Mobility Decisions in Heterogeneous Wireless Access Networks

Aaditeshwar Seth Nabeel Ahmed Srinivasan Keshav

School of Computer Science
University of Waterloo
Canada N2L 3G1

Abstract: With the proliferation of heterogeneous wireless networks, intelligent systems are likely to be implemented in smart phones that help in making good decisions with regard to selection of appropriate networks and other infrastructure resources. We propose a client-driven multi-factor utility function based framework for making decisions by using different heuristics to maximize the utility. We conduct extensive simulation studies to evaluate these online heuristics against each other, and compare them with an offline utility maximization scheme. Our offline scheme uses a dynamic programming approach that runs on a decision graph to find the best sequence of choices. We also study the interplay between different network design parameters and our proposed schemes, and obtain some interesting results. Apart from our study of heuristics for making mobility decisions, our work also provides important insights with respect to network design parameters in the presence of intelligent client side decision making systems.

Index Terms: Vertical handoffs, decision graph, simulations, dynamic programming

I. INTRODUCTION

With the proliferation of wireless networks and the standardization of many different wireless technologies, heterogeneous access networks for data services are becoming very common. Furthermore, consolidation of these different wireless technologies is unlikely to happen in the near future because each technology is better suited in certain scenarios where other technologies cannot be appropriately applied. For example, WiFi networks are cheaper and offer a better throughput than cellular data networks like EVDO and UMTS. However, the coverage area of a WiFi access network cluster is much smaller, and is therefore not very suitable for data transfer in high mobility scenarios.

For this reason, smart phones are being developed that have multiple interfaces for connecting to different access networks. This is shown in Fig. 1 where a mobile user carrying a smart phone traces a path through an overlay of different access network technologies. Depending on the user preferences and the movement pattern, the smart phone should be capable of making the best vertical handoff [1] decisions of attaching to the most favorable network.

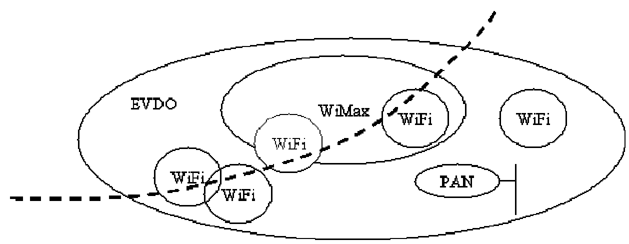

Fig. 1: Mobility across heterogeneous networks

The decision is dependant on many parameters like the movement pattern, network characteristics, user preferences, and application requirements. The decision also depends on the underlying design of the infrastructure for supporting mobility. We attempt to identify such decision parameters and develop a client-side framework for decision making that does not require assistance from the access network. We then evaluate the efficiency of different heuristics to make good handoff decisions, and compare them with each other on varying network scenarios by means of simulations.

In Section II, we describe our assumptions and the various decision parameters that we take into account. We describe our framework in Section III, followed by the simulator model in Section IV. We then present our simulation results in Section V. Related work is mentioned in Section VI, and a discussion on future work is presented in Section VII.

II. DECISION PARAMETERS

We group the relevant parameters in six categories: movement pattern, network characteristics, user preferences, application requirements, device characteristics, and infrastructure related parameters. We then describe our assumptions on the availability of knowledge oracles that can provide knowledge of these parameters.

1. Movement pattern related

User behavior and speed of movement indicate the expected residence time in a network. A large residence time implies a relatively stationary user with respect to the coverage area of the network. For example, if the residence time is large in a WiFi network, then it will be practical to switch to the high throughput WiFi network.

2. Network characteristic related

Different network technologies have widely different per-user throughput guarantees, QoS support features, connection establishment latencies, and economic cost. Each of these factors can play an important role in network selection. Some of these factors may even be time and location dependant. For example, networks in a busy marketplace are likely to be congested depending on the time of day. Different networks of the same technology may likely have different characteristics too.

3. User preferences

Different users have different preferences related to network selection. For example, some users may value cost and would be averse to switching to an expensive good network in the presence of a cheaper albeit poor performing network.

4. Application requirements

Applications may have varying requirements that indicate preferred networks and decisions. For example, streaming applications may have requirements for a certain minimum amount of bandwidth, and would be averse to rapid switching across multiple networks because significant service disruptions may occur.

5. Device characteristics

Since different network technologies have different power requirements, energy constraints on the device will also affect the decisions.

6. Infrastructure related

The architectural design also affects the decisions. For example, if two network providers are partnered with each other and have an integrated authentication system in place [2, 3, 4], then switching between the two providers will be cost effective in terms of minimizing connection establishment latencies.

Another aspect of the infrastructure comes into play with a MIP [5] or HMIP [6] like location management system [7] for network layer mobility. Most such systems have the notion of an anchor point that needs to be notified of IP address changes. Hence, selection of an anchor point that is close to most networks likely to be visited is expected to be beneficial in terms of minimizing the RTT latencies for the IP address updates.

A similar reasoning also applies when anchor points are considered to be proxy storage servers as suggested in the Tetherless Computing Architecture [8]. A closer anchor point will have a lower RTT, which will result in a better throughput [9] between the anchor point and the mobile receiver. However, frequent anchor point changes will bring about considerable thrashing by shuffling the context and data from one anchor point to another. In this study, we only evaluate the stand-alone mobility anchor points, and do not consider any context or data switching costs that may be incurred if the anchor points would store data as well.

A. Decision choices

To summarize, decision making has to occur on two fronts whenever multiple choices are available.

1. What is the best network to choose?
2. What is the best anchor point to choose?

All the parameters enumerated above need to be analyzed in light of these two choices.

B. Assumptions

1. Movement prediction

Intelligent systems can be implemented that make use of user behavior and movement patterns to predict the mobility path of a mobile user. Coupled with knowledge of the coverage areas of different networks, such a system can predict the expected amount of residence times in different networks.

We assume the existence of such a system that can predict user movements in terms of network traversal within a certain probability of accuracy.

2. Prediction of network characteristics

Past history can potentially be used to predict network characteristics like the expected throughput. For example, genetic programming algorithms can help in establishing relationships between the throughputs of a network, and the location and time of day – networks along a busy street are likely to be congested and provide low throughput during evening hours.

We assume that such a system exists that can predict knowledge of the expected network characteristics within a certain probability of accuracy.

3. Round-trip-time estimates

We have assumed that movement in geographical space is predictable to a certain extent. If the geographical coordinates can be mapped to the expected RTTs between different access networks and anchor points, then this knowledge can potentially be used to select the best anchor point. In Section VII, we describe our efforts towards the development of a system that can establish mappings between geographical coordinates and Vivaldi [10] or GNP [11] like network coordinates.

Although we do not assume the existence of such a system for our study, we do assume that all access networks and anchor points are within 100ms RTTs from each other. We expect that this will cover most realistic scenarios because anchor points are likely to be collocated in the same city or region as the access networks under consideration, and therefore likely to be close to each other in network coordinate space as well. For the same reason, we do not consider any variation in the throughput due to changes in the selection of anchor points, because we assume the RTT to the sender >> RTT to the mobility anchor points.

III. Decision Making Framework

We propose a client-side decision making framework that can support multiple applications and decision heuristics. Such a framework can be implemented as an independent module on a smart phone drawing its inputs from the movement and network prediction systems. We also describe a full knowledge offline scheme that can be used as a yardstick to evaluate the performance of different online decision heuristics against each other. The decision making process is based on the utility maximization approach described below.

A. Utility based approach

Utility functions similar to those used for econometric analysis are ideal tools to specify desired preferences when multiple factors are involved in the decision making process. For example, such functions can be used to relate the amount of throughput in a network with the economic cost incurred by utilizing the network. Utility functions can even be application dependant, as shown below.

1. Data-based applications

The goal of such applications is to maximize the amount of data received. Therefore, the utility acquired while present in a certain network can be expressed as:

$$U_i = \textit{throughput} \times (\textit{residence time} - \textit{switching latency}) - k_1.\textit{cost} - k_2.\textit{power}$$

Here, k_1 and k_2 are user preference factors that express the value a user attaches to the cost versus amount of data transferred. A similar utility function is defined in [14] where the relationship between the cost and throughput parameters is logarithmic.

2. Streaming applications

Such applications are adversely affected by high switching frequencies that cause undesirable service disruptions. Therefore, the utility function for such applications can be expressed as:

$$U_i = -\textit{switching latency} - k_1.\textit{cost} - k_2.\textit{power}$$
$$\textit{For throughput} > \textit{a certain minimum threshold}$$
$$U_i = -\textit{infinity}$$
$$\textit{For throughput} < \textit{minimum threshold}$$

In both the examples above, U_i is stated as being the utility acquired by switching to the i^{th} network. If the movement prediction system is able to yield mobility across multiple networks in the future, then the general utility function can be stated as:

$$U = \sum U_i$$

Maximization of the utility function will yield the best choice of networks to be selected according to user mobility.[1] Instead of a maximization heuristic, other policies can also be used for decision making that depend on the applications at hand, and try to maximize the overall utility approximately based on the current available knowledge.

Note that the utility functions of different applications can be combined into a single function by attaching priorities to each application. The priorities can even be increased temporarily if an important application becomes active in-between.

Such functionality can be implemented as a Connection Manager (CM) module that lies in a shim layer between the TCP and IP stacks on the client device. The CM will intercept all incoming and outgoing packets, and tunnel them from/to the appropriate network interfaces. The implementation can be similar to the MSocks [12] or TCP Migrate [13] approach, which supports connection migration across access network interface changes.

We have not implemented the CM for the simulations illustrated in our study – we conduct our experiments for a single application only. We feel that extension of our studies to multiple applications is an implementation issue where user preference factors will exercise the

[1] The utility function can be made probabilistic as well, with the probability based on the accuracy of prediction of the movement and network parameters.

greatest effect in deciding the interplay of different applications with each other.

B. Offline full knowledge scheme

If a scheme has accurate and future knowledge of all the parameters that constitute the utility function from the start to the end, then maximization of the complete utility function will give the sequence of best choices to be made. We find that we can represent this as a decision graph where each edge weight represents a component of the overall utility. Hence maximization of the utility corresponds to finding the longest path in the graph.

An example decision graph is shown in Fig. 2. An EvDO network is always present along the line of mobility of a user. The user enters a WiFi network coverage area at time T_1 and leaves it at time T_3. Similarly, a Bluetooth network is encountered at time T_2. T_0 and T_4 simply indicate the start and end of the graph, which can be regarded as being analogous to the activity period of an application. For simulation studies, we consider the application to be active throughout the simulation run. Hence, T_0 and T_4 represent the start and end of the entire simulation run. Based on the edge weights assigned to the decision graph, a utility maximization algorithm will give the sequence of decisions that the user should make for getting the best performance.

The graph can be visualized more generally as follows.

1. External events like when the user enters or leaves the coverage area of a network define the instances of time when a decision has to be made.
2. Each node represents a switch event from/into another/same network. It can be uniquely represented as a tuple of (*network*, *time*). Note that all networks are considered separately, irrespective of whatever network technology they belong to.
3. Each edge represents a decision of either staying in a network or switching over to a different network.
4. The weights carried by the edges are a component of the utility derived by traversing the edge. For example, considering:

   U_i = *throughput* (*residence time – switching latency*)

   Then:
   $w_1 = \quad throughput_{evdo}(T_1 - T_0)$
   $w_2 = \quad throughput_{evdo}(T_2 - T_1)$
   $w_3 = - throughput_{evdo}(L_{wifi})$
   $w_4 = \quad throughput_{wifi}(T_2 - T_1 - L_{wifi})$
   …

   Where L_{wifi} = *latency of switching into Wifi*. This includes the connection establishment time and the authentication delay, and is assumed to be constant.

5. Voluntary disconnection decisions can also be supported by adding extra nodes that represent voluntary disconnections.
6. The graph can be expanded to include anchor point selection decisions as well. Each node will then be uniquely represented as a tuple of (*network, anchor point, time*).
7. Other external events like throughput changes in networks can also be included easily.
8. If the latency is not a constant, but is also dependant on the previous source and destination network pairs, then we do not mention this component in the decision graph but we take it into account when evaluating the longest path in the graph. Such a case may arise when the latency also includes anchor point update delays. The dynamic programming approach explained below allows us to consider such *path dependant* edge weights easily.

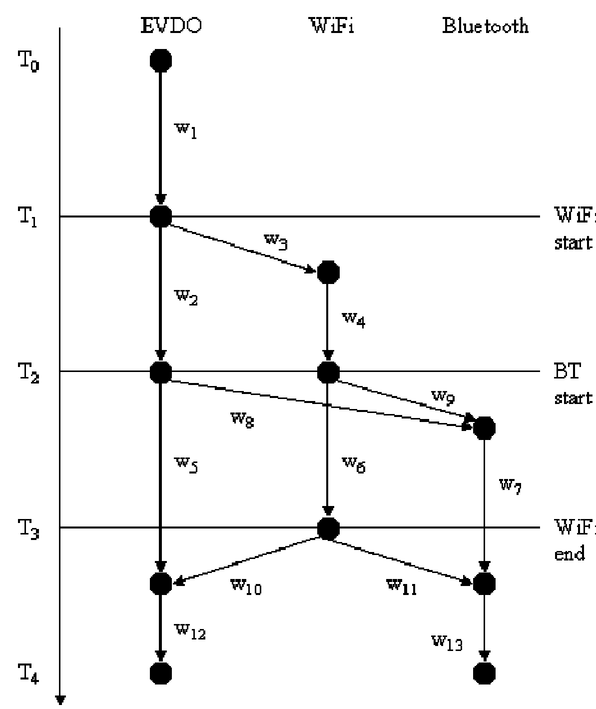

Fig. 2: Full knowledge decision graph

It is interesting to note that the decision graph is a Directed Acyclic Graph (DAG) that is topologically sorted. A linear time dynamic programming algorithm exists that can find the longest path in a DAG [15]. The algorithm is based on the following theorem.

If a path ($v_1 \ldots v_{n-1}\ v_n$) to a node (v_n) is the longest path in the DAG up to that node, then the path ($v_1 \ldots v_{n-1}$) to the previous node (v_{n-1}) is the longest path in the DAG for the previous node ($v_1 \ldots v_{n-1}$).

Therefore, the longest path computation simply amounts to a traversal of the graph in topological order, while also maintaining pointers at each node to the previous node in the longest path up to that node. Summation of the edge weights on the longest path gives the maximum amount of utility that can be derived by the best sequence of decisions. Since we only need to maintain one maximal path for each node, the dynamic programming algorithm allows us to take edge weights into account that depend on the maximal path taken up to that node. This permits us to model the *path dependant* latencies mentioned earlier.

We use this result to serve as a yardstick to compare different heuristics that have access to lesser and inaccurate amounts of knowledge.

C. Online heuristics

Heuristic based decision making is triggered through external events like the start of a new network or the end of the current network or a throughput change in the current network. Signal strength measurements can be used to determine whether a mobile device is within range of a network or not. We assume that online schemes have access to only the knowledge of parameters of the current networks surrounding a mobile user. We leave the design of heuristics based on future knowledge to extensions explained in Section VII. Note that all the online schemes are still subject to the prediction errors incurred by the movement and network characteristic prediction systems for all the current networks.

For the purposes of this study, we only experiment with the data oriented application illustrated in the previous section, and do not consider the cost or power parameters. We conduct simulations on the following decision heuristics suited for such data intensive applications.

1. Maximum throughput, closest anchor (MTCA)

This scheme selects the closest anchor point, and the network with the maximum throughput. MTCA is an aggressive greedy scheme, and is likely to incur large switching costs due to a high frequency of switching.

2. Maximum throughput, same anchor (MTSA)

This scheme is similar to MTCA except that instead of selecting the closest anchor each time, it chooses an anchor point at the beginning and sticks to it for the entire duration of the simulation run.

3. Closest anchor (CA)

This scheme selects the closest anchor each time, but tries to lower the frequency of network changes by switching to the network with the highest value of *throughput* x (*residence time – switching latency*), thus assuming to remain in the network for the rest of the residence time.

4. Same anchor (SA)

This scheme works similar to CA, but chooses an anchor point at the beginning and sticks to it for the entire simulation run.

In order to minimize the thrashing that aggressive schemes may cause, we use the concept of makeup time first introduced in [14]. The makeup time is defined as the minimum amount of time that a user needs to remain in a network to make up for the utility lost whilst making a switch into that network. If we assume that no utility is gained in the actual process of making a switch, the makeup time can be calculated from the following equation.

$$U_{current}\,(T_{makeup} + L) = U_{better}\,T_{makeup}$$

Note that we consider vertical handoffs as being hard handoffs and do not assume anchor points to do bi-casting as networks are changed. Therefore, no utility is gained during the switching process. The switching latencies are calculated as being the sum of connection establishment latencies, authentication delays, and the update latency incurred to inform the anchor point of the network change. If the anchor point is changed, it also includes the latency incurred to inform the sender of the anchor point change.

D. Effect of network design

We compare the performance of the online heuristics with the offline full knowledge scheme in various simulation scenarios. The simulation scenarios are categorized on the basis of the following parameters.

1. Number of anchor points

Intuitively, a large choice of anchor points will provide more scope for improvement by making good choices, but the improvement should follow the law of diminishing returns as further increasing the number of anchor points does not bring about a corresponding increase in performance.

2. Mobility domain size

All access networks within the geographical domain of a single provider are likely to have the same RTTs to the anchor points. Hence, mobility across networks within the same provider domain should not require a change in the anchor point – the closest anchor point can be selected for each domain, and retained for all movements within the domain. Such a scenario may arise in a corporate building where all WiFi networks would be under the same provider. Intuitively, larger the mobility domain size, lesser will be the number of anchor point changes, and hence, the more will be the utility.

3. Density of high throughput networks

Intuitively, more the density of high throughput networks, lesser will be the amount of time spent in lower throughput networks.

4. Probability of error

Intuitively, more the probability of error, more will be the chances of selecting a bad network, and subsequently, more will be the switching costs incurred in moving from bad networks to good networks.

We study the interplay between these parameters and the decision processes, and find some interesting trends that we later explain in our simulation results section.

IV. Simulation Model

A. Simulator implementation

We have implemented a discrete-event simulator in Java. The simulator implementation is divided into two parts. The first part is involved with the placement of networks over a grid of specified size. It also facilitates the generation of external events according to the mobility pattern of the mobile user. The second part involves the actual discrete event simulations.

1. Network setup

a) Network and anchor point placement

In order to facilitate the creation of a generalized network map, we have implemented a tool that randomly places all networks on a 2-dimensional grid of specified size. The number and properties of each particular kind of network technology (e.g. network bandwidth, coverage area, number of networks), must be supplied as inputs to the tool. Network coordinates are assigned to the access points of each network to calculate the RTTs for switching latencies. In order to simulate realistic network scenarios, we have also defined the notion of a *network provider*. All same technology networks owned by the same provider have the same network coordinates. This is likely to be the case because network providers will most generally partner with the same ISP, and hence will have very similar network coordinates.

Anchor points are not placed physically on the grid, but are assigned network coordinates randomly. The number of anchors points must be specified as an input.

b) Mobility pattern

We use the random walk mobility model to define the movement of the mobile user over the network map. We do this by specifying multiple mobility segments; i.e. a movement over a straight line between any 2-points in our 2-dimensional network grid. These segments, when combined, define the complete movement pattern of the mobile user.

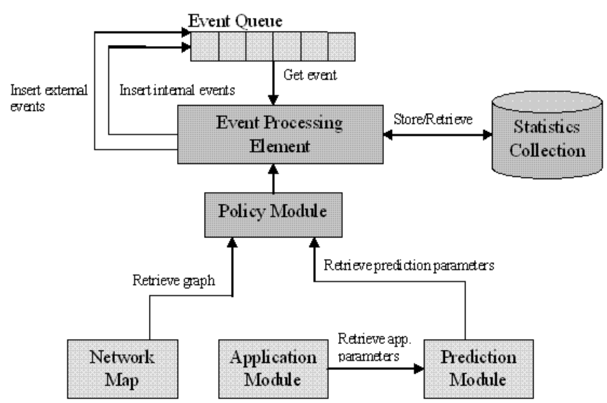

Fig. 3: Simulator Architecture

2. Simulator Components

Fig. 3 illustrates the overall architecture of the simulator. Each of the individual components and their specifics are discussed further:

a) Event processing element

The event processing element is the kernel of the simulator and is the starting point in the execution of the simulation. It uses information from other modules as input and processes each of the events in a sequential fashion. The simulator recognizes 3 basic types of external events:

1. *Start*: Defines the time when a user enters a new network. This is a global event that triggers an instantaneous adaptation phase to evaluate the benefit from switching into the new network.
2. *End*: Defines the time when a user moves out of a network. This is localized event that occurs in the

current network, and triggers an adaptation phase to find the best network to switch into.

3. *Throughput change*: Occurs periodically in the current network, and triggers an adaptation in the current network to find a better network to switch into. We use a Markov state model to perform throughput changes in a given network, where each state represents a certain throughput value and switches can occur between two consecutive states with a probability of 0.5.

Each of the external events can trigger an instantaneous adaptation event, or can schedule one in the future.

1. *Adaptation*: Defines the time at which a mobile adapts to changes in the network state (e.g. throughput change, network end, etc).

b) Policy module

The policy module provides the implementation for each of the online heuristics. The implementation of the different online heuristics and the full knowledge schemes is discussed further:

i) *Optimal scheme*

The optimal scheme is the only scheme that is run in *offline* mode. It uses accurate information (e.g. throughput, residence time) about all network states to compute the best path for utility maximization.

ii) *Maximum throughput, closest anchor (MTCA)*

MTCA runs in online mode and uses instantaneous knowledge of the current network state that it obtains from the *prediction module* to compute the best network to switch to. In our study, we are assuming that the mobile users already have knowledge of the network co-ordinates of the anchor points in the network. The closest anchor to switch to then simply reduces to a distance calculation problem between the mobile and the anchor point.

iii) *Maximum throughput, same anchor (MTSA)*

MTSA follows the same principles as MTCA, but the anchor point needs to be specified in advance of running the simulation.

iv) *Closest anchor (CA)*

CA also runs in online mode, but invokes the application module to obtain information about the *remaining utility* for any potential network that it is currently evaluating during its decision making phase.

v) *Same anchor (SA)*

SA evaluates each network using the same approach that CA employs, but it uses a single anchor point through-out the simulation run.

As can be seen, the CA, and SA schemes account for application flexibility by allowing the application to specify its own choice of utility function. This can be achieved by means of the application module, which is discussed next.

c) Application module

The application module interfaces with the prediction module and specifies the utility function and parameters that are important based on the application of interest. In our simulations, we assume that only one data-oriented application is running throughout the simulation run.

d) Prediction module

The prediction module has been implemented as a black-box tool so that actual prediction of parameters need not be done. Since we primarily focus on the evaluation of the policies themselves, we implement the prediction module by probabilistically introducing an error in the true value of the parameters. This simulates the effect of an arbitrary prediction scheme that attempts to estimate the true value of a parameter, but makes mistakes.

e) Network map

The network map component is used to query the output generated by the mobility pattern and graph construction tool. It is used by the optimal scheme to construct the decision graph, and can also be used to output a partial decision graph of the future for the online schemes. However, the event graph is currently not being used for the online schemes and has been left as a part of future work.

f) Statistics Collection

Statistics are gathered throughout the simulation and are stored in a separate statistics object that is later referenced to output the results for each simulation run.

B. Simulation Metrics

Three metrics have been chosen to evaluate each of the online heuristics against the full knowledge scheme.

a) Achieved utility

This is the overall utility derived by the schemes during the entire simulation run. For the data-oriented application that we have simulated, this is measured in terms of bytes/sec.

b) % residence time

This is the percentage of overall simulation time spent in a particular network or technology. We use this metric to analyze the behaviors of the heuristics with respect to the different network design parameters discussed in Section III D. This metric also allows us to study the correctness and accuracy of each of the schemes in selecting the best network possible, by, for example, validating that the time spent in low throughput networks is less than the time spent in higher throughput networks.

c) Switching frequency

We use this metric to primarily study and explain the stability or adaptation behavior of the different heuristics. It is measured as the number of network or anchor point switches/hour.

V. SIMULATION RESULTS

After careful consideration, we have designed the following scenarios within which to test our schemes:

1. *Moving between buildings with single technology* The network grid is divided into smaller mobility domains for each provider, and all networks within the same domain are assumed to belong to a single provider having the same network coordinates (Fig. 4). Each mobility domain picks up an anchor point closest to it in terms of network coordinates. This is used to study the impact of the number of anchor points and the mobility domain size on the utility and the frequency of anchor point and network switches. This is analogous to walking between buildings where each building is under a different provider and the only technology available is WiFi.

2. *Walking in a large building with two technologies and a single provider*
   A small grid-size is chosen containing a cloud of EVDO coverage and random layouts of WiFi networks (Fig. 5) belonging to the same provider. This scenario is used to study the impact of network density on the attained utility and the frequency of network switches. This is analogous to walking in an office building where multiple WiFi access points are present and a cloud of EVDO coverage is also available.

   We find that the simulation results and trends are very similar to those exhibited by next scenario, and only present those results later.

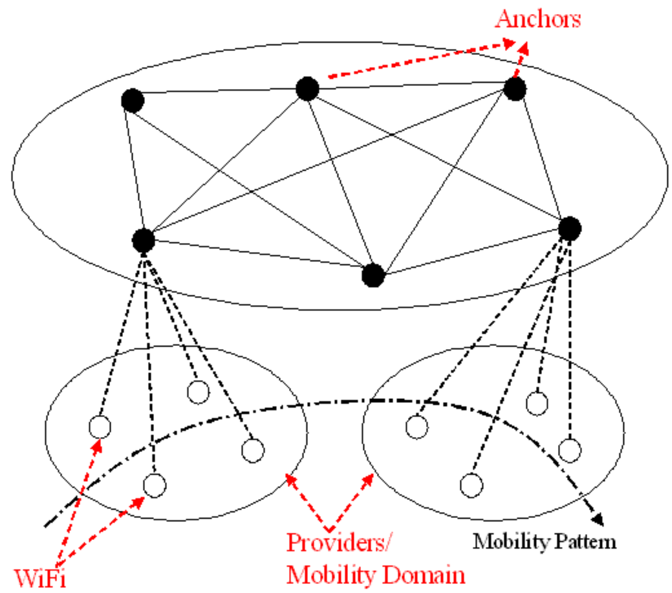

Fig. 4: Moving across multiple mobility domains

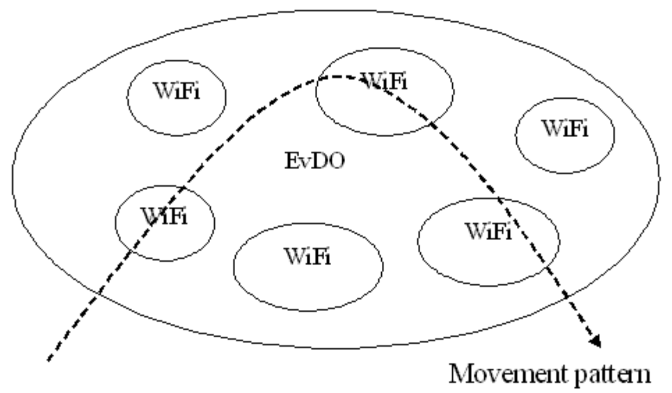

Fig. 5: Moving in a building

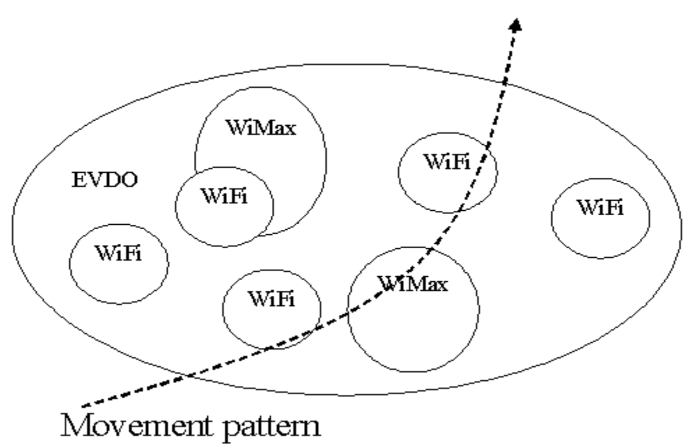

Fig. 6: Moving in a city

3. *Walking in a large building with two technologies and multiple providers*
Scenario 3 is identical to the previous scenario with the added exception that multiple providers are present, and the networks are assigned randomly to the providers. In this case, we want to analyze the effect of increasing the number of anchor points and providers on the network density and frequency of networks switches. This is analogous to walking in a large shopping mall.

4. *Moving in a city with multiple technologies and multiple providers*
A large grid-size is chosen containing multiple network technologies that are spread out randomly over the grid, along with random assignment of providers (Fig. 6). We use this scenario to analyze the performance of the schemes against the accuracy of the prediction module. This scenario is analogous to driving thru a city downtown that contains multiple access technologies.

All simulations are conducted by averaging the results over 5 different network maps, with 10 random walks on each map.

A. Moving between buildings with single technology

A 5000m x 5000m map is constructed having only WiFi networks (2500 access points) with coverage area diameters between 50m and 100m. The number of providers in this scenario is 20. We assume the per-user throughput in WiFi networks to vary between 400kbps and 1mbps. The map is divided into square mobility domains where all the networks within a mobility domain belong to the same provider. The probabilistic error in prediction of network parameters is assumed to be 0.

a) Network and anchor point switches

Fig. 7 plots the frequency of anchor point switches against the mobility domain size, with the number of anchor points fixed at 7. As would be expected, the frequency decreases with an increasing mobility domain size because the closest anchor point can be retained while roaming within a mobility domain to reduce the update latencies to the anchor point. This also serves as a validation for our simulations because the drop-off in *switching frequency* is roughly inversely proportional to the *mobility domain size*, which is expected because the *switching frequency* will be proportional to the *number of mobility domains* (= *grid size* / *mobility domain size*).

Fig. 8 plots the frequency of anchor point switches against the number of anchor points, with the mobility domain size being fixed at 600m. It is interesting to see that the switching frequency required for good performance saturates at around 10 anchor points. This indicates that over provisioning of anchor points may not offer any substantial advantage when the anchor points are all collocated within small RTTs from the access networks.

For validation purposes, we also plot the network switching frequency with respect to the number of anchor points and the mobility domain size, and observe the switching frequency to be largely insensitive to these parameters as expected.

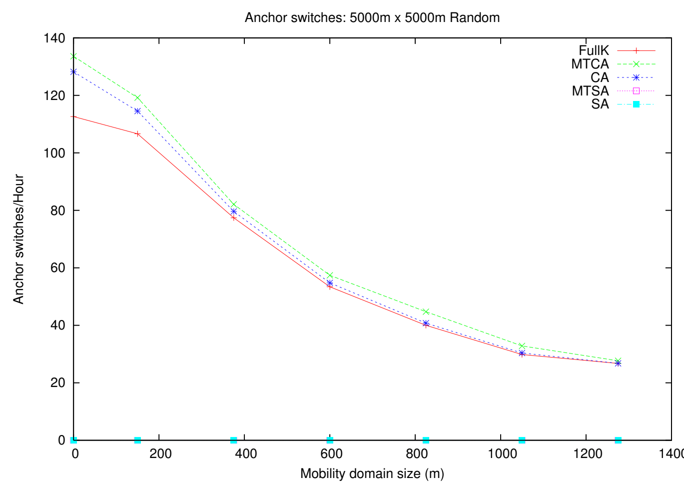

Fig. 7: Anchor switches against mobility domain size

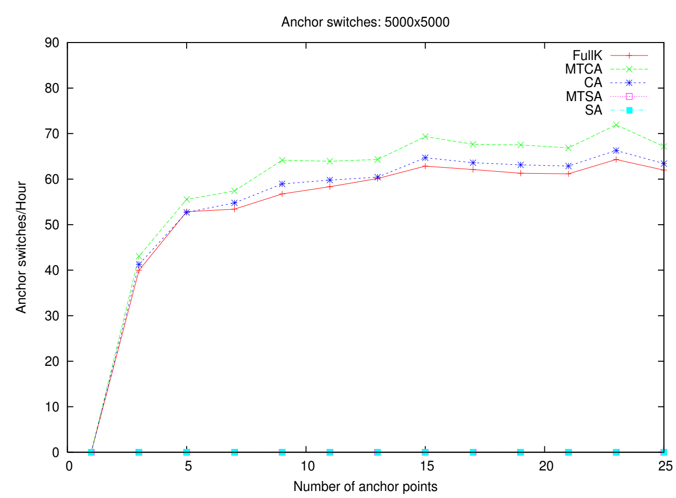

Fig. 8: Anchor switches against number of anchor points

b) Utility

Fig. 9 and Fig. 10 plot the achieved utility against the number of anchor points and the mobility domain size. It is observed that the utility is insensitive to either of the two parameters because the difference caused due to a better selection of anchor point is largely insignificant because of the minor differences in RTT update delay latencies (< 100 ms) involved. These results indicate that staying with the same anchor point throughout might do just as well as selecting the closest one to the mobile.

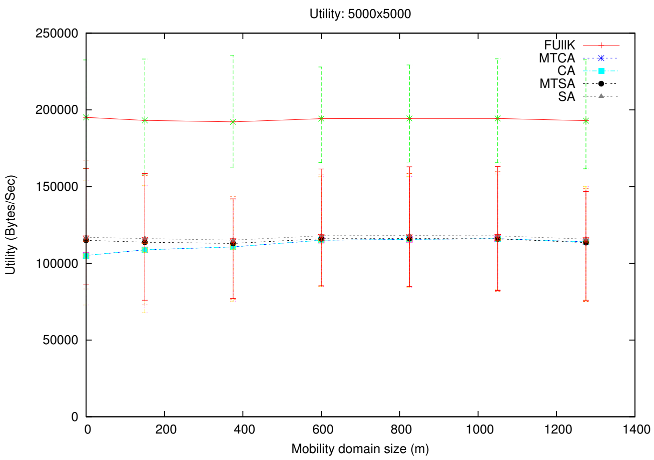

Fig. 9: Utility against mobility domain size

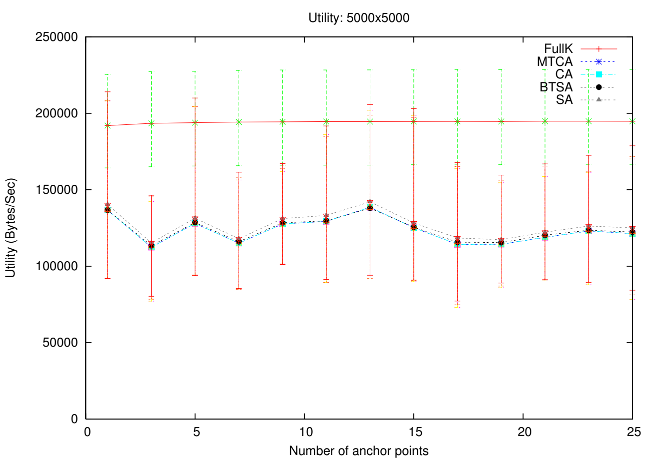

Fig. 10: Utility against number of anchor points

B. *Walking in a large building with two technologies and multiple providers*

A 500m x 500m network map is constructed that has a complete EVDO coverage, along with a high density of higher throughput WiFi networks. The EvDO per-user throughput is assumed to lie between 100kbps and 750kbps. 7 anchor points and 20 providers are considered in this scenario. The probabilistic prediction error in network parameter values is assumed to be 0.

a) Utility

Fig. 11 plots the overall utility against an increasing density of WiFi networks. As shown by the figure, the overall utility of the user increases with an increase in the network density, but eventually saturates when complete network coverage is realized. The MTCA/MTSA schemes perform close to the optimal and do much better than CA/SA. This indicates that in such scenarios, aggressive opportunistic switches to the highest throughput network outweigh the cost associated with the switching latency. This is noticed because we consider the switching cost to be dependant only on the latency incurred. However, if native TCP like schemes are modeled without any mobility enhancements, then the switching costs should be greater due to the slow start phase that begins after conservative cut-downs in the window size are made during large network switching latencies. We do not model such scenarios because extensive research has already been done on TCP enhancements over mobility, and various existing mechanisms can be used to keep the switching costs low.

We also observe that the MTSA scheme does marginally better than the aggressive MTCA scheme. Coupled with results from the previous scenario, this shows that when networks with the same closest anchor point are collocated close to each other, then same anchor point and closest anchor point schemes perform relatively similarly (Fig. 9). However, when networks can have randomly different anchor points closest to them, then stable schemes that choose the same anchor point perform slightly better than the closest anchor point schemes (Fig. 11).

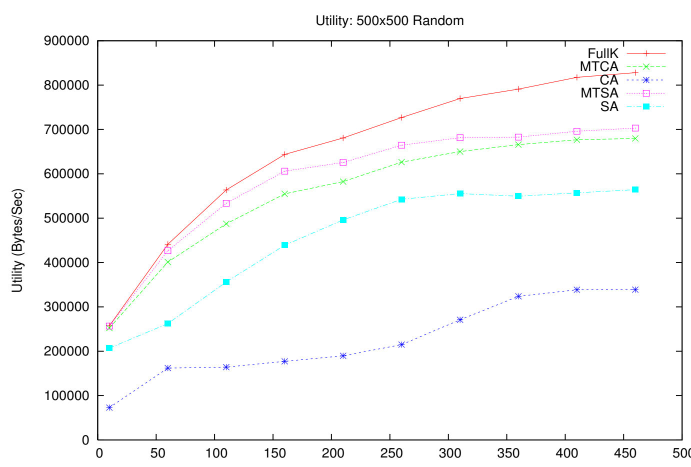

Fig. 11: Utility against network density

b) Per technology network switches

Fig. 12 plots the number of network switches per hour into EvDO and WiFi networks. We notice an interesting behavior that the switching frequency initially increases as the number of WiFi networks is increased, but starts dropping later. We explain the network switching behavior in Fig. 12 by dividing the graph into a set of zones. Using this approach, zone 1 starts at the graph's origin to the point where EVDO switches/hour peak. Zone 2 extends from the end of zone 1 up to the point where the WiFi curves peak. Likewise zone 3 starts from the end of zone 2 and progresses to the point at which the EVDO curve approximates to zero. Finally, zone 4 proceeds from the end of zone 3 onwards.

In zone 1, we see that with a limited number of WiFi networks, the mobile rapidly switches between EvDO and WiFi, causing an increase in both WiFi switches and EvDO switches. By consulting the corresponding percentage residence time graph per technology (Fig. 13), we also see an exponentially fast decrease in the residence time in EvDO and a likewise increase in the residence time of WiFi. The peak of EvDO switches (Fig. 12) corresponds to the crossover point in the residence time spent in WiFi and EvDO (Fig. 13).

In zone 2, the larger increase in the number of WiFi networks causes the mobile device to stay in the WiFi networks more often, causing a reduction in the EVDO switches and an increase in the number of WiFi switches (due inter-WiFi switching). This is seen as a further decrease/increase in the corresponding residence times of EVDO/WiFi.

Zone 3 indicates a region where the number of WiFi networks with larger average coverage areas increases, causing a decrease in the number of both WiFi and EvDO switches. We verify our intuition for the selection of larger coverage area networks through Fig. 14 which shows the utility breakdown derived by grouping networks into bins of 50 seconds each for the residence time spent in each network. We see that as the WiFi networks increase, greater utility is obtained from networks with larger residence times. Interestingly, we also note that the intersection point of (residence time >100s) networks cuts at the same point at which the original WiFi switch curves had peaked (Fig. 12).

Zone 4 defines the region in which EVDO coverage is completely eliminated and a greater number of larger coverage area WiFis networks reduce the effective number of switches in the system.

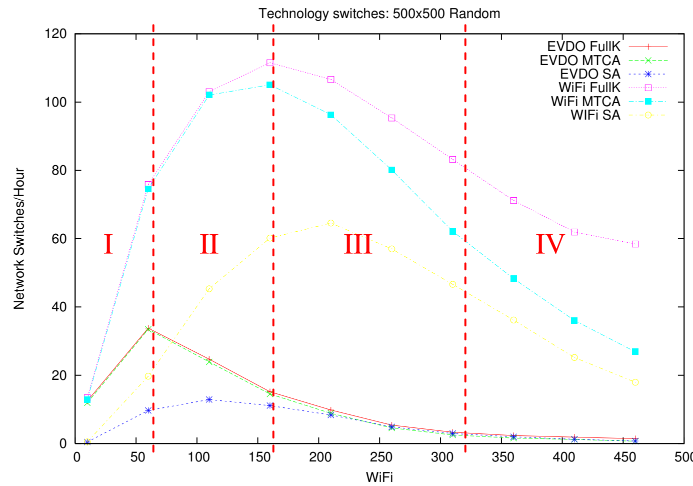

Fig. 12: Network switches against network density

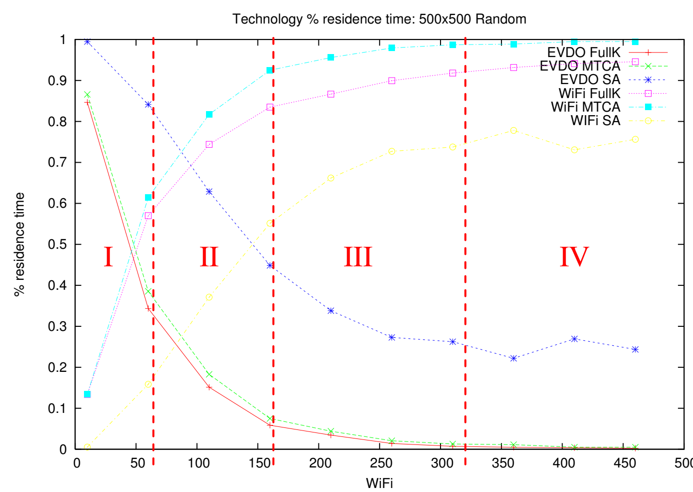

Fig 13: Percentage residence time per technology against network density

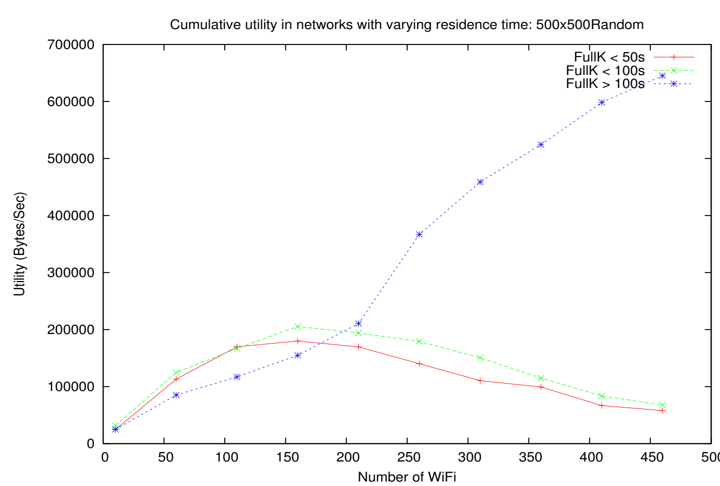

Fig 14: Percentage residence time per network

c) Residence time

Fig. 13 plots the percentage residence time of the mobile user in each network technology. We see that as the number of WiFi networks increase, the percentage residence time in WiFi increases whereas the percentage residence time in EVDO decreases. This serves as a good validation for our simulator since we expect that as the number of high-throughput networks increase, the mobile selects networks with higher throughput. We also see that MTCA has better selection accuracy than CA which is more sluggish in its response to selecting WiFi.

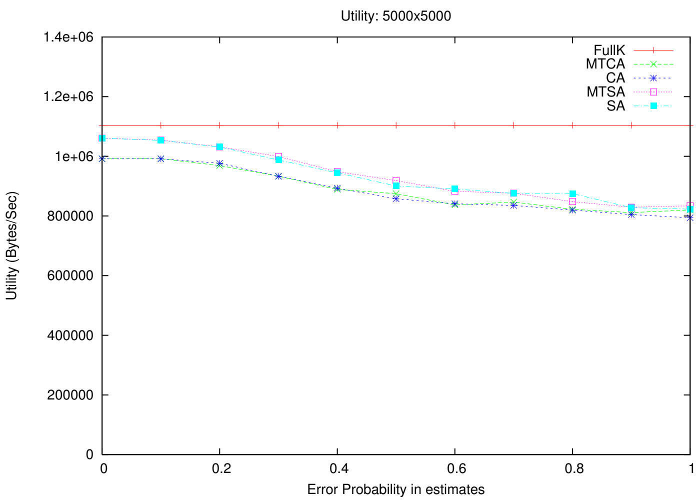

Fig 15: Utility against error probability

C. *Moving in a city with multiple technologies and multiple providers*

A 5000m x 5000m network map is constructed, having EVDO, GPRS, WiMax, and WiFi coverage. We use 7 anchor points and 20 providers in this scenario. We use the following number of networks for each technology: WiFi=2500, GPRS=15, EVDO=15, WiMax=25. We assume the WiMax throughput to lie between 500kbps and 2mbps, and the GPRS throughput to lie between 5kbps and 10kbps.

a) Utility

Fig. 15 plots the utility against the error probability in the prediction modules. Oddly enough, we see that even a dramatic increase in the prediction error does not affect the utility of the online schemes to a great extent. We believe this to be a consequence of the values of simulation parameters we have modeled. In the present model, we introduce errors in the decision computation by altering the observed throughput and expected residence times in different networks by a factor equal to the error probability plotted on the x-axis in Fig. 15. However, since the throughput ranges of the networks do not considerably overlap with each other, hence even the introduction of errors results in the same decisions made by the adaptation phases. For example, the expected throughput of GPRS is between 1kbps to 10kbps and the expected throughput of WiFi is between 200kbps to 1mbps. Hence, even if the throughputs are reported erroneously, WiFi will always have a better throughput and will be considered as a more favorable network than GPRS.

We do however observe that as the error probability increases, all schemes converge to a similar performance in terms of utility. Therefore, we believe that the relative performance of any of the schemes is highly dependent on the accuracy of the prediction scheme used.

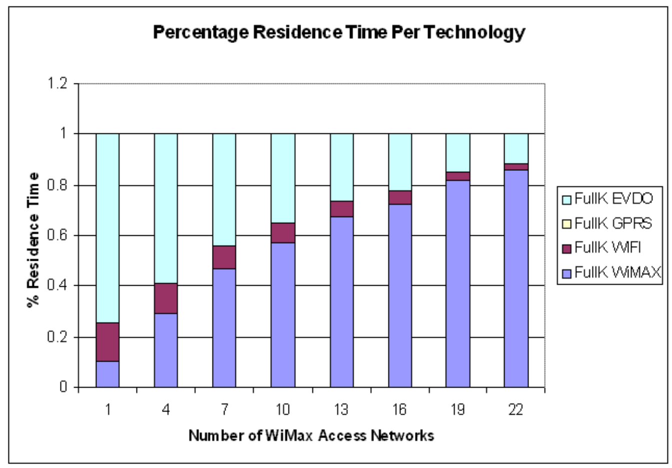

Fig 16: Residence time per technology with increasing WiMax networks (FullK)

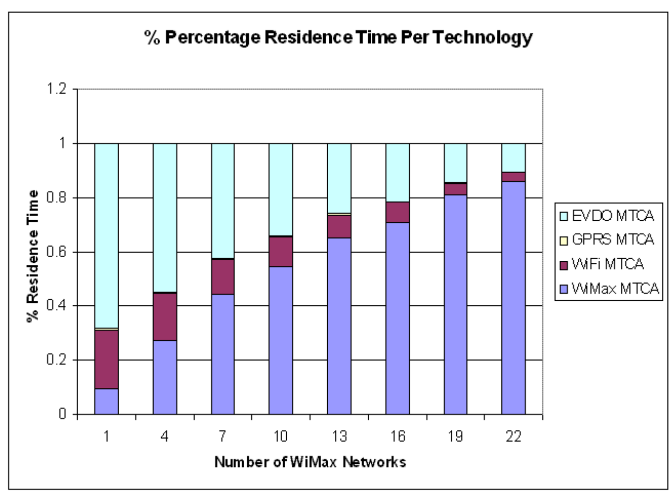

Fig. 17: Residence time per technology with increasing WiMax Networks (MTCA)

b) Residence Time

Fig. 16 and Fig. 17 plot the percentage residence time spent in different technologies, against an increase in the density of the highest throughput network (which is WiMax in this scenario). We only present results for varying the number of WiMax networks for MTCA and full knowledge schemes. The results of the other cases follow the same trend. We see that as the number of WiMax networks increases, other lower-throughput networks are effectively *squeezed* out. It is worth noting that this even occurs in cases where other high-throughput but low coverage networks such as WiFi are present. From this result, we hypothesize that high-throughput networks tend to dominate while maximizing the utility, provided that there are a fairly large number of such networks available.

VI. Related Work

The work that comes closest to our study is the utility function based vertical handoff proposal in [14]. We follow a similar notion of constructing a generic utility function that can be application specific, but augment it by defining a full knowledge scheme that can serve as a benchmark for evaluating any other schemes. Our work is also substantially different because we evaluate multiple heuristics that aim to approximately maximize the overall utility derived during the entire application activity period, while [14] considers the maximization of only the immediate utility for making decisions. We also study the performance with respect to many network design parameters, which, to the best of our knowledge, has not been done so far.

[16, 17] also follow a similar approach of multi-factor minimum-cost based utilities, and introduce the notion of network elimination that dictates whether or not a candidate network can satisfy certain application level constraints. The ideas can be applied to our approach as well by defining edge costs of – *infinity*.

[18] proposes the use of signal strength measurements to make switching decisions. Although we follow a different approach through the use of utility maximization to make switching decisions, we implicitly assume a signal strength sensing mechanism to be in place. This mechanism triggers a decision making process whenever it detects a strong signal from a new network, or a progressively fading signal from the current network. If desired, our utility function based approach can easily incorporate the signal strength as a selection parameter as well.

A fuzzy decision making approach for heterogeneous wireless networks is described in [19]. In our work, we do not directly apply fuzzy logic, but we have abstracted it away into the movement and network characteristic prediction modules. We describe this further in the next section.

VII. Discussion and Future Work

As smart phones and heterogeneous networks continue to proliferate, the need for intelligent algorithms for making mobility decisions will continue to grow. We have shown in our study that making decisions is not an easy task and involves accurate knowledge and prediction of many parameters. A major requirement is therefore to develop good movement and network characteristic prediction tools, as described in Section II. Even if accurate prediction tools are not developed, there is still scope for building more intelligent heuristics that keep adapting themselves based on past experience. Our formulation of the decision graph can be used as a data structure to store all past decisions, evaluate the efficiency of the heuristics used in making the decisions, and adapt the heuristics accordingly.

In our study, we have assumed client-side decisions to always prevail. However, this might not always be the case because network providers can exercise admission control and force the user to stay in a particular network. One solution is to generalize the utility function to a multi-objective social welfare scheme where both the client as well as the infrastructure utility is simultaneously maximized. However, this approach requires knowledge of the utility preferences of the providers, which may change with time and might not be available either. It will be interesting to see how different heuristics perform in the presence of such constraints.

Building RTT estimation tools that function during mobility requires the establishment of a mapping from geographical to network coordinates [10, 11]. We are currently exploring this further to differentiate between regions where network coordinates do not depend on the geographical coordinates, and regions where a dependency does exist. An extensive literature survey has indicated promising trends.

So far we have assumed the throughput to be independent of the RTT between anchor points and the mobile. Although this is likely to be true for the scenarios we model, we plan to conduct experiments in the TCA scenario where anchor points may be far from the mobile users, and may act as data caching proxies as well. Therefore, a good choice of anchor point will become important and noticeable in the experiments.

We are also working on the development of an analytical model that can mathematically explain our results, and for example, find the drop-off point at which the WiFi density suffocates EvDO completely.

VIII. CONCLUSIONS

From the simulation results, we are able to conclude that aggressive and greedy heuristics for choosing the best network perform better than stable schemes that try to minimize the switching costs. We also observe that if all anchor points are collocated within small RTTs from the access networks, then choosing the closest anchor point is not always better than a conservative approach that chooses one anchor point and sticks to it.

Our experiments on the density of high throughput networks suggest interesting findings that networks like WiFi and WiMax may proliferate to such an extent that lower throughput albeit higher coverage area networks like EvDO may not be required at all.

IX. REFERENCES

[1] M. Stemm, R. Katz. "Vertical Handoffs in Wireless Overlay Networks," *In Mobile Networks and Applications,* Volume 3, Number 4, Pages 335-350, 1998.

[2] K. Ahmavaara, H. Haverinen, R. Pichna. "Interworking Architecture between 3GPP and WLAN Systems," *In IEEE Communications Nov 2003.*

[3] M. Buddhikot, G. Chandranmenon, S. Han, Y.W. Lee, S. Miller, L. Salgerelli. "Design and Implementation of a WLAN/CDMA2000 Interworking Architecture," *In IEEE Communications Nov 2003.*

[4] G. M. Koien, T. Haslestad. "Security Aspects of 3G-WLAN Interworking," *In IEEE Communications Nov 2003.*

[5] Charles Perkins. "IP Mobility Support for Ipv4," *http://www.ietf.org/rfc/rfc3344.txt*, Aug 2002.

[6] H. Soliman, C. Castelluccia, K. El-Malki, and L. Bellier**.** **"**Hierarchical Mobile IPv6 Mobility Management (HMIPv6)," *Internet draft, work in progress*, Feb 2004.

[7] I.F. Akyildiz, Jiang Xie, S. Mohanty. **"**A survey of mobility management in next-generation all-IP-based wireless systems**,"** *In IEEE Wireless Communications*, Volume: 11, Issue: 4, Aug. 2004, Pages: 16 - 28.

[8] A. Seth, P. Darragh, S. Keshav. "A Generalized Architecture for Tetherless Computing in Disconnected Networks," *Work in progress.*

[9] J.F. Kurose, K.W. Ross. "Computer Networking," *Addison Wesley*, Third Edition, Page 271.

[10] F. Dabek, R. Cox, F. Kaashoek, R. Morris. "Vivaldi: A Decentralized Network Coordinate System," *In Proceedings of SIGCOMM,* August 2004.

[11] E. Ng, H. Zhang. "Predicting Internet Network Distance with Coordinates-Based Approaches," *In Proceedings of INFOCOM*, June 2002.

[12] D. A. Maltz, P. Bhagwat. "MSOCKS: An Architecture for Transport Layer Mobility," *In Proceedings of the 1EEE INFOCOM '98*, Pages. 1037-1045, San Francisco, CA, 1998.

[13] A. Snoeren. "A Session-Based Approach to Internet Mobility," *PhD Thesis, MIT, http://www.cse.ucsd.edu/~snoeren*, Dec 2002.

[14] H. Wang, R. Katz, J. Giese. "Policy-Enabled Handoffs across Heterogeneous Wireless Networks," *In Mobile Computing Systems and Applications,* 1999.

[15] R. Sedgewick. "Algorithms in Java, Part 5: Graph Algorithms," *Addison Wesley*, Third Edition, Section 21.6.

[16] J. McNair, F. Zhu, "Vertical Handoffs in Fourth Generation MultiNetwork Environments," *In IEEE Wireless Communications,* Vol. 11, No. 3, June 2004, Pages 834-848.

[17] F. Zhu, J. McNair "Optimizations for Vertical Handoff Decision Algorithms," *In Proceedings of WCNC*, Pages 867-872, March 2004.

[18] R. Inayat, R. Aibara, K. Nishimura, "A Seamless Handoff for Dual-interfaced Mobile Devices in Hybrid Wireless Access Networks," *In Proceedings of AINA 2004,* Vol. 1, March 2004.

[19] P.M.L Chan, Y.F. Hu, R.E. Sheriff, "Implementation of Fuzzy Multiple Objective Decision Making Algorithm in a Heterogeneous Mobile Environment," *In IEEE Wireless Communications and Networking Conference,* March 2002.