovers with small gains occur. If there exists a network that can give a gain of 1.3 or more, the device will handover to that network.

4.4 Simulation Results

4.4.1 Effect on Overall User Satisfaction

Both scenarios were run 5 times and the average of user satisfaction levels (sum of satisfaction levels divided by the number of users at any point of time) is plotted over time for one chosen run in Fig.4.10 for scenario 1 and Fig. 4.12 for scenario 2. Also, the maximum and average increase in user satisfaction obtained by the use of the proposed system was shown in Fig. 4.11 and Fig. 4.13.

We notice that the satisfaction level achieved by the use of the proposed system is near to that of the case that doesn't make use of it in the beginning and in the end of the simulation. This is because all networks are very lightly loaded at the beginning and very heavily loaded at the end, whether the device chooses a network according to its load conditions or not, satisfaction levels will be high (beginning) and low (end).

In the first scenario, a maximum satisfaction increase of around 0.29 is achieved when using the proposed system. This value is noticed to exceed 0.40 in the second scenario due to the different network setting of that scenario; devices can always stay with the network providing the highest level of satisfaction and do not need to handover for coverage reasons..

It's also noticed that the satisfaction levels of scenario 1 change in relatively large oscillations compared to that of scenario 2. This is due to the fact that handovers for mobility reasons (moving out of coverage of some network) occur often in this scenario. This means that the device will have no option but to handover, sometimes to a network with less utility since it's the only available option. This type of handovers does not take place in scenario 2 and thus less oscillations is seen in the satisfaction levels.

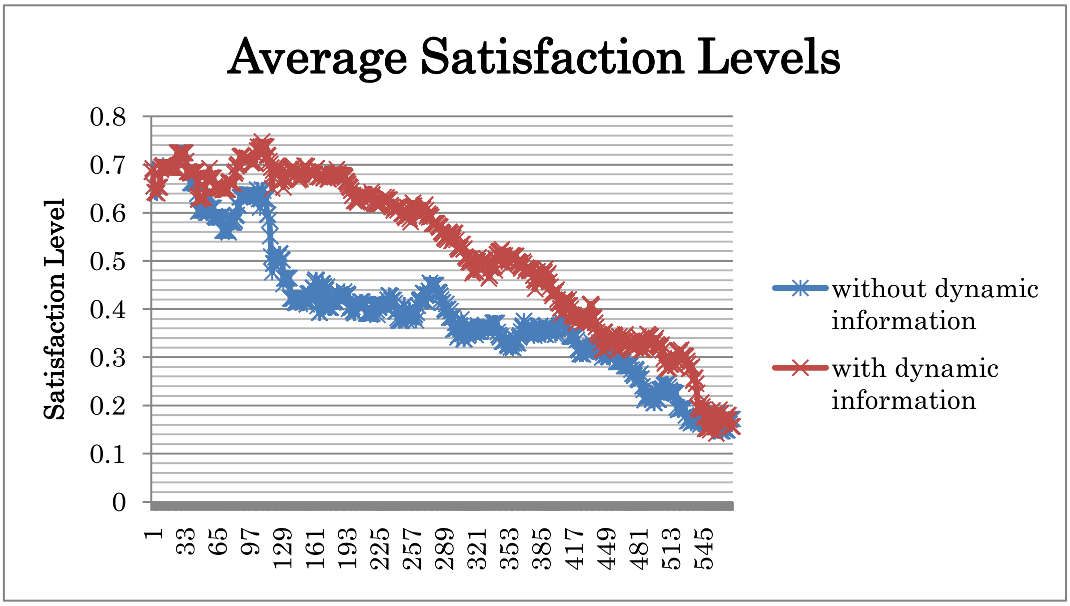

Fig. 4.10 Average satisfaction levels of users over time for one run of scenario 1

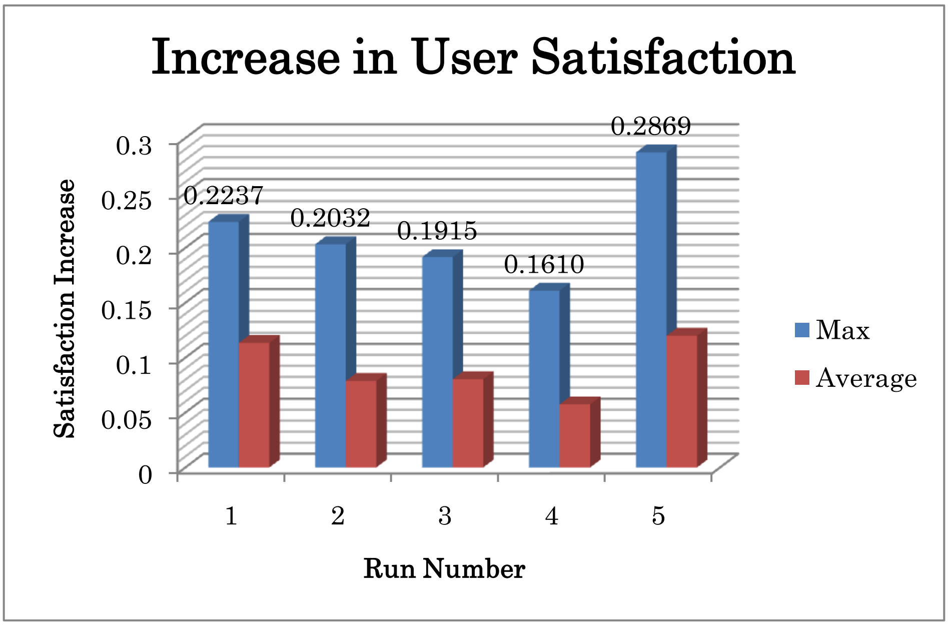

Fig. 4.11 Maximum and Average Increase of user Satisfaction for different runs of Scenario 1

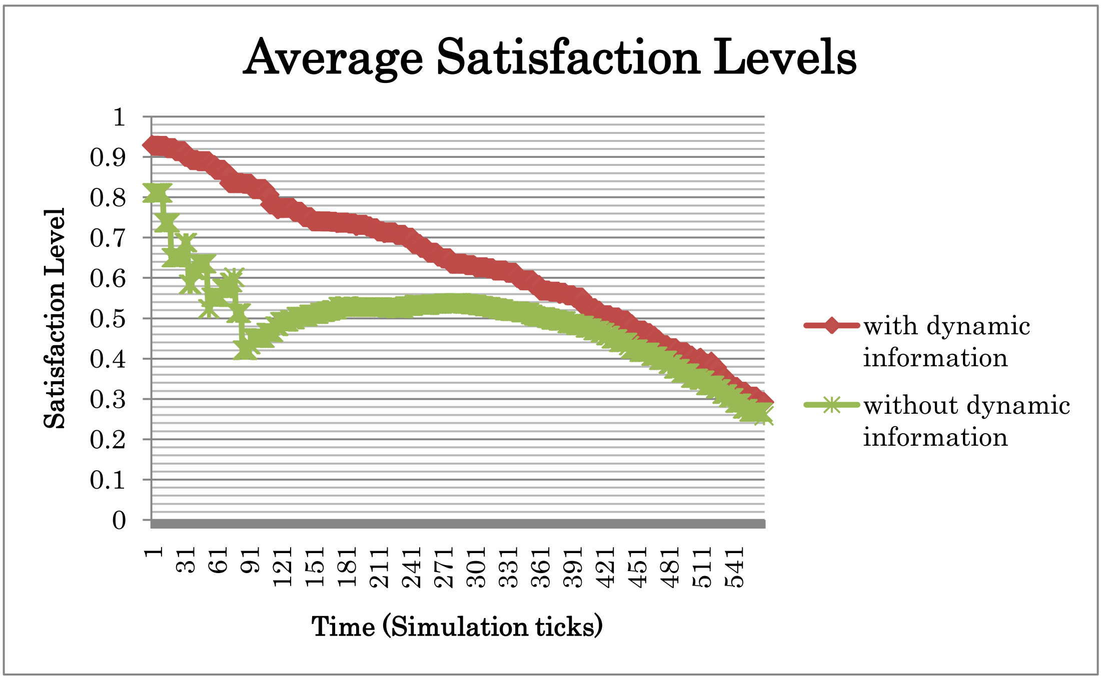

Fig. 4.12 Average satisfaction levels of users over time for one run of scenario 2

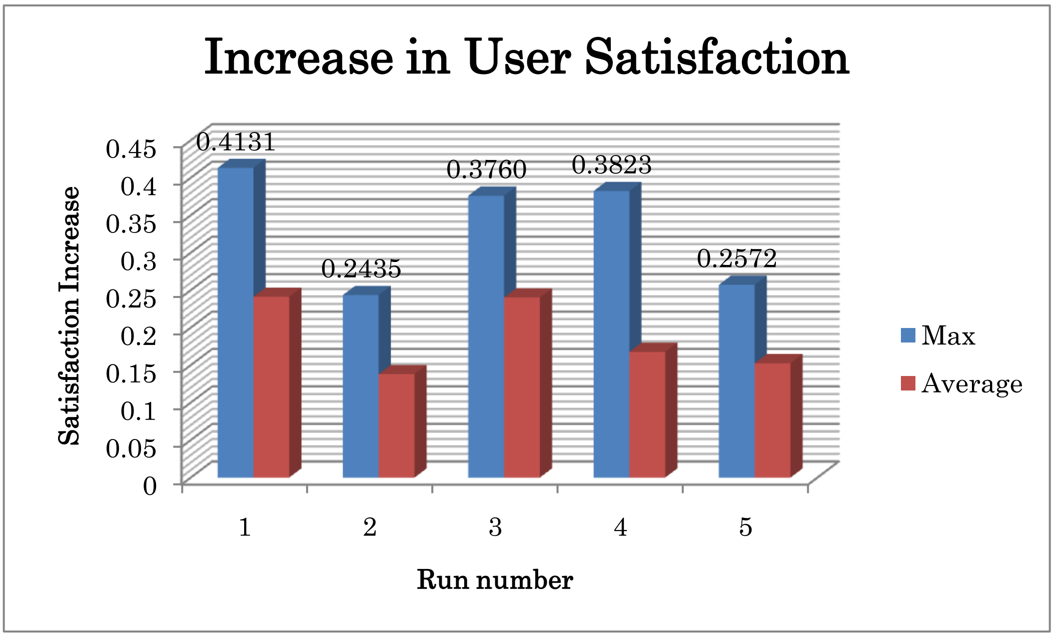

Fig. 4.13 Maximum and Average Increase of user Satisfaction for different runs of Scenario 2

4.4.2 Effect on Handover Blocking Ratio

The effect of using the proposed system was also investigated over the overall handover blocking ratio (Fig. 4.14 and Fig. 4.15) which is one of the problems discussed in the previous chapter. It is shown that our system can drastically decrease the blocking ratio for both scenarios, which are expected results since devices with the knowledge of network load won't try to handover to a heavily loaded network. Fig. 4.16 shows the effect of handover blocking on the elementary user satisfaction.

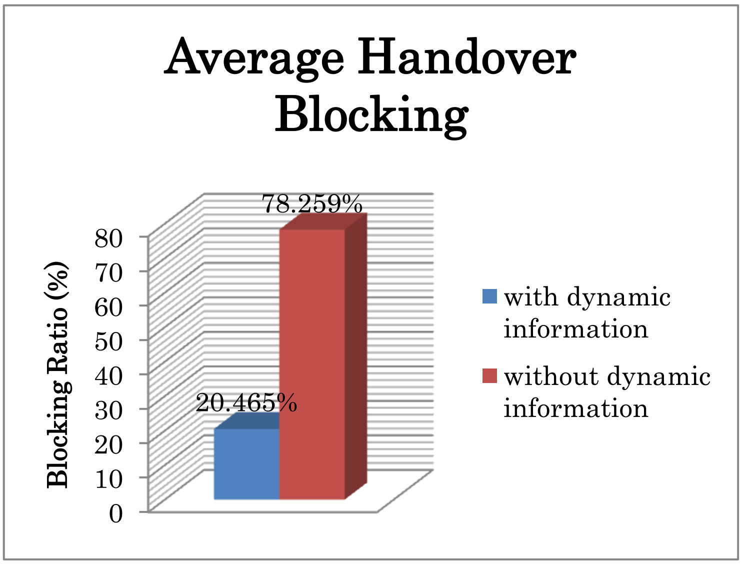

Fig. 4.14 Average handover blocking for Scenario 1

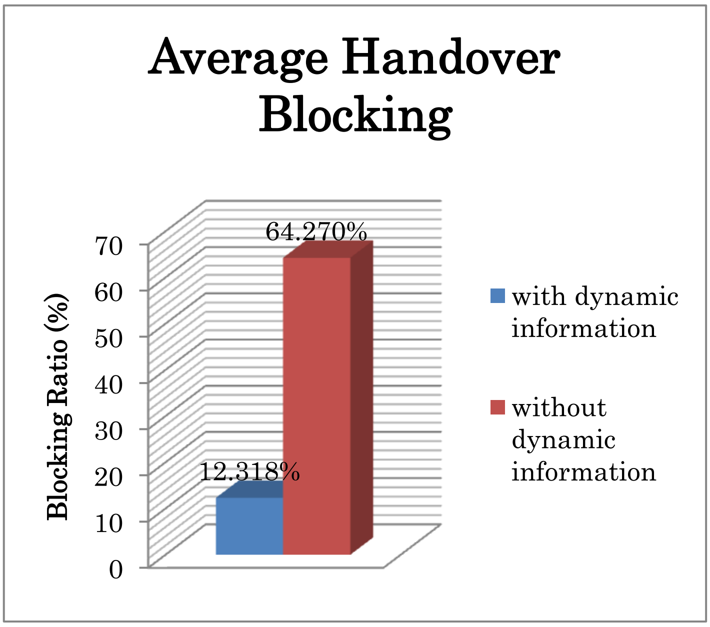

Fig. 4.15 Average handover blocking ratio for Scenario 2

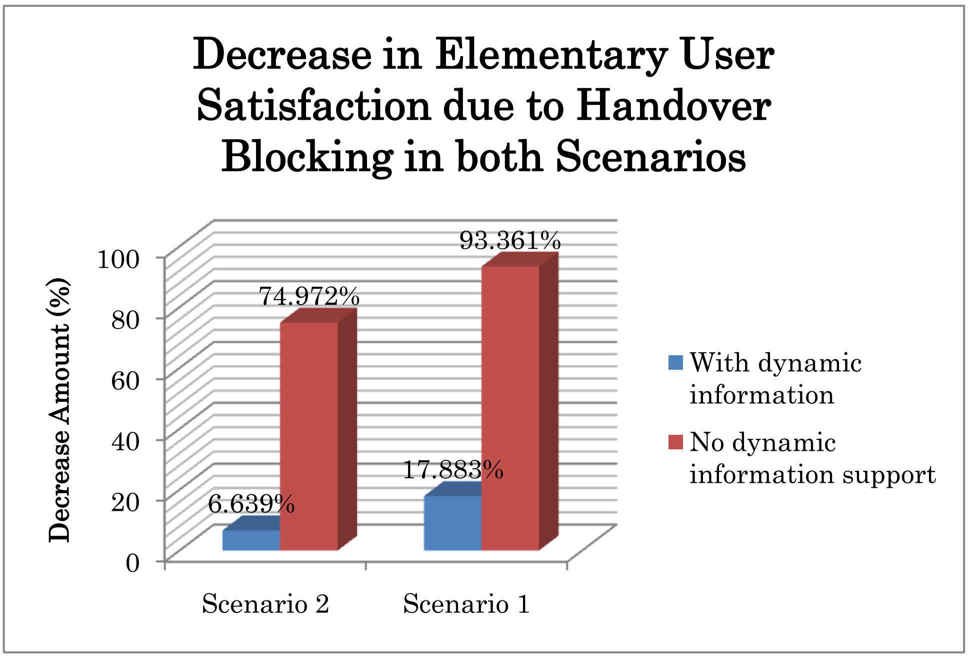

Fig. 4.16 Decrease in user satisfaction level due to Handover Blocking (χ_α=0, χ_β=90 , χ_0 =45 and ζ = 2)

4.4.3 Effect of Information Update Rate on the Server Error

The effect of the information update rate on the freshness of information on the server side is investigated in this subsection. This is an evaluation that is more close to an evaluation of the server side performance, that's why one scenario is used for this purpose. Scenario 2 is chosen since the arrangement of networks won't have much effect on the result.

Fig. 4.17 shows the amount of error in load information on the server for different values of polling periods for the 3 networks of scenario 2. As expected, errors increase as the polling period increases. Notice also that the amount of this error is different for the 3 networks, since the number devices connected to these networks will differ according to the utility value of these networks. Net1 had the largest errors in all cases. This is an indication on the network high activity.

Fig. 4.18 shows the results with the update rate tuning support. The tuning support gave results that can be somewhere between the values obtained from poll period = 13 and poll period = 20.

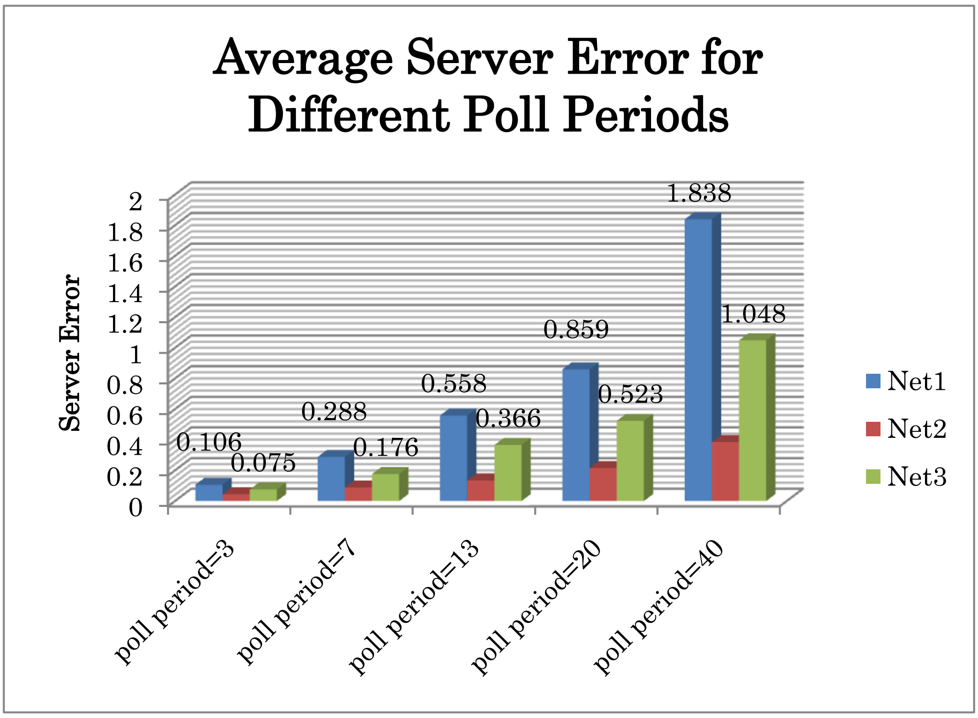

Fig. 4.17 Average server error for the 3 networks of scenario 2 with different poll periods

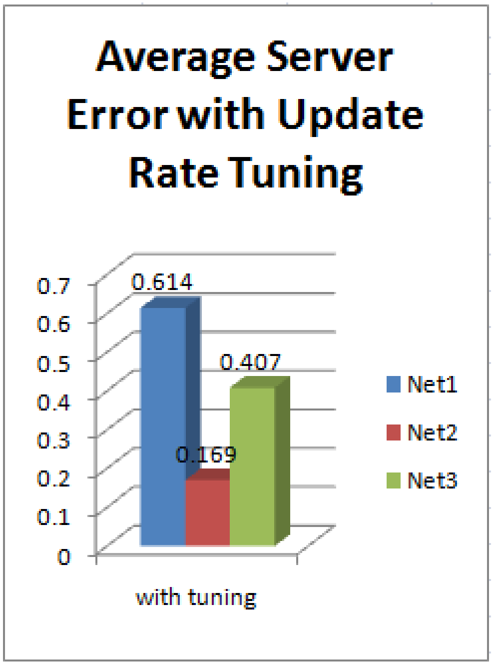

Fig. 4.18 Average server error for the 3 networks of scenario 2 with Update rate tuning

4.4.4 Effect of Update Rate Tuning on the demand on Server and Devices

Our proposed method depends heavily on the periodic updates sent from devices to the server and thus the amount of traffic arriving at the server from many devices might be of concern from a scalability point of view. The same goes for the amount of polls sent to devices. Fig. 4.19 shows these amounts with constant interarrival periods for devices, ie. 3 , 7 and 12 seconds.

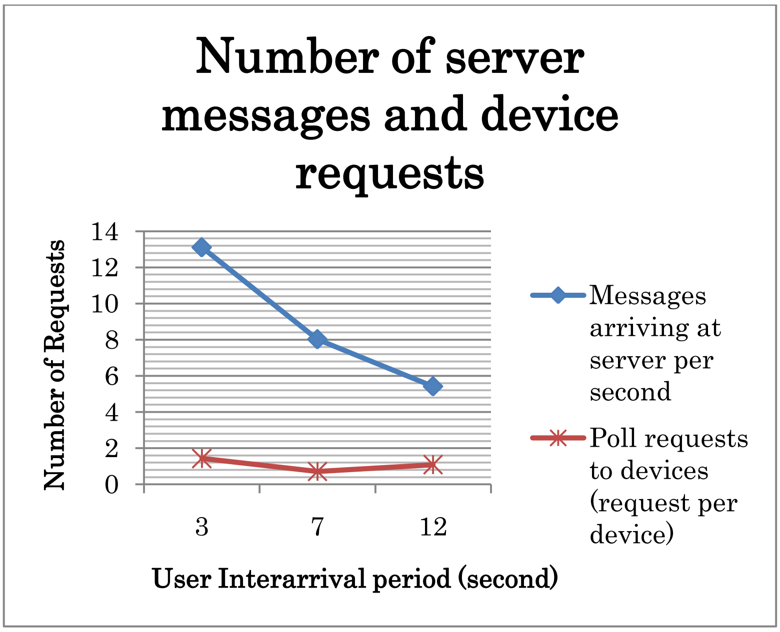

Fig. 4.19 Number of messages arriving to server and devices with different interarrivals (100 devices)

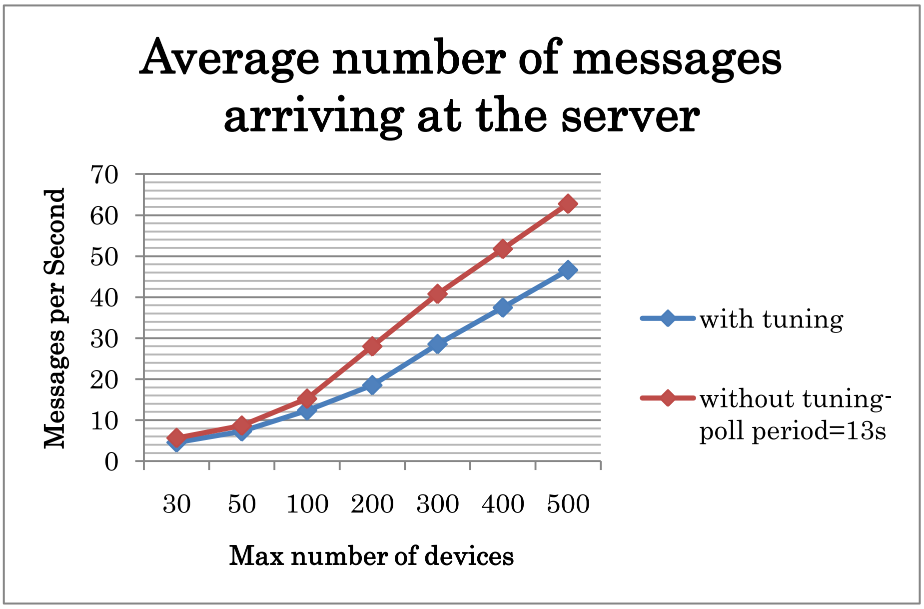

Fig. 4.20 Average number of messages arriving at the server with increasing number of devices

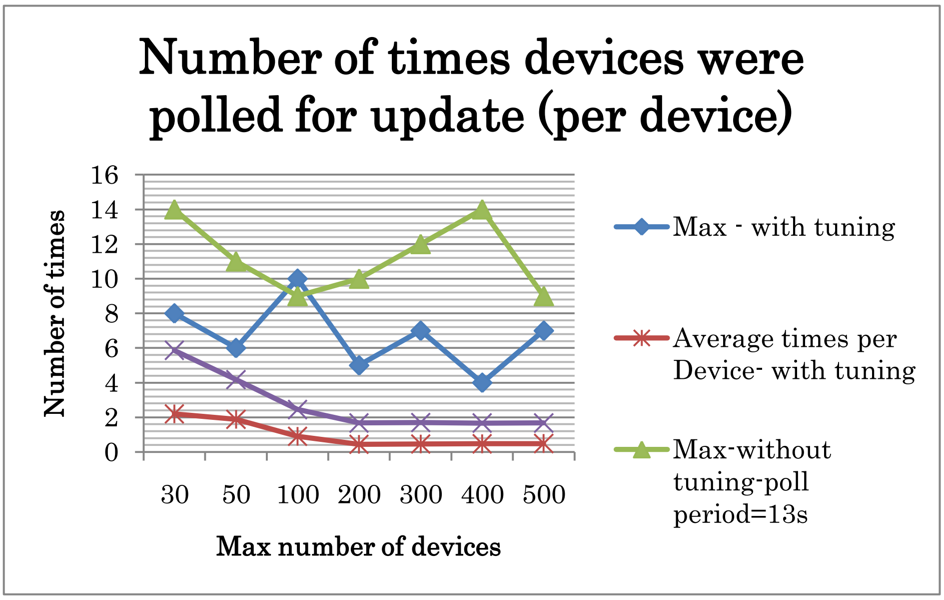

Fig.4.21 Effect of increasing number of users on the number of times devices were polled for update

From Fig.4.20 and Fig. 4.21 it is obvious how using update rate tuning can reduce the traffic on both the server and the device sides. These results were obtained with variable interarrival periods during the simulation time. The change of interarrivals is shown in Fig. 4.22 and the adaptation of update period can be seen on the same graph.

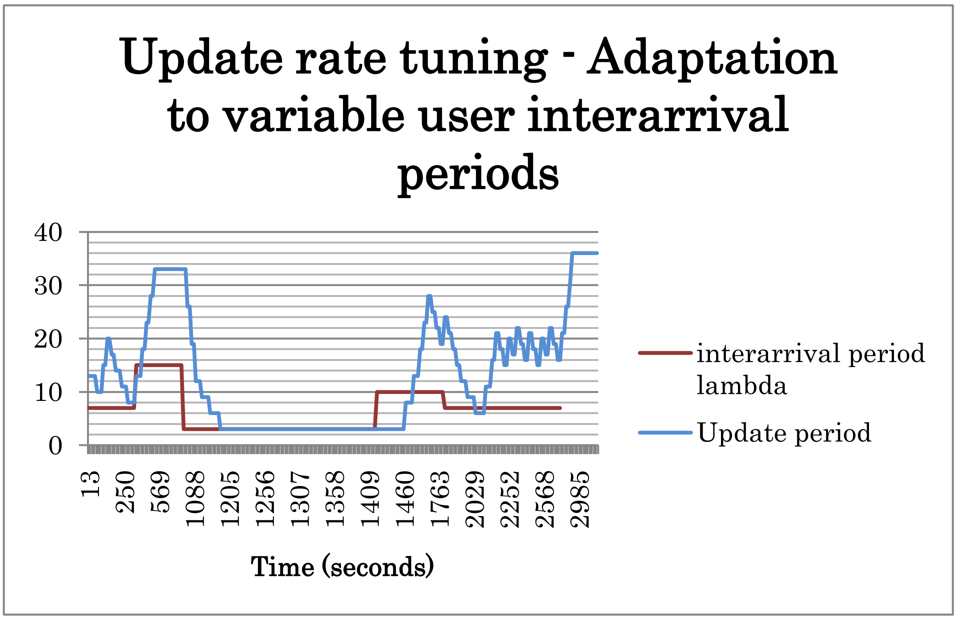

Fig.4.22 Tuning algorithm adaptation to variable user interarrival periods

4.5 Further Considerations

There are still a number of points that need further investigation. These are briefly discussed in this section.

4.5.1 Type of Dynamic Information

In this work, only the network load was considered as dynamic information to be reported to the server and devices. This refers to the number of users connected to a certain network and was used as one of the selection criteria. As might be noticed, this is not always a good indicator on how the network is actually busy since this will depend on the activities of individual users which might be different. In other words, the applications that the users are running have a big effect on the actual degree of how much a network is busy, rather than the plain number of users in that network.

4.5.2 Dynamic Information Update Method

Dynamic information is sent to servers periodically by devices. These devices first wait to receive a poll message from the server before a report message is sent. Another way to implement the information update process is to add some intelligence on the device side that can allow them to report load information to the servers only when there is a significant change compared to the previous update. This way, servers will need less processing power since they don't need to track when and to which device a poll message is to be sent. But, on the other hand, implementing the update functionality in this way would need the devices to decide the reporting order among themselves without the need to contact the server. This might incur more processing needs on the device side and more traffic sent over the network.

4.5.3 User Categories and Information Agents

As was mentioned before, the system is best utilized when there are many users that take network load into consideration.

There are a number of scenarios that can reduce the benefits obtained from the proposed system. For example, scenario where there is a network with no devices connected to it will leave the server with outdated information about that network. The same thing can be said about networks with users only utilizing traditional networks (non heterogeneous devices). Even thought the proposed system can prove to be beneficial for such devices as well, the existence of such system is not crucial for such devices, thus, such devices will not need the system help and accordingly won't be participating in the update process as Information Agents. This will put more burden on devices that are willing to be users of this system, and in the worst case one network will be having only this type of devices leaving the information about that network outdated on the server.

Chapter 5
Conclusions and Future Work

5.1 Conclusions

In this thesis, a system for providing network information both dynamic and static was proposed as a way for network discovery in heterogeneous networks instead of the traditional time and power consuming methods. One type of dynamic information was considered in this work; that is the network load information. By taking this parameter into consideration, devices can achieve better handover decisions that lead to networks that better match their needs.

The effect of this system on the user satisfaction levels was shown by means of simulations that took into consideration many different cases. It was shown that the proposed system can help in achieving higher user satisfaction levels compared to cases where only static information is considered.

Since the system depends heavily on the device's cooperation to periodically send updates about dynamic conditions of the networks, the amount of updates needed to be sent by devices was also explored and shown along with the amount of traffic reaching the server. These results can help to give a rough idea of the amount of requests that the server needs to deal with so they can be taken into consideration for real implementations of the system.

A simple update rate tuning algorithm was also implemented to allow the system to adapt to networks with varying levels of activity. This also helps in reducing unnecessary updates and accordingly reduce the burden on the server and devices.

5.2 Future Work

5.2.1 Including Received Signal Strength Readings (RSS) in Report Messages

As another type of dynamic information, we consider adding the RSS reading as seen by the device from its current location. By including RSS readings in device updates, coverage patterns can be more accurately represented on the server.

The server can have this information and recommend or dis-encourage a handover to some networks based on the device location and the RSS in that location. With this information sent to the server, a map as the one shown in Fig. 5.1 can be constructed.

One obstacle for such a step is the accuracy of localizing methods used by the device. If considering GPS as the localizing method, this approach might be very useful for outdoor environments but will not be accurate in indoor environments.

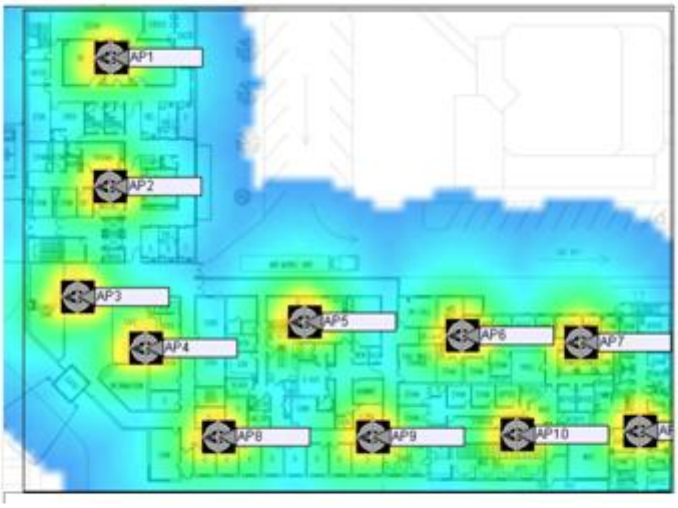

Fig. 5.1 Providing RSS readings as a future work

5.2.2 Adding more metrics to better define the Network Load

As pointed out by the considerations subsection of the previous chapter, only the network load was considered as dynamic information to be reported to the server and devices. And this is defined as the number of users connected to a certain network at a certain time and was used as one of the selection criteria. This plain number of users is not always a good indicator on how the network is actually busy since this will depend on the activities of individual users which might be different. Some users might be running bandwidth demanding applications such as video streaming while others perform simple activities like web browsing. In this case, the number of users will not be representative to the actual network load.

As future work, adding functionality for the devices to estimate that actual throughput that can be achieved by one user is considered. Devices can do this by sending a fixed number of test data that can be used to estimate the actual achievable data rate for some network. This can be combined with the number of users, which is the parameter used in this work, to get more accurate estimations.

References

[1] Anita Singhrova and Nupur Prakash, “A Review of Vertical Handoff Decision Algorithm in Heterogeneous Networks”, Proc. of the 4th Intl. Conf. on Mobile Technology, Applications and Systems, 2007.

[2] C. Liu and C.Zhou, “Challenges and Solutions for Handoff Issues in 4G Wireless Systems, An Overview”, LACCEI 2004, June 2004.

[3] W.-S. Lim et al. ,“Efficient WLAN Discovery Schemes Based on IEEE802.21 MIH Services in Heterogeneous Wireless Networks”, proceedings of GLOBECOM 2008.

[4] Hye-Soo Kim et. al., “Selective Channel Scanning for Fast Handoff in Wireless LAN using NeighborGraph”, ITC-CSCC2004,July 2004.

[5] Q.-T. Nguyen-Vuong et al., “A user-centric and context-aware solution to interface management and access network selection in heterogeneous wireless environments”, Computer Networks, 2008.

[6] Lin C. Jian, Tsai I-Ta and Lee C. Yung, “An adaptive fuzzy predictor based handoff algorithm for heterogeneous network”, IEEE Annual Meeting of the Fuzzy Information, vol. 12, 2004, pp. 944-947

[7] R. Koodli (Ed.), “Fast Handovers for Mobile IPv6,” Internet draft, draft-ietfmipshop-fast-mipv6-03.txt, Oct. 2004.

[8] Youngkyu Choi and Sunghyun Choi, “Service Charge and Energy-Aware Vertical Handoff in Integrated IEEE 802.16e/802.11 Networks,” IEEE INFOCOM 2007, May 2007.

[9] Yaqub, R.; Haq, I.U.; Yahya, K.M.; , "Network discovery approach in heterogeneous radio access networks," Emerging Technologies, 2006. ICET '06. Nov. 2006

[10] V. Gupta et al., “The IEEE 802.21 Standard : Media Independent Handover Services,” IEEE Computer Society (2009).

[11] W.-T. Chen and Y.-Y. Shu, “Active Application Oriented Vertical Handoff in Next-Generation Wireless Networks,” IEEE WCNC 2005, Mar. 2005.

[12] Inoue, M.; Mahmud, K.; Murakami, H.; Hasegawa, M.; , "MIRAI: a solution to seamless access in heterogeneous wireless networks," Communications, 2003. ICC '03.May 2003

[13] Soohong Park, Minho Lee and Cheolju Hwang, “Location Information Look-Up for Intelligent Handover Decision”, 22nd International Conference on Advanced Information Networking and Applications - Workshops, 2008.

[14] Balachandran, A.; Bahl, P.; Voelker, G.M.; , "Hot-spot congestion relief in public-area wireless networks," Mobile Computing Systems and Applications, 2002. Proceedings Fourth IEEE Workshop on , vol., no., pp. 70- 80, 2002

[15] Bejerano, Y.; Seung-Jae Han; , "Cell Breathing Techniques for Load Balancing in Wireless LANs," Mobile Computing, IEEE Transactions on , vol.8, no.6, pp.735-749, June 2009

[16] “Cisco aironet access points features”, Cisco Systems Inc.,2005

[17] Gang Wu, Fanglu Guo, Tzi-cker Chiueh, "Transparent and Accurate Traffic Load Estimation for EnterpriseWireless LAN," Network Computing and Applications, NCA 2007, 12-14 July 2007

[18] PlaceEngine Webpage, http://www.placeengine.com

[19] Amani ABUQDAIS and Hitoshi AIDA, “A Global Network Information Architecture for Handover Support in Heterogeneous Networks”, Information Processing Society of Japan 50th Anniversary, Proc. 72th Conference of IPSJ, Mar.2010

.

List of Publications

[1] Amani ABUQDAIS and Hitoshi AIDA, "A Global Network Information Architecture for Handover Support in Heterogeneous Networks", Information Processing Society of Japan 50th Anniversary, Proc. 72th Conference of IPSJ, Mar.2010