Mobility pattern detection for Quality of Service enhancements

Stefan Michaelis

University of Dortmund, Germany

E-mail: `stefan.michaelis@udo.edu`

Abstract

The classical network analysis using statistical methods has been accompanied during latest research by enhanced data mining methods for quality of service guarantees.

The EU-funded project INTERMON ([1]), which is about to end in spring 2004, has spent some of its efforts on the detection of regular patterns inside interdomain traffic traces. These detection and analysis algorithms targeted at large scale and high bandwidth interdomain traffic in fixed networks. As mobile networks providing higher bit rates like the well known GSM networks become available, a new dimension causing problems according to user movements raises.

This paper presents some ideas for extending the plain trace pattern detection with the analysis of user movements in mobile networks.

Quality of Service (QoS) has become one of the major buzz words in the competition between telecommunication companies. Networking independent of time and place in combination with sophisticated applications like video conferencing raises the demand for guaranteed connection parameters. Beside the simple bit rate needs parameters like latency, transmitting time, jitter etc. depending on the certain type of application are extended for wireless networks with the mobility factor. As some techniques target at the estimation of network loads, this paper targets at the estimation of user movement in combination with the generated traffic.

This paper is divided into sections showing the major problems and approaches to mobility detection. Section 1 discusses the general characteristics of user equipment mobility. The following section 2, Simulation, presents a frame work especially targeting at movement of user equipment. The simulation delivers the input data for the further network and movement analysis. Section 3 describes some general learning approaches which can be useful to detect regular patterns during user movement. Section 4 completes the details of the former sections by putting the single topics into their related contexts.

The mobile network and mobility models discussed here do not concentrate on a certain kind of network beside the assumption that the network managed is a cellular based network instead of mobile ad-hoc networks. The ideas presented concentrate on the common denominators of mobile networks using base stations as access points.

1. Mobility

The major difference of mobile networks (GSM, UMTS, WLAN Hotspots) to fixed networks consists of the mobility of the end users (and sometimes even the sources, e.g. peer-to-peer applications where both users are mobile). Mobility of the user equipment makes it hard to keep the network consistent and to always determine valid routes.

GSM networks evolved over a long period of time and many of the problems concerning mobility have been solved here. Protocols targeting at routing and equipment location (e.g. MAP for GSM networks or MobileIP for Internet based networks) provide a sufficient solution for the mobility handling. In contrast the next generation networks allow due to higher available bit rates applications similar to those known from the classical fixed networks. These applications have different demands on maximum/average throughput, latency, jitter and reliability. These networks are not limited to a certain set of QoS-classes like voice and fixed bit rate data services and therefore integrate many different applications into a single net, raising the problems of reliable connections in combination with user mobility.

1.1. Movement classification

The problems occurring during user movement can be divided into different classes representing selected levels of complexity. As the user moves, in some cases cell borders will be crossed which triggers rerouting events to ensure data still reaches the user in time and without data losses. The cell changes can be differentiated by dividing the events into *horizontal* and *vertical* cell changes.

See figure 1 for an impression of different events which can possibly happen while the user crosses certain areas and triggers horizontal changes.

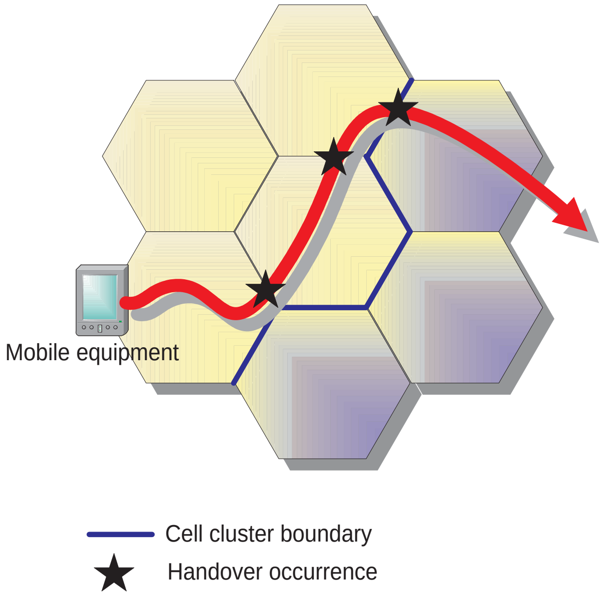

Figure 1. User mobility

Horizontal change events are the main events usually occurring while the user moves. The simplest case is the user moving only in a small area, staying in the same cell for the whole period of time while the mobile application is running. From the rerouting point of view no events happen, the route stays the same. The first event causing a rerouting occurs when the user crosses the border of a cell and connects to the next neighbouring cell. Most core networks controlling the base stations are hierachically clustered, e.g. for the UMTS case a Radio Network Controller responsible for a cluster of NodeBs and these RNCs also controlled by a set of Support Nodes. Therefore a change into a cell which belongs to the same radio controller is not or not as critical as the change crossing the border to another base station cluster. Similar assumptions of course apply to WLAN based hot spot networks. The impact on the routing algorithms and the side effects from and to other traffic inside the core network strongly relates with the border which is crossed during the handover between to radio cells.

In addition to the horizontal movement, mobility along a vertical path is also possible. Vertical movement in this context does not mean that the user is physically moving vertically, but that the connection and data transmission is switched seamlessly between different heterogenous mobile networks. From the current point of view in Europe three major mobile networks are available: GSM (and its extensions like HSCSD, GPRS etc.), UMTS (or third generation mobile network) and WLAN hot spots. Some companies see these technologies competing, others as in a friendly co-existence. These technologies have different reachability and performance parameters. GSM mobile networks have a reachability of near to 100%, UMTS networks much lower (e.g. in Germany by the end of the year 2003 not much more than 25%) and WLAN provides only hot spot access in selected areas started from public or private initiatives. In contrast the maximum available bit rate varies starting from WLAN at 54 MBit/s over UMTS down to the 9,6 KBit/s of GSM. Vertical movement describes the seamless handover process between two or more of these mobile networks. As this a hard to manage from a radio interface position because of the different frequency ranges, transmission modes etc. only very few equipment is able to handle more than one technology. Additionally the protocols are usually not able to guarantee a lossless handover and correct charging between these networks.

But as chip design evolves the combination of heterogenous networks surely will add a new dimension of complexity to handover and routing mechanisms as well as to add more data for pattern detection algorithms.

1.2. Movement characteristics

At the first sight movement of users leads to an unpredictable change of routing and path information through time. But having a look at experiments analysing mobility, the main streams of movement are not totally random. Thinking of people driving along in a train or considering concentration in hot spots like stadiums.

In [6] movements of users in ad-hoc networks were tracked. Here the measurements showed that in many cases the movement stopped for a certain time. So detecting a user staying in a cell and letting the time in this cell reaching a threshold, it becomes very likely that the user will stay in this cell much longer. For mobility management this reduces the degree of randomness and shows some regularities in the movement of the users. Similar assumptions apply to cellular networks.

This leads to a summary of the movement characteristics:

- Partly linear movement
- Stability for a distinct time interval
- Predictable behaviour

To ensure the quality of service class of running applications during user movement the rerouting mechanism has to be fast and reliable. The integration of different types of applications with different demands makes it difficult to fulfil each of the QoS requests, even if there would be enough capacity. Our current research attempts to distribute and manage the network capacity more efficiently using knowledge about the movement of the users and precognitivly reserving capacity.

2. Simulation

The pattern detection and data mining algorithms need as an additional input source traces of mobile user's movements. All data mining and pattern detection algorithms have one thing in common: They strongly depend on the quality of the input data. This means a sufficient amount of data, correct data and applying the right filters before importing it into the pattern detection algorithms (e.g. treatment of missing values, normalisation...)

To allow the generation of movements for complex networks a specialised simulator is under development (MOOSE, MObile Object Simulation Environment, see figure 2), which allows to apply different movement models (e.g. random or path-following walkers) to the network users. Many available simulators concentrate only on single cell networks and respect movement only by incoming and leaving users. Others targeting at ad-hoc networks focus on mobility without respecting impacts on routing and core networks. To address these problems, the proposed simulation environment is currently under development.

The simulation environment has beside the network protocol simulation to provide detailed information about applications used on the mobile equipment and different models for movement simulation. The following subsections give an overview for different approaches on these fields.

2.1. Movement models

Here some models for movement and mobility simulation are discussed to allow the decision about realistic or artificial movement patterns. To check the quality of pattern detection algorithms, each result has to be validated against the realism of the selected movement scenario. See [2] and [3] for another discussion of movement models.

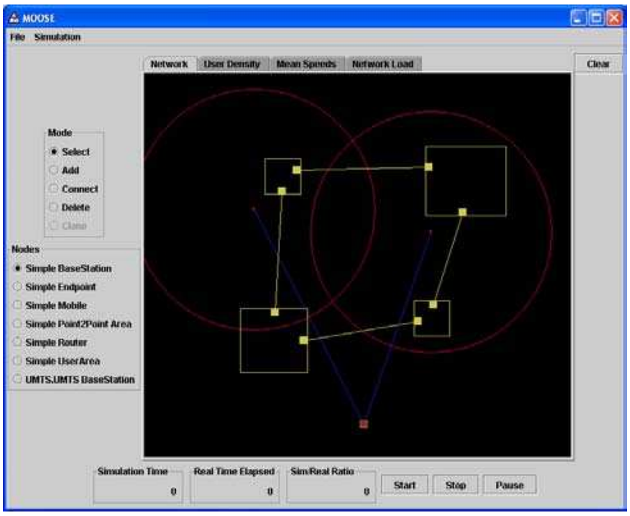

Figure 2. MOOSE simulator screenshot

The simplest movement model thinkable of is a random walk model. Here the position changes of the users appear random along time.

This model is easy to implement, as the users inside the areas do not correlate with each other. Figure 3 shows a result picture from simulating random user movement in a single large area. The figure shows the mean user density for each position possible in this area, the lighter the colour, the more users have stayed at this position for certain points of time (heat distribution model). This figure also shows one of the biggest disadvantages using a plain random walk model: If direction and speed changes of the users occur independently of their position they tend to concentrate near the borders of the area. Here they are blocked till the random direction changer tries to walk into a valid direction inside the area.

More sophisticated random walk models therefore use some kind of determinism to change directions if the come near the area borders and keep speed and direction constant for a short interval of time.

As experience tells random walk models build a limited realistic movement model. Most users have a target while moving and directly head for their destination. Random walk models are suitable for large open places like market places in inner cities or when the granularity of movement does not necessarily need have to respect natural boundaries like buildings etc.

This leads to a more complex movement model simulating path following mobility (to simulate streets and railways). The figures 4 to 6 show a sample mobility network consisting of four small area and connecting streets. First figure is the plain movement networks (not considering base stations or core network), second figure the user density as before and the third figure the mean speeds of the users at their distinct positions (for sake of simplicity speed remained unchanged during simulation time).

The density picture shows how the main points of user concentration change from the areas to the connecting streets, where the users have to move on a smaller area in comparison to the larger areas at the streets endpoints.

Movement models considering streets give a better approximation of real user movements, do they allow to build path following models where the users have to reach a certain location before they can move on to the next one. This enables complex travel modelling.

Nevertheless, other movement models try to further enhance the simulation. Introducing correlations between the users inside an area/street gives the possibility to simulate crowded areas and even traffic jam. Gravity models calculate a value representing the attractivity of an area. The more attractive an area is, the higher the probability that more users will try to move into this area. This could be used to model regular travellers, moving in the morning from home to work and in the evening back to home.

As these enhanced movement models consume much more computational power due to the calculation of correlations between area and/or users, each time before simulation it should be checked if the gain in finer granularity and realism is worth the raised effort.

2.2. Application models

Modelling of applications is another important part for a realistic simulation as the applications are responsible for some of the highest loads inside the network ([4], [5]). In combination with user mobility this leads to data flows on single links with a high jitter as the data path changes along time.

Critical point here is how to estimate the applications and the traffic generated by them on mobile networks. The classical GSM case has been comparably easy, voice and data calls using only a single channel and time slot for the whole application running time. These cases could be treated under certain circumstances similar to telephony blocking probability. For mobile networks allowing higher data rates and variable bit rate traffic these assumptions seldom hold.

The estimation of possible applications in third gen-

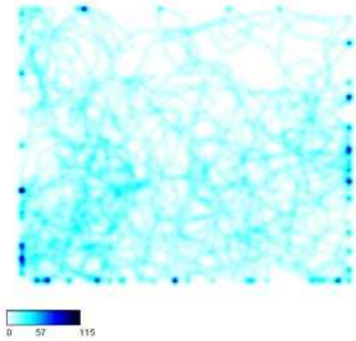

Figure 3. User distribution and density

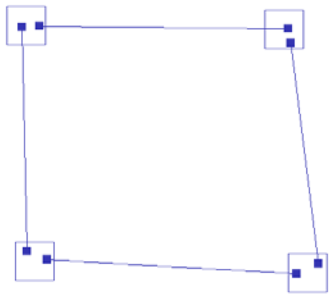

Figure 4. Sample movement area network

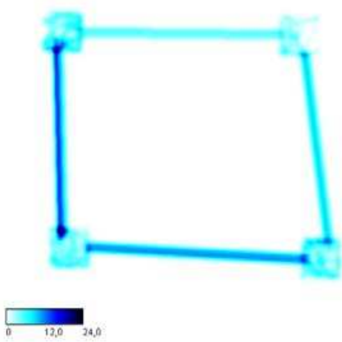

Figure 5. User distribution and density

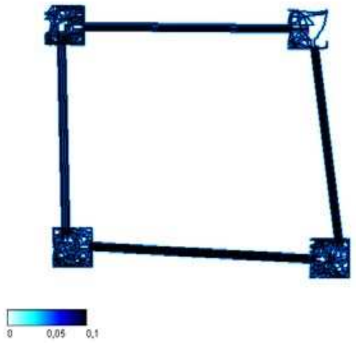

Figure 6. Mean speeds

eration mobile networks is critical as at the current point of time no one has a complete overview about applications available. UMTS and WLAN started to bring the services to the user in the same quality he expects from the wired Internet. But will the applications look the same? Will their characteristics be the same? Only few people where able to see the business potential of the GSM Short Messaging Service in Europe as GSM started. For UMTS networks the Multimedia Messaging Service tries to renew the success of SMS, but will it be that way?

Beside the question if the mobile applications will be the same as the known applications, the question rises if the characteristics stay the same. Will the holding time be the same? The usage frequency? Finding here realistic models ensures a realistic simulation output which is in addition important for the result quality of pattern detection algorithms.

What we can assume from the well known classical applications are that several traffic classes usually occur parallel in the same network:

- Low to medium bit rate and high latency services (e.g. email services or ftp downloads)
- Low to medium bit rate services with minimum latency (e.g. voice over IP applications)
- High bit rate services with minimum latency (e.g. video conferencing)

The simulation has to consider these traffic classes, estimate the usage of the different applications and inject the according traffic into the network.

3. Pattern detection

Pattern detection, data mining, artificial intelligence and statistical learning algorithms are a wide field of research. Areas of application start from medical analysis up to automated starfield clustering and network trace analysis. All these algorithms target at extracting information in large data sets which are not obvious. Many different data mining algorithms are available: Decision trees, support vector machines, neural networks etc. In most cases it is personal taste which algorithm suits best. The selection of concrete pattern detection algorithms is out of scope of this paper, a very good introduction to the different classes can be found in [7].

As experience shows the selection of the implementation of the pattern detection algorithm is not as critical as the availability of high quality input data (compare [8]). If inside the data regular patterns are available, most pattern detection algorithms are able to find them, regardless of their algorithmic background. The quality of the patterns detected may differ inside a certain threshold, but experience also showed that the more accurate algorithms have as a trade-off a longer computational time or larger memory dumps. Experiments on plain trace data showed a difference in load forecast of 10%-20%, but all lay above a 50% chance of estimating the right trend of link load.

Assuming some kind of regular behaviour in user movement and application usage this offers the chance of detecting patterns inside the trace data. These patterns could then be used to recognise problematic areas with high network traffic loads before they occur or estimate the future movement of the user to allow a pre-reservation of capacity on the links which will be involved along the users path. Similar techniques like those researched in [9], [10] could be applied here to learn network behaviour and especially user movement along time.

The possible patterns detected and used here are not limited to the plain traffic trace analysis. For example, if the case regularly happens that the movement of many people is heading to a certain location (e.g. a large stadium) the algorithms can recognise this fact and learn that the users will gather there and stay there for a while. Another example would be to learn that passing some locations will certainly lead to a movement into a fixed direction (thinking of entering a train at a railway station). Whatever regular behaviour of user movement exists, in many cases this can be detected and used to get an impression of the future behaviour of the associated users.

For the application in a real network and therefore also in the simulation the pattern detection algorithms have to be integrated into the network to either collect trace data (movement and traffic) and inject the information they extracted back into the routing tables. For the simulation this means that each node inside the simulated network (router, base station, application server...) has the ability to use protocol stacks implementing standard network protocols (IP, TCP etc.). This is also the point where the pattern detection algorithms apply. As each packet (or event like handover occurrence) is parsed through the node's protocol stack, the pattern detection algorithms can gather information about the load in nodes as well as user mobility. Important here is that the pattern detection algorithm share the same visibility (packets and events) as the routing node where the algorithm is installed.

4. Mobility tracking lifecycle

Figure 7 shows an overview of the complete lifecycle of the mobility tracking and reaction following the cross-industrial software process model for data mining (CRISP-DM, [11]).

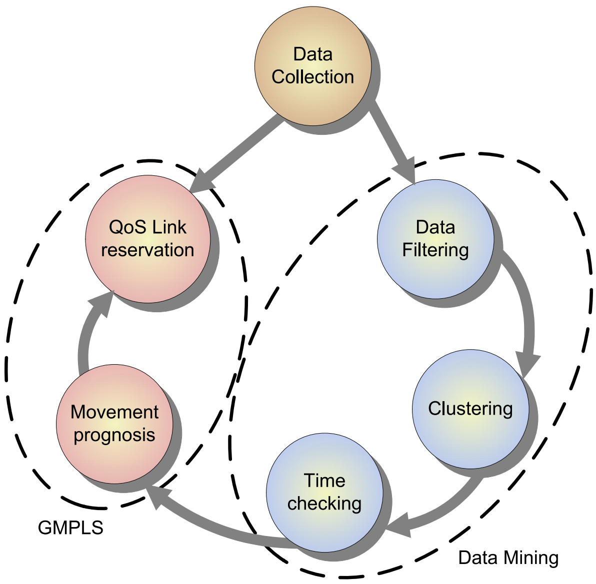

Figure 7. Process model for movement and link reservation analysis

To initialise the process it starts at the top with the *data collection*. Measurements or simulations of the cell loads and user movements serve as an input to identify the state of the network. The traces are entering the data mining section of the process with *data filtering*. The filtering guarantees the validity of the entered data as it is processed. This includes smoothing the data by eliminating peaks and gaps in the measurements, which seem not to belong to the usual and repeating patterns. For the time dependent cluster prognosis only recurring patterns are of interest. Traces repeating very often can be eliminated since they have no high impact on the quality of the clustering process. Filtering should enhance the quality of the data input for the clustering process and reduce the complexity of the measured data to support the following cluster process.

Clustering is the state of finding clusters in the filtered data traces. It identifies the range and weight of areas with a considerable traffic occurrence. This delivers new traces along time, but with a coarser granularity than single cell measurements.

The last stage in the data mining part performs the *time checking*. This part brings the clustered cell traces in their timely dependency. Taking time into account leads to an estimation of general mobile user behaviour.

Afterwards the process uses the discovered cluster models to make a *movement prognosis* on the hot spots. Here begins the routing part of the process, which is responsible for building new or additional paths and reserving capacities based on data mining estimations. The movement prognosis leads to the knowledge, which areas will have an increased demand on bit rate and reliability regarding to the quality of service classes in the near future. The rerouting has to happen fast and transparent to the network services and protocols above, so as an example GMPLS could be selected as the binding middle layer between the routing and the physical link layers.

It is job of the GMPLS adjustment to make the transparent *QoS link reservation* to fulfil the user's needs in these areas. This part is also based on the ongoing data collection process since this data is needed to recognise the mobility patterns detected in the data mining stages. Continuing the data collection process and feeding this new data into the pattern detection algorithms again ensures the actuality of movement prognosis even for changing mobility behaviour.

5. Conclusions

Managing networks and setting up routing is a complicated task. Great efforts have been spent to encourage the development of adaptive and flexible routing mechanisms ensuring the quality of service classes suitable for each application. Latest research targets at enhanced data mining and pattern detection algorithms to analyse where possible problems inside the network occur and give a forecast on future load situations.

This paper presents some ideas on how to extend these pattern detection techniques to the field of emerging high bit rate mobile networks. The main topics needed to provide sufficient data as an input for data mining algorithms as well as the overall process starting from simulation over pattern detection up to pro-active network management have been discussed.

Of course there is still much work to do and many things have to be verified before it could be seen as a complete framework for simulation, data mining and adaptive routing. Intention of this paper was not only to present the current and future research efforts on this area, but also to encourage the reader to try pattern detection algorithms in similar research field to gain knowledge about the behaviour of network traffic.

References

[1] INTERMON project: *Advanced architecture for INTERdomain quality of service monitoring, modelling and visualisation*, http://www.ist-intermon.org/

[2] Markoulidakis, J. et al.: *Mobility modelling in third generation mobile telecommunication systems*, IEEE Personal Communications, 1997

[3] Zoonozi, M.; Dassanayake, P.: *User mobility modelling and characterization of mobility patterns*, IEEE Journal on selected areas in communications, 1997

[4] Mah, B.: *An Empirical Model of HTTP Network Traffic*, IEEE Transactions Communications, PP. 410-422, 1997

[5] Mena, A.; Heidemann, J.: *An empirical study of real audio traffic*, Proceedings of IEEE Infocom, March 2000

[6] Toh, Chai-Keong: *A Novel Distributed Routing Protocol To Support Ad-Hoc Mobile Computing*, IEEE 1996

[7] Holsheimer, M.; Siebes, A.: *Data Mining: the search for knowledge in databases*, Centrum voor Wiskunde en Informatica 1994, Amsterdam

[8] Bindrim, W.; Michaelis, S.: *Approaches to Pro-Active Network Management using Tele-Traffic Theory and Data Mining Techniques*, Proceedings of the World Multiconference on Systemics, Cybernetics and Informatics; IIIS, Orlando 2001

[9] Miloucheva, I.; Anzaloni, A.; Müller, E.: *A practical approach to forecast Quality of Service parameters considering outliers*, First international workshop on Inter-domain performance and simulation, IPS 2003, Salzburg 20-21 February

[10] Miloucheva, I.; Hofmann, U.; Aranda Gutiérrez, P.A. *Spatio-Temporal QoS Pattern Analysis in large scale Internet environment*, Conference on Multimedia Interactive Protocols and Systems, 18-21 November 2003, Napoli, Italy

[11] Chapman, Pete et al.:*The CRISP-DM Process Model*, CRISP-DM consortium 1999